Theor Chem Acc (2010) 126:371–382
DOI 10.1007/s00214-009-0699-5

REGULAR ARTICLE

The optimum contraction of basis sets for calculating spin–spin coupling constants

Frank Jensen

Received: 6 October 2009 / Accepted: 10 November 2009 / Published online: 28 November 2009

Abstract The previously proposed pcJ-*n* basis sets, optimized for calculating indirect nuclear spin–spin coupling constants using density functional methods, are re-evaluated for finding the optimum contraction scheme as a compromise between computational efficiency and minimizing contraction errors. An exhaustive search is performed for the H_2, F_2 and P_2 molecules, and candidates for optimum contraction schemes are evaluated for a larger test set of 21 molecules. Using the criterion that the contraction error should not exceed the basis set error relative to the basis set limit, the optimum contraction is defined for each basis set. The results show that it is difficult to contract basis sets for calculating spin–spin coupling constants to any significant degree without losing the inherent accuracy. The work provides guidelines for searching for optimum contraction schemes for other properties and/or at theoretical levels where a systematic search is impractical.

Keywords Basis set contraction · Spin–spin coupling constants · DFT

1 Introduction

The calculation of nuclear magnetic spin–spin coupling constants is a challenging theoretical task, as the results depend significantly on the quality of the wave function, the basis set, the molecular structure, as well as vibrational and solvent effects [1–13]. We have in recent work shown that it is possible to design basis sets that systematically reduce the basis set error in the calculation of spin–spin coupling constants by either density functional [14] or high level wave function (CCSD) methods [15]. These pcJ-*n* ($n = 0, 1, 2, 3, 4$) and ccJ-pVXZ (X = D,T,Q,5) basis sets are derived from the corresponding pc-*n* [16–19] and cc-pVXZ basis sets [20] by addition of high-exponent basis functions to the uncontracted basis set followed by a recontraction. The contraction was done using guidelines from previous work, and it was found that only a small number of basis functions could be contracted without loss of accuracy. This leads to a relatively high computational cost which limits the use for calculations on large systems, and in the present work we describe a more systematic search for the optimum contraction of the pcJ-*n* basis sets.

The notation for the spin–spin optimized polarization consistent basis sets is pcJ-*n*, where *n* indicates the level of polarization beyond the atomic system, i.e. pcJ-0 is unpolarized, pcJ-1 includes a single polarization function, pcJ-2 has two types of polarization functions, etc. The uncontracted pcJ-*n* basis sets with $n = 0, 1, 2, 3, 4$ provide a systematic convergence towards the basis set limiting results, which allows a definition of the inherent error at a given *n*-level as the error relative to the basis set limit. Basis set contraction improves the computational efficiency, but introduces additional errors as the basis set becomes less flexible with respect to describing differences due to the molecular environment. The inherent error at a given *n*-level provides an objective measure of how strongly the basis set can be contracted. The goal is therefore to find the highest level of contraction where the contraction error is less than, but comparable to the inherent basis set error, as there is no point in reducing the contraction error to a small value for a basis set with a large inherent error.

There are two limiting types of basis set contraction, denoted segmented and general [21–23]. In a segmented

F. Jensen (✉)
Department of Chemistry, University of Aarhus,
Langelandsgade 140, 8000 Århus, Denmark
e-mail: frj@chem.au.dk

contraction, each primitive basis function is only allowed to contribute to one contracted function, while in a general contraction, all primitive functions are allowed to contribute to all contracted functions. Segmented contractions are thus a subset of general contractions. In a segmented contraction, the partition of the primitive functions into contracted functions is usually defined a priori, and the basis set exponents and contraction coefficients are normally optimized simultaneously, while in a general contraction the basis set exponents are optimized for the uncontracted basis set, and kept fixed in the subsequent contraction. The contraction coefficients for a general contraction are normally obtained from atomic SCF coefficients, but have in some cases been obtained from molecular calculations [24–26]. The latter may improve the performance for systems resembling the reference molecule, but can potentially bias the results for a more diverse set of systems.

The optimization function for basis sets designed to describe properties related to the potential energy surface, like molecular structures and vibrational frequencies, is an atomic or molecular energy, and the contraction error can be quantified by an energetic criterion. As the core orbitals are relatively insensitive to the molecular environment, the basis functions with large exponents can be contracted with little loss of accuracy, while the functions with small exponents describing the valence orbitals must be left uncontracted. The optimum general contraction can thus be decided by including more and more valence functions into the contracted functions until the contraction error exceeds the inherent basis set error. For basis sets designed to describe molecular properties, the selection of contraction strategy is less clear, since the contraction error is defined by the molecular property. Spin–spin coupling constants are sensitive to the description of the electronic structure near the nucleus and contain contributions from four different operators, each of which have different basis set requirements, and a sequential contraction based on the magnitude of basis set exponents is therefore not necessarily the optimum strategy. Lacking a physically motivated strategy for the basis set contraction, we have in the present work opted for a brute force search over a large number of possible contraction schemes. For the smaller basis sets (pcJ-0, -1 and -2) a practically exhaustive search has been performed, and based on these results, a search over all reasonable contraction schemes has been done for the pcJ-3 and -4 basis sets.

The contraction error depends on the molecular system, but an unacceptable contraction error for a single molecular system is sufficient to disqualify the contraction scheme for all systems. We have thus used a few systems for discarding many unacceptable contractions schemes, and only subjected a much smaller number of possible contraction schemes to a larger variety of molecular systems. The systems used for the initial screening of contraction schemes should have coupling constants containing significant contributions from all four operators, and be small enough that thousands of calculations, of which some employ quite large basis set, should be feasible. We have in the present work used the H_2 (B3LYP basis set limit value = 321 Hz), F_2 (B3LYP basis set limit value = 15,516 Hz) and P_2 (B3LYP basis set limit value = 354 Hz) systems, which fulfill these criteria.

2 Computational details

The indirect spin–spin coupling constant is calculated using Eq. 1.

$$\mathbf{J} = \langle \Psi_0 | \mathbf{H}^{\mathrm{DSO}} | \Psi_0 \rangle + \sum_{k}^{\mathrm{PSO,SD,FC}} \sum_{n \neq 0} \frac{\langle \Psi_0 | \mathbf{H}^k | \Psi_n \rangle \langle \Psi_n | \mathbf{H}^k | \Psi_0 \rangle}{E_0 - E_n} \tag{1}$$

The four operators [diamagnetic spin–orbit ($\mathbf{H}^{\mathrm{DSO}}$), paramagnetic spin–orbit ($\mathbf{H}^{\mathrm{PSO}}$), spin–dipolar ($\mathbf{H}^{\mathrm{SD}}$) and Fermi-contact term ($\mathbf{H}^{\mathrm{FC}}$)] are defined in Eqs. 2–5 [27].

$$\mathbf{H}^{\mathrm{DSO}} = \sum_i \frac{g_A g_B \mu_N^2}{2c^4} \frac{\left(\mathbf{r}_{iA}^t \mathbf{r}_{iB} - \mathbf{r}_{iB} \mathbf{r}_{iA}^t\right)}{r_{iA}^3 r_{iB}^3} \tag{2}$$

$$\mathbf{H}^{\mathrm{PSO}} = \sum_i \frac{g_A \mu_N}{c^2} \frac{\mathbf{r}_{iA} \times \mathbf{p}_i}{r_{iA}^3} \tag{3}$$

$$\mathbf{H}^{\mathrm{SD}} = -\sum_i \frac{g_e \mu_B g_A \mu_N}{c^2} \frac{\left(\mathbf{r}_{iA}^t \mathbf{r}_{iA} - 3\mathbf{r}_{iA} \mathbf{r}_{iA}^t\right)}{r_{iA}^5} \mathbf{s}_i \tag{4}$$

$$\mathbf{H}^{\mathrm{FC}} = \sum_i \frac{8\pi g_e \mu_B g_A \mu_N}{3c^2} \delta(\mathbf{r}_{iA}) \mathbf{s}_i \tag{5}$$

Here, $\mathbf{r}_{iA}$ denoted the position vector between electron i and nucleus A, $\mu_{B/N}$ is the Bohr and nuclear magneton, $g_{e/A}$ is the electron and nuclear g-factor and $\mathbf{s}_i$ is the electron spin operator. The isotropic spin–spin coupling constant used in the present case is defined as 1/3 of the trace of $\mathbf{J}$.

All calculations have been performed with the Dalton [28] program package using the B3LYP exchange-correlation functional [29, 30], as this has been one of the more successful for predicting NMR properties [31, 32]. The GIAO technique has been employed to ensure gauge independence of the calculated results [33–36].

We have only considered contraction of the s- and p-functions, while polarization functions are left uncontracted. The contraction has been done using atomic SCF coefficients calculated with the B3LYP functional, and all exponents are kept fixed at their optimized value for the uncontracted basis sets. In the analysis phase for deciding the optimum contraction, we have used a measure of the contraction error defined as the sum of the absolute errors for each of the four contributions to improve the robustness, as this prevents artificial low contraction errors arising from random error cancellations. For the final quality analysis, we have used both the (unsigned) error in the total spin–spin coupling constants, as well as the sum of absolute values for each of the four contributions. The basis set limit has been taken as the results with the uncontracted pcJ-4 basis set.

3 Results and discussion

We will employ the following notation in the present paper. Contractions are classified according to the number of contracted functions as 1-contracted, 2-contracted and 3-contracted. The number of primitive functions in each contracted function is indicated in parentheses, i.e. (n,1,1,1,…) for a 1-contraction having n primitive basis functions in the contracted function, (n,m,1,1,…) for a 2-contraction having n and m primitive basis functions in the two contracted functions and (n,m,k,1,…) for a 3-contraction having n, m and k primitive basis functions in the three contracted functions. For a segmented 2-contraction, the sum of n and m must equal the number of primitive functions contracted ($n + m = N^c_{\mathrm{prim}}$), with the first contracted function containing the primitive functions 1 to n, and the second contracted function including the primitive functions $n + 1$ to $n + m$ and similarly for a 3-contraction. For a general 2-contraction, both n and m can vary between 2 and N^c_{prim} with the restriction that $n + m \geq N^c_{\mathrm{prim}}$, and simply denoting the number of primitive functions in each contracted function is not a unique classification. We will use the notation (n,m,1,1,…) [n_{start}, m_{start}] to indicate a contraction where the n_{start} to $n_{\mathrm{start}} + n$ primitive functions contribute to the first contracted function, while the m_{start} to $m_{\mathrm{start}} + m$ primitive functions contribute to the second contracted function and similarly for a 3-contraction. In order to facilitate the discussion, the first contracted function is taken to always include the innermost primitive function, i.e. n_{start} is always 1. This is not a restriction of the possible contraction schemes, but merely allows for an easier organization of the results. The next section will illustrate the notation for a specific example.

3.1 The F_2 molecule, contraction of the p-functions

It is generally found that the contraction of the s- and p-function space is virtually independent and can be considered separately. A search over possible contractions for a given primitive basis set is a combinatorial problem that is perhaps best illustrated by an example. The uncontracted pcJ-1 basis set for F is 9s5p2d in composition compared to 7s4p1d for the pc-1 basis set. The s- and p-functions for the latter were contracted to 3s2p based on the performance for molecular structures and energies, where it was found that the outer s- and p-function must be left uncontracted [16, 17]. The maximum contraction for the five p-functions in the pcJ-1 basis set is thus to two functions, where the inner four functions are contracted to one function, and the outermost function is left uncontracted. For the F_2 molecule, this leads to a contraction error for the spin–spin coupling constant of 340 Hz which is unacceptably large compared to the inherent basis set error of 87 Hz.

While the contraction to two p-functions is unique when the outermost function is left uncontracted, the contraction to three functions can be done in six different ways, as illustrated in Fig. 1. The first three corresponds to segmented contractions denoted (3,1,1), (2,2,1) and (1,3,1), while the latter three are general contractions denoted (3,3,1) [1,2], (3,2,1) [1,3] and (2,3,1) [1,2]. Note that these are the only unique contractions, as the primitive functions of a contracted function cannot be a subset of the primitive functions of another contracted function when only a single set of contraction coefficients is available (2p SCF coefficients). A general contraction corresponding to (3,4,1) [1,1], for example, is equivalent to a segmented (3,1,1) contraction, since the inner three functions of the second contracted function can be removed by a simple linear transformation [37]. The six possible contractions to three p-functions illustrated in Fig. 1 have contraction errors of

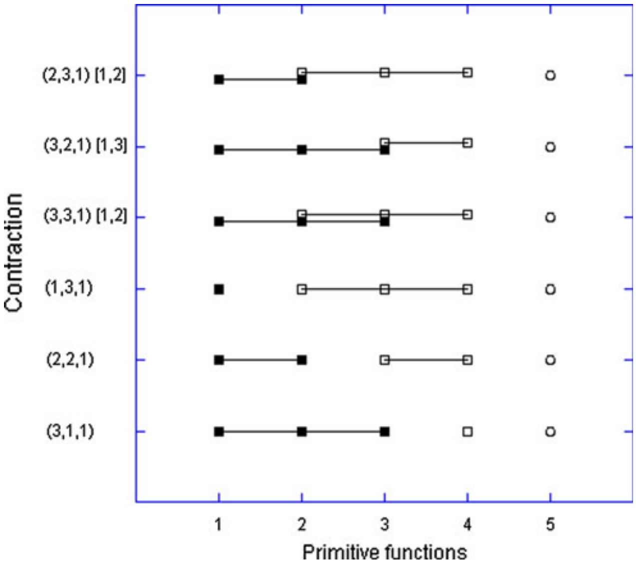

Fig. 1 Illustrating six different contractions of five primitive functions to three contracted functions. *Solid squares* indicate basis functions in the first contracted function, *open squares* indicate basis functions in the second contracted function, while *open circles* indicate the third (uncontracted) function. *Lines between the symbols* indicate that the ratio between these functions is fixed by the contraction coefficients

84, 167, 310, 1,574, 5,142 and 1,090 Hz, respectively. While the contraction error of 84 Hz for the (3,1,1) contraction is just below the inherent error of 87 Hz, it is still unacceptably large since the s-functions will contribute a similar contraction error.

There are five possible contractions of five p-functions to four functions, three segmented contractions denoted (2,1,1,1), (1,2,1,1) and (1,1,2,1) and two general contractions denoted (2,2,1,1,1) [1,2] and (1,2,2,1,1) [2,3]. The corresponding contraction errors are 4, 62, 169, 1,302 and 3,740 Hz, respectively, where the lowest value of 4 Hz is well below the inherent error of 87 Hz. The maximum contraction of the five p-functions in the pcJ-1 basis set is thus to four functions, and the optimum contraction corresponds to a segmented (2,1,1,1) contraction.

While the number of segmented contractions for a given combination of primitive and contracted functions is relatively small, the number of general contractions increases rapidly with the number of primitive functions. For the smaller basis sets, where a complete search is possible, we observe that combinations where high exponent primitive functions are left uncontracted, while smaller exponent functions are contracted, always have larger contraction errors than combinations where all the uncontracted functions are the outermost primitive functions. In the above example, the (3,1,1) and (2,2,1) combinations have lower contraction errors than the (1,3,1) combination, the segmented (2,1,1,1) contraction is better than (1,2,1,1) and (1,1,2,1), and the general (2,2,1,1,1) [1,2] contraction is better than (1,2,2,1,1) [2,3]. This is in line with normal principles for basis set contraction, but as mentioned in the introduction, it was not clear from the onset of this work that this strategy was also optimum for molecular properties depending on the near-nucleus region. We view the finding as the result of a zero-sum game: for a fixed number of contracted functions, the uncontraction of a core basis function dictates contraction of an inner-valence basis function. The valence region indirectly determines the spin–spin coupling constant by changing the electron density near the nucleus (this is the origin of the molecular dependence of the spin–spin coupling constant), and the error from contracting a valence basis function always exceeds the error from contracting a core basis function. In order to reduce the number of possible contractions, we have thus limited the search to combinations where uncontracted functions always are taken as the outermost primitive functions.

Using these principles for the p-functions for all the pcJ-*n* basis sets gives the results shown in Table 1. The inherent error for each basis set is given both as the error in the total J-value and as the sum of the absolute errors for each of the four contributions. For the pcJ-0 basis set, the two values are 185 and 1,647 Hz, showing that significant error cancellation occurs in the total J-value. It is in all cases observed that segmented contractions give a lower error than general contractions. Using the criterion that the contraction error must be below or comparable to the inherent error, gives the recommended contractions levels indicated in bold. For the pcJ-3 basis set, a contraction to 6p is suitable for F_2, but a 7p contraction is required for

Table 1 Contraction errors (sum of absolute errors for the four contributions in Hz) for 1- and 2-contractions of p-functions of the pcJ-*n* ($n = 0$, 1, 2, 3, 4) basis sets for the F_2 molecule

Basis set ($N_{primitive}$) [Inherent error] {Inherent absolute error}	Contraction	1-Contracted	2-Contracted	
		Segmented (*n*,1,1,…)	Segmented (*n*,*m*,1,…)	General (*n*,*m*,1,…) [n_{start},m_{start}]
pcJ-0 (4p) [185] {1,647}	2p	390 (3,1)	–	–
	3p	**9.1 (2,1,1)**	–	2,929 (2,2,1) [1,2]
pcJ-1 (5p) [87] {387}	3p	84 (3,1,1)	167 (2,2,1)	1,090 (2,3,1) [1,2]
	4p	**3.6 (2,1,1,1)**	–	1,382 (2,2,1) [1,2]
pcJ-2 (7p) [4.7] {28.1}	4p	42 (4,1,1,1)	25 (3,2,1,1)	117 (2,4,1,1) [1,2]
	5p	**6.1 (3,1,1,…)**	16.5 (2,2,1,…)	189 (2,3,1,…) [1,2]
	6p	0.4 (2,1,1,…)	–	218 (2,2,1,…) [1,2]
pcJ-3 (10p) [21.3] {21.3}	5p	37.1 (6,1,1,…)	6.6 (5,2,1,…)	16.8 (6,6,1,…) [1,2]
	6p	9.3 (5,1,1,…)	6.5 (4,2,1,…)	13.0 (2,5,1,…) [1,2]
	7p	2.0 (4,1,1,…)	**0.8 (3,2,1,…)**	15.0 (2,4,1,…) [1,2]
pcJ-4 (12p)	7p	7.8 (6,1,1,…)	4.1 (5,2,1,…)	5.4 (6,6,1,…) [1,2]
	8p	0.53 (5,1,1,…)	0.74 (4,2,1,…)	5.6 (2,5,1,…) [1,2]
	9p	0.18 (4,1,1,…)	**0.13 (3,2,1,…)**	4.4 (2,4,1,…) [1,2]

Only the combinations with lowest errors are shown. The recommended contractions are given in bold

The inherent error in brackets is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is the sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result other systems (Sect. 3.5). Both 8p and 9p contractions can be considered for the pcJ-4 basis set, but since the computational cost will be dominated by the multiple polarization functions (7d4f2g1h), the savings from a 8p contraction will be marginal compared to 9p. Since the pcJ-4 basis set is aimed at establishing basis set limiting results, we recommend the 9p contraction.

3.2 The F_2 molecule, contraction of the s-functions

The contraction of the s-functions follows the same principles as illustrated for the p-functions above, except that there now are two sets of contraction coefficients available (1s and 2s SCF coefficients). Contraction of the nine s-functions of the pcJ-1 basis set to four functions or less gave unacceptable large contraction errors (>173 Hz), and the search for an optimum contraction scheme is illustrated below for the contraction to five s-functions.

The simplest contraction scheme to five s-functions consists of one contracted function with five primitive functions, leaving the outermost four functions uncontracted, i.e. (5,1,1,1,1). There are two such 1-contracted functions, where either the 1s or 2s SCF coefficients are used for the contraction, with the associated contraction errors of 33.7 and 40.3 Hz shown in Table 2. It is perhaps slightly surprising that contraction with the 2s-coefficients gives almost the same result as contraction with the 1s-coefficients. This reflects that the (5,1,1,1,1) contraction only contracts primitive functions with large exponents (>100 for F) and that the two sets of coefficients are nearly linear dependent in the region close to the nucleus, but differ of course in the valence region.

At the 2-contraction level, the contraction is denoted (*n*,*m*,1,1,1), when *n* and *m* indicate the number of primitive functions (from 2 to 6) contracted with either 1s or 2s SCF coefficients. If only one set of coefficients are used for both contracted functions (1s,1s or 2s,2s), there are 13 unique combinations, of which three are of the segmented type while ten are of the general type. If both sets of contraction coefficients are used, there are 32 different contraction schemes when the innermost function always is contracted with the 1s coefficients (1s,2s), of which three are of the segmented type and 29 are of the general type. A further 31 contractions (one is redundant) can be generated if the innermost functions always are contracted with the 2s coefficients (2s,1s). There are thus a total of 89 possible ways for contracting nine primitive functions to five contracted functions when the outermost three functions are left uncontracted.

The 2-contracted combinations with the smallest contraction errors are shown in Table 2. The segmented contraction corresponding to (4,2,1,1,1) gives a lower error than (3,3,1,1,1) and (2,4,1,1,1), in agreement with usual principles for basis set contraction. The use of only 1s-coefficients for both contracted functions in the (4,2,1,1,1) combination gives a lower contraction error (28.8 Hz) than any combination involving 2s-coefficients (88.4, 31.6, 89.5 Hz). The best general contraction using 1s-coefficients for both functions is a (4,5,1,1,1) [1,2] contraction with an error of 43.8 Hz, while the use of 2s-coefficients for both functions gives a slightly larger contraction error (55.6 Hz). Significantly lower contraction errors are found when the two functions are contracted using different set of coefficients, i.e. 7.9 Hz for the (6,4,1,1,1) [1,2] contraction using (1s,2s) coefficients, and 6.8 Hz for the (6,3,1,1,1) [1,3] contraction using (2s,1s) coefficients. It is again slightly surprising that the use of 2s-coefficient for contracting the innermost primitive functions gives a slightly

Table 2 Contraction errors (sum of absolute errors for the four contributions in Hz) for 1-, 2- and 3-contractions of nine primitive s-functions to five contracted functions for the pcJ-1 basis set for the F_2 molecule

Contraction level	Contraction type	1s-coefficients	1s- and 2s-coefficients			2s-coefficients
		(1s) (1s,1s) (1s,1s,1s)	(1s,2s) (1s,1s,2s)	(2s,1s) (1s,2s,2s)	(2s,1s,2s)	(2s) (2s,2s) (2s,2s,2s)
1-Contraction	Segmented (*n*,1,1,1,1)	33.7 (5,1,1,1,1)				40.3 (5,1,1,1,1)
2-Contraction	Segmented (*n*,*m*,1,1,1)	28.8 (4,2,1,1,1)	88.4 (4,2,1,1,1)	31.6 (4,2,1,1,1)		89.5 (4,2,1,1,1)
	General (*n*,*m*,1,1,1) [n_{start},m_{start}]	43.8 (4,5,1,1,1) [1,2]	7.9 (6,4,1,1,1) [1,2]	6.8 (6,3,1,1,1) [1,3]		55.6 (5,5,1,1,1) [1,2]
3-Contraction	Segmented (*n*,*m*,*k*,1,1)	284 (2,3,2,1,1)	425 (2,3,2,1,1)	416 (2,3,2,1,1)	425 (2,3,2,1,1)	416 (2,3,2,1,1)
	General (*n*,*m*,*k*,1,1) [n_{start},m_{start},k_{start}]	81.7 (3,5,4,1,1) [1,2,4]	2.1 (6,2,7,1,1) [1,6,1]	6.1 (6,5,2,1,1) [1,1,6]	0.9 (6,5,6,1,1) [1,2,2]	32.9 (5,5,5,1,1) [1,2,3]

Only the combinations with lowest errors are shown smaller error than using the 1s-coefficients, although for practical purposes, the difference is not significant.

At the 3-contraction level, the contraction is denoted (n,m,k,1,1), where n, m and k can have values from 2 to 7, and each containing either 1s or 2s SCF coefficients. If the 1s-coefficients are used for all three contracted functions, there are 73 combinations of which three are segmented and 70 are of the general contracted type. Using also the 2s-coefficients, and exploring the combinations (1s,1s,2s), (1s,2s,2s), (2s,1s,2s) and (2s,2s,2s) increases the total number of combinations to 1,554. We have performed a complete search over all these combinations and those with the lowest contraction errors are shown in Table 2.

The best segmented 3-contraction is a (2,3,2,1,1) combination, which gives a contraction error of 284 Hz when using only 1s-coefficients, but significantly larger errors when the 2s-coefficents are included (416–425 Hz). General contractions using only one set of coefficients (1s,1s,1s or 2s,2s,2s) have contraction errors (81.7 and 32.9 Hz) that are significantly larger than those using mixed combinations (1s,1s,2s, 1s,2s,2s or 2s,1s,2s) (2.1, 6.1 and 0.9 Hz).

For deciding the optimum contraction scheme, it is important that the contraction is robust, i.e. low contraction errors that are due to fortuitous error cancellations for the target system(s) should be discarded. The best 3-contraction with the (1s,1s,2s) coefficient combination is (6,2,7,1,1) [1,6,1] with a contraction error of 2.1 Hz, where the first two contracted functions have only a single primitive function in common. Removing this function overlap to give a (5,2) segmented contraction increases the error to 79.5 Hz, showing that the (6,2,7,1,1) [1,6,1] contraction is not robust. Similar analyses suggest that 3-contractions in general do not offer an improvement over 2-contractions, as 3-contraction combinations with lower errors than 2-contractions are not robust. The optimum choice for a contraction of nine s-functions to five contracted functions is thus either a segmented or general 2-contraction, which have contraction errors (28.8, 7.9 and 6.8 Hz) well below the inherent error of 87 Hz. The final choice between a segmented or general contraction will be discussed in the following.

We have performed analogous complete searches for the pcJ-0, pcJ-1 and pcJ-2 basis sets with all 1-, 2- and 3-contractions to between three and seven contracted functions. For the larger pcJ-3 and pcJ-4 basis sets the number of possible contractions increases rapidly, especially at the 3-contraction level. For the pcJ-3 basis set, for example, the contraction of 16 primitive s-functions to seven contracted functions can be done in 339 different ways at the 2-contraction level, and ~18,000 different ways at the 3-contraction level. In order to limit the search space, we have restricted the number of combinations based on the analysis for the pcJ-0, pcJ-1 and pcJ-2 basis sets. As mentioned earlier, general 3-contractions that produce lower errors than 2-contractions are not robust, and we have therefore only considered segmented 3-contractions, in addition to segmented and general 2-contractions. The number of segmented 2-contractions is small, and we have considered both (1s,1s) or (1s,2s) combinations of contraction coefficients, while only the latter has been used for general 2-contractions. As shown in Table 2, the general 2-contractions using either (1s,1s) or (2s,2s) coefficients have significantly larger errors than (1s,2s) and (2s,1s). The latter two give very similar errors for the optimum contractions, and we have selected the (1s,2s) combination based on physical arguments, i.e. the innermost primitive functions are always contracted using the 1s-coefficients. For the segmented 3-contractions we have considered the (1s,1s,1s), (1s,1s,2s) and (1s,2s,2s) coefficient combinations.

Table 3 shows the lowest contraction errors for the F_2 molecule for each of the pcJ-n ($n = 0,1,2,3,4$) basis sets as a function of number of contracted functions. The optimum contraction for the pcJ-0 basis set is to four s-functions, with either a segmented (4,1,1,1) or general (5,3,1,1) [1,2] contraction with errors of 118 and 66 Hz, respectively. The optimum contraction for the pcJ-1 basis set is to five s-functions, with either a segmented (4,2,1,…) or general (6,4,1,…) [1,2] contraction with errors of 28.8 and 7.9 Hz, respectively, as discussed in more detail above. The optimum contraction for the pcJ-2 basis set for the F_2 molecule is to six s-functions, but tests for other systems (Sect. 3.5) showed that a contraction to seven functions is required, where a segmented (5,2,1,…) and a general (7,2,1,…) [1,4] contraction have errors of 7.6 and 3.0 Hz, respectively. A contraction to seven s-functions for the pcJ-3 basis set gives contraction errors of ~1 Hz, but tests for other systems (Sect. 3.5) showed that a contraction to nine functions is required to sufficiently reduce the contraction errors. Both 9s-, 10s- and 11s- contractions for the pcJ-4 basis set can be considered based on contraction errors, and we recommend the 11s-contraction since the pcJ-4 is designed to produce high accuracy results and the computational savings by stronger contractions will be marginal.

Table 3 shows only the contractions with the lowest errors at a given contraction level and type, but there are in many cases several other contractions with almost the same error, and the exact combination that produces the lowest error will depend on the specific system. By analyzing the results for all the general 2-contractions, it is found that a close to optimum contraction can be located by starting from a fully contracted basis set ($N^c_{\mathrm{prim}}, N^c_{\mathrm{prim}}, 1, 1, \ldots$), and uncontracting the first function from the outermost primitives in combination with uncontracting the second function from the innermost primitives, until the

Table 3 Contraction errors (sum of absolute errors for the four contributions in Hz) for 1- and 2-contractions of s-functions of the pcJ-*n* ($n = 0$, 1, 2, 3, 4) basis sets for the F_2 molecule

Basis set ($N_{primitive}$) [Inherent error] {Inherent absolute error}	Contraction	1-Contracted	2-Contracted		3-Contracted
		Segmented (*n*,1,1,1,…)	Segmented (*n*,*m*,1,1,…)	General (*n*,*m*,1,1,…) [n_{start},m_{start}]	Segmented (*n*,*m*,*k*,1,…)
pcJ-0 (7s) [185] {1,647}	3s	984 (5,1,1)	531 (4,2,1)	186 (6,3,1) [1,4]	
	4s	**118 (4,1,1,1)**	169 (3,2,1,1)	66 (5,3,1,1) [1,2]	
	5s	49.4 (3,1,1,1,1)	–	21.9 (3,3,1,1,1) [1,2]	
pcJ-1 (9s) [87] {387}	4s	351 (6,1,1,1)	215 (5,2,1,1)	173 (7,2,1,1) [1,6]	2388 (2,4,2,1)
	5s	33.7 (5,1,1,1,1)	**28.8 (4,2,1,1,1)**	7.9 (6,4,1,1,1) [1,2]	284 (2,3,2,1,1)
	6s	23.2 (4,1,1,1,1,1)	5.8 (3,2,1,1,1,1)	1.7 (5,2,1,1,1,1) [1,4]	26.7 (2,2,2,1,1,1)
pcJ-2 (12s) [4.7] {28.1}	5s	320 (8,1,1,1,1)	135 (7,2,1,1,1)	60.1 (9,5,1,1,1) [1,3]	195 (6,2,2,1,1)
	6s	32.4 (7,1,1,…)	10.4 (5,3,1,…)	9.2 (8,7,1,…) [1,1]	47.8 (3,4,2,…)
	7s	36.4 (6,1,1,…)	**7.6 (5,2,1,**…)	3.0 (7,2,1,1…) [1,4]	14.1 (3,3,2,…)
pcJ-3 (16s) [21.3] {21.3}	6s	150 (11,1,1,…)	27.0 (9,3,1,…)	10.8 (12,5,1,…) [1,6]	
	7s	50.8 (10,1,1,…)	1.5 (9,2,1,…)	1.0 (11,2,1,…) [1,10]	15.2 (6,4,2,…)
	8s	8.5 (9,1,1,…)	2.6 (6,4,1,…)	0.36 (10,3,1,…) [1,5]	1.1 (5,4,2,…)
	9s	4.5 (8,1,1,…)	**0.89 (6,3,1,**…)	0.76 (6,8,1,…) [1,2]	0.37 (5,3,2,…)
pcJ-4 (19s)	8s	35.8 (12,1,1,…)	4.8 (11,2,1,…)	4.1 (13,6,1,…) [1,5]	14.0 (6,5,3,…)
	9s	24.1 (11,1,1,…)	1.8 (10,2,1,…)	0.33 (12,10,1,…) [1,1]	0.76 (7,4,2,…)
	10s	16.3 (10,1,1,…)	1.1 (7,4,1,…)	0.28 (11,7,1,…) [1,4]	0.21 (5,5,2,…)
	11s	12.7 (9,1,1,…)	**0.57 (6,4,1,**…)	0.39 (7,8,1,…) [1,3]	0.83 (5,4,2,…)

Only the combinations with lowest errors are shown. The recommended contractions are given in bold

The inherent error in brackets is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is the sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result

contraction error reached a minimum. This procedure significantly reduces the number of possible contraction schemes to be considered.

Considering only the segmented contractions in Table 3, the optimum contraction moves from 1-contraction for the smallest basis set towards 3-contraction for the largest basis set. For the pcJ-0 basis set, the (4,1,1,1) contraction is better than (3,2,1,1), for the pcJ-1 basis set the (4,2,1,…) contraction is slightly better than (5,1,1,…), while for the pcJ-2 basis set the (5,2,1,1,…) contraction is significantly better than (6,1,1,1,…) and (3,3,2,1,…), the latter being the best 3-contraction. For the pcJ-3 basis set, the 2-contraction (9,2,1,1,…) is better than the 3-contraction (6,4,2,1,…) when contracted to seven s-functions, while the (5,4,2,1,…) 3-contraction is marginally better than the (6,4,1,1,…) 2-contraction for a contraction to eight s-functions. For the contraction of the pcJ-4 basis set to nine s-functions, the 3-contraction (7,4,2,1,…) is marginally better than the 2-contraction (10,2,1,1,…). For the contraction to 10s-functions, a (5,5,2,1,…) 3-contraction is somewhat better than a (7,4,1,1,…) 2-contraction, but the former contains a fortuitous error cancellation, as the corresponding (5,4,2,1,…) contraction to 11s-functions has a larger error.

When comparing the contraction errors for the best segmented and best general contraction in Table 3, the latter always has the lowest error, but the differences are small. For the pcJ-2 basis set contracted to seven s-functions, for example, the contraction error for the best segmented contraction is 7.6 Hz, compared to 3.0 Hz for the best general contraction. This difference is insignificant compared to the (absolute) inherent error of 28.1 Hz. For the smaller pcJ-1 and pcJ-0 basis sets, the difference between the optimum segmented and general contraction is larger, but still well below the inherent error. There does not appear to be any systematic trend in the composition of the best general contracted function, and the slightly smaller contraction errors are therefore likely due to fortuitous error cancellation. Opting in favor of simplicity, we thus recommend the segmented contractions indicated in bold in Table 3.

Table 4 shows the contraction errors for the combined contraction of the s- and p-functions in Tables 1 and 3, and compares the errors to those from the previously recommended contractions where only 1-contractions were considered [14]. The improvements in terms of contraction errors and number of contracted functions are small.

Table 4 Contraction errors for the total J-value, with the sum of the absolute errors for each of the four contributions given in curly brackets (Hz) for recommended contractions of the pcJ-*n* ($n = 0, 1, 2, 3, 4$) basis sets for the F_2 molecule

Basis set ($N_{\text{primitive}}$) [Inherent error] {Inherent absolute error}	Primitive	Previous contraction [14]	Contraction error	New contraction	Contraction error
pcJ-0 [185] {1,647}	7s4p	4s3p (4,1/2,1)	123 {123}	4s3p (4,1/2,1)	123 {123}
pcJ-1 [87] {387}	9s5p	5s4p (5,1/2,1)	34.2 {36.7}	5s4p (4,2/2,1)	16.5 {28.4}
pcJ-2 [4.7] {28.1}	12s7p	7s5p (6,1/3,1)	38.1 {42.3}	7s5p (5,2/3,1)	8.0 {13.0}
pcJ-3 [21.3] {21.3}	16s10p	10s8p (7,1/3,1)	4.9 {5.2}	9s7p (6,3/3,2)	1.0 {1.5}
pcJ-4	19s12p	15s10p (5,1/3,1)	3.1 {3.2}	11s9p (6,4/3,2)	0.48 {0.69}

The inherent error in brackets is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result

3.3 The P_2 molecule, contraction of the p-functions

The contraction strategy for second row systems like P_2 follows the guidelines discussed for F_2 in the previous section. The contraction of the p-functions using 2p and 3p SCF coefficients is analogous to the contraction of s-functions for F_2, with the results shown in Table 5. For the pcJ-1 the inherent error in the total J-value is relatively low, although the sum of absolute errors shows that this is partly due to error cancellations, and 5p is the highest contraction level which produces acceptable contraction errors. For the pcJ-0 and pcJ-2 basis sets, the contraction errors suggest that the highest contraction level is 3p and 5p, respectively. We feel, however, that there should be a smooth progression in the overall quality of the pcJ-*n* basis sets as a function of *n*, and therefore suggest contractions to 4p and 6p, respectively, instead. For the pcJ-3 basis set, a contraction to either a 6p or 7p produces acceptable errors for the P_2 system, and based on the progression argument, we have chosen the 7p contraction. The pcJ-4 basis set is designed to provide high accuracy, and 9p is the maximum contraction that has acceptable errors. At the chosen contraction levels, there is little difference in performance between a segmented and general contraction and we recommend the segmented contractions given in bold.

3.4 The P_2 molecule, contraction of the s-functions

A contraction of the s-functions for P_2 can employ information from the 1s, 2s and 3s SCF coefficients, and

Table 5 Contraction errors (sum of absolute errors for the four contributions in Hz) for 1- and 2-contractions of p-functions of the pcJ-*n* ($n = 0$, 1, 2, 3, 4) basis sets for the P_2 molecule

Basis set ($N_{\text{primitive}}$) [Inherent error] {Inherent absolute error}	Contraction	1-Contracted	2-Contracted		3-Contracted
		Segmented (*n*,1,1,1,…)	Segmented (*n*,*m*,1,1,…)	General (*n*,*m*,1,1,…) [n_{start},m_{start}]	Segmented (*n*,*m*,*k*,1,…)
pcJ-0 (7p) [739] {739}	3p	17.9 (5,1,1)	245 (4,2,1)	22.2 (6,6,1) [1,1]	121 (2,3,2)
	4p	38.7 (4,1,1,1)	**15.9 (3,2,1,1)**	3.4 (4,5,1,1) [1,1]	9,998 (2,2,2,1)
pcJ-1 (9p) [7.2] {81.9}	4p	60.6 (6,1,1,1)	36.3 (5,2,1,1)	15.3 (7,7,1,1) [1,1]	69.6 (4,2,2,1)
	5p	13.2 (5,1,1,1,1)	**4.5 (4,2,1,1,1)**	3.2 (5,6,1,1,1) [1,1]	39.6 (3,2,2,1,1)
pcJ-2 (11p) [17.8] {56.5}	4p	44.2 (8,1,1,1)	32.2 (7,2,1,1)	34.1 (9,9,1,1,1) [1,1]	
	5p	45.8 (7,1,1,1,1)	6.8 (5,3,1,1,1)	1.9 (6,8,1,1,1) [1,1]	51.1 (5,2,2,1,1)
	6p	16.6 (6,1,1,1,…)	**2.3 (4,3,1,1,…)**	0.69 (5,7,1,1,…) [1,1]	2.1 (2,4,2,1,…)
pcJ-3 (14p) [9.7] {9.7}	5p	12.9 (10,1,1,1,…)	65.2 (9,2,1,1,…)	21.5 (11,11,1,1,…) [1,1]	
	6p	38.4 (9,1,1,1,…)	3.6 (7,3,1,1,…)	1.9 (7,10,1,1,…) [1,1]	16.6 (7,2,2,1,…)
	7p	21.9 (8,1,1,1,…)	**1.5 (6,3,1,1,…)**	0.68 (9,8,1,1,…) [1,2]	2.0 (5,3,2,1,…)
pcJ-4 (17p)	7p	38.1 (11,1,1,1,…)	4.8 (8,4,1,1,…)	0.80 (10,9,1,1,…) [1,4]	19.2 (9,2,2,1,…)
	8p	23.7 (10,1,1,1,…)	1.2 (7,4,1,1,…)	0.18 (11,9,1,1,…) [1,3]	1.5 (7,3,2,1,…)
	9p	8.6 (9,1,1,1,…)	**0.63 (6,4,1,1,…)**	0.08 (10,8,1,1,…) [1,3]	0.34 (5,3,3,1,…)

Only the combinations with lowest errors are shown. The recommended contractions are given in bold

The inherent error in brackets is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result (general) 3-contractions may now offer a real improvement over (general) 2-contractions. A complete search over all possible general 3-contractions is a significant effort, as even the smaller pcJ basis sets contain a fair number of primitive functions. We have limited the search over 2-contractions to combinations employing both 1-s and 2-s coefficients, as this was found to provide the lowest contraction errors for F_2, and over 3-contractions to combinations where each of the three functions are contracted using different sets of coefficients (1s,2s,3s). Within these limitations, we have furthermore only searched 3-contractions where the first function was sequentially uncontracted from the outermost primitive functions, the third function was sequentially uncontracted from the innermost primitive functions, while both inward and outward uncontraction was used for the second function relative to a fully contracted basis set ($N^c_{\text{prim}}, N^c_{\text{prim}}, N^c_{\text{prim}}, 1, \ldots$). The inner uncontraction for the third function was required to be at least the same as for the second function. The pcJ-3 basis set contains 19 primitive s-functions, and even with the aforementioned restrictions on the contractions, there are ~9000 possible combinations for contracting to seven s-functions, and ~7000 combinations for contracting to eight s-functions. We have not attempted searching for the best general 3-contraction for the pcJ-4 basis set due to an even larger number of possible combinations, and limited the search to segmented 3-contractions.

The best 2- and 3-contractions for a given number of contracted functions and contraction type are shown in Table 6. 1-contractions were found to produce significantly larger contractions errors and are not reported. Considering the data in Table 6, we note that the error reduction upon going from a general 2-contraction to a 3-contraction is only significant for the most contracted forms (4s and 5s) of the pcJ-0 and pcJ-1 basis sets. As discussed earlier, the large number of possible 3-contractions offers rich opportunities for random error cancellations when applied to a single system, and analysis of contractions similar to the optimum combinations in Table 6 suggest these results are due to fortuitous error cancellations. There are therefore no indications that 3-contractions offer a real improvement over 2-contractions.

The pattern for the segmented contractions follows the results for F_2. For the smaller pcJ-0, pcJ-1 and pcJ-2 basis sets, a 2-contraction is better than a 3-contraction for the recommended contractions (five, six and seven s-functions, respectively). Note that the best segmented 3-contraction for the pcJ-2 basis set corresponds to a (2,3,6,1,…) partitioning, while a (5,4,2,1,…) combination has a somewhat higher error (26.2 Hz). Combinations where the outer contracted function(s) contain more primitive functions than the inner contracted function(s), are considered to be artifacts resulting from random error cancellations. For the pcJ-3 and pcJ-4 basis sets the 2- and 3-contractions have very similar errors.

Table 6 Contraction errors (sum of absolute errors for the four contributions in Hz) for 1- and 2-contractions of s-functions of the pcJ-*n* ($n = 0$, 1, 2, 3, 4) basis sets for the P_2 molecule

Basis set ($N_{\text{primitive}}$) [Inherent error] {Inherent absolute error}	Contraction	2-Contracted		3-Contracted	
		Segmented (n,m,1,1,…)	General (n,m,1,1,…) [n_{start},m_{start}]	Segmented (n,m,k,1,…)	General (n,m,k,1,…) [n_{start},m_{start},k_{start}]
pcJ-0 (10s) [739] {739}	4s	482 (5,3,1,1)	287 (8,8,1,1) [1,1]	36.8 (2,4,3,1)	3.5 (9,5,7,1) [1,3,3]
	5s	30.4 (4,3,1,1)	**2.2 (6,7,1,1) [1,1]**	50.4 (4,2,2,1)	1.0 (8,4,4,1) [1,5,5]
pcJ-1 (13s) [7.2] {81.9}	5s	153 (7,3,1,1,1)	76.1 (7,6,1,1,1) [1,5]	83.1 (7,2,2,1,1)	6.9 (9,7,4,1) [1,3,8]
	6s	13.0 (5,4,1,1,…)	**2.5 (8,9,1,1,**…) **[1,1]**	67.2 (6,2,2,1,…)	1.0 (10,6,4,1,…) [1,1,7]
pcJ-2 (15s) [17.8] {56.5}	6s	27.0 (5,6,1,1,…)	3.7 (11,8,1,1,…) [1,4]	26.8 (5,5,2,1,…)	1.8 (10,8,9,1,…) [1,4,4]
	7s	**4.7 (7,3,1,1,**…)	1.3 (8,8,1,1,…) [1,3]	6.7 (2,3,6,1,…)	0.22 (11,8,4,1,…) [1,4,8]
pcJ-3 (19s) [9.7] {9.7}	7s	11.1 (9,5,1,1,…)	1.3 (10,13,1,1,…) [1,2]	7.2 (9,3,3,1,…)	0.19 (13,10,2,1,…) [1,4,14]
	8s	5.9 (9,4,1,1,…)	3.8 (13,10,1,1,…) [1,4]	1.1[a] (5,5,4,1,…)	0.31 (14,9,3,1,…) [1,3,12]
	9s	**2.3 (8,4,1,1,**…)	0.67 (12,10,1,1,…) [1,3]	0.42 (8,3,2,1,…)	0.18 (13,2,2,1,…) [1,10,12]
pcJ-4 (22s)	8s	13.7 (11,5,1,1,…)	3.7 (12,15,1,1,…) [1,2]	7.8 (10,4,3,1,…)	
	9s	5.6 (10,5,1,1,…)	2.0 (15,12,1,1,…) [1,4]	3.0 (4,8,4,1,…)	
	10s	5.2 (10,4,1,1,…)	1.0 (4,14,1,1,…) [1,1]	0.47 (10,2,3,1,…)	
	11s	**0.33 (9,4,1,1,**…)	0.48 (7,12,1,1,…) [1,2]	0.31 (8,4,2,1,…)	

Only the combinations with lowest errors are shown. The recommended contractions are given in bold

The inherent error in brackets is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result

[a] 7.5 Hz for a (9,3,2,1…) contraction

Table 7 Contraction errors for the total J-value, with the sum of the absolute errors for each of the four contributions given in curly brackets (Hz) for recommended contractions of the pcJ-*n* ($n = 0, 1, 2, 3, 4$) basis sets for the P_2 molecule

Basis set ($N_{\text{primitive}}$) [Inherent error] {Inherent absolute error}	Primitive	Previous contraction [14]	Contraction error	New contraction	Contraction error
pcJ-0 [739] {739}	10s7p	5s4p (6,6,1/4,1)	4.8 {42.5}	5s4p (6,7,1/3,2)	16.7 {16.7}
pcJ-1 [7.2] {81.9}	13s9p	6s5p (8,8,1/5,1)	14.4 {28.8}	6s5p (8,9,1/4,2)	0.7 {2.0}
pcJ-2 [17.8] {56.5}	15s11p	8s6p (8,8,1/6,1)	13.9 {28.3}	7s6p (7,3,1/4,3)	1.8 {4.9}
pcJ-3 [9.7] {9.7}	19s14p	11s10p (9,9,1/5,1)	1.0 {1.0}	9s7p (8,4,1/6,3)	1.9 {3.4}
pcJ-4	22s17p	16s12p (7,7,1/6,1)	0.29 {0.49}	11s9p (9,4,1/6,4)	0.40 {0.64}

The inherent error in bracket is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result

At the contraction levels where the contraction error becomes acceptable relative to the inherent errors, there is little difference between the best segmented and general 2-contractions for the pcJ-2, -3 and -4 basis sets, and we recommend the segmented contractions indicated in bold in Table 6. A contraction to eight s-functions for the pcJ-3 basis set is suitable for the P_2 molecule, but tests for other systems (Sect. 3.5) indicated that this basis set should not be contracted to less than nine functions. For the smaller pcJ-1 and -0 basis sets, a general contraction offers a real advantage over a corresponding segmented one, and the former are thus recommended.

Table 7 shows the contraction errors for the combined contraction of the s- and p-functions in Tables 5 and 6, and compares the errors to those from the previously recommended contractions where only 1- and a subset of 2-contractions were considered [14]. The new contraction schemes offer some reduction in the contraction errors for the pcJ-1 and -2 basis sets for essentially the same number of functions, while a small reduction in the number of contracted functions is achieved for the pcJ-3 and -4 basis sets without significantly increasing the contraction errors.

3.5 Other molecules

The H_2 molecule was subjected to a similar analysis for redefining the contraction for hydrogen with the results shown in Table 8. Except for the most strongly contracted versions of the pcJ-2 and pcJ-3 basis sets, segmented contractions are found to have smaller contraction errors than general contracted.

Table 8 Contraction errors (sum of absolute errors for the four contributions in Hz) for 1- and 2-contractions of p-functions of the pcJ-*n* ($n = 0, 1, 2, 3, 4$) basis sets for the H_2 molecule

Basis set ($N_{\text{primitive}}$) [Inherent error] {Inherent absolute error}	Contraction	1-Contracted	2-Contracted	
		Segmented (*n*,1,1,…)	Segmented (*n*,*m*,1,…)	General (*n*,*m*,1,…) [n_{start},m_{start}]
pcJ-0 (5s) [16.1] {19.3}	**3s**	**2.4 (3,1,1)**	30.9 (2,2,1)	7.7 (2,3,1) [1,2]
	4s	0.63 (2,1,1,1)	–	29.5 (2,2,1,1) [1,2]
pcJ-1 (6s) [9.2] {10.1}	3s	15.0 (4,1,1)	12.5 (3,2,1)	13.2 (3,4,1) [1,2]
	4s	1.7 (3,1,1,1)	**1.4 (2,2,1,1)**	18.5 (2,3,1,1) [1,2]
pcJ-2 (8s) [2.5] {2.5}	4s	27.6 (5,1,1,…)	6.9 (4,2,1,…)	0.65 (4,5,1,…) [1,2]
	5s	1.8 (4,1,1,…)	**0.18 (3,2,1,…)**	1.5 (2,4,1,…) [1,2]
	6s	0.09 (3,1,1,…)	0.30 (2,2,1,…)	7.7 (2,3,1,…) [1,2]
pcJ-3 (11s) [0.54] {0.54}	5s	186 (7,1,1,…)	3.0 (5,3,1,…)	0.83 (3,6,1,…) [1,3]
	6s	12.8 (6,1,1,…)	**0.21 (4,3,1,…)**	2.4 (4,6,1,…) [1,2]
	7s	8.9 (5,1,1,…)	0.40 (4,2,1,…)	0.60 (3,5,1,…) [1,2]
pcJ-4 (12p)	7s	23.5 (6,1,1,…)	1.6 (4,3,1,…)	2.7 (4,6,1,…) [1,2]
	8s	2.2 (5,1,1,…)	**0.05 (3,3,1,…)**	0.78 (5,5,1,…) [1,2]
	9s	0.33 (4,1,1,…)	0.09 (3,2,1,…)	4.3 (2,4,1,…) [1,2]

Only the combinations with lowest errors are shown. The recommended contractions are given in bold

The inherent error in brackets is the error relative to the uncontracted pcJ-4 result for the total spin–spin coupling constant. The inherent error in curly brackets is sum of the absolute errors of the four components to the spin–spin coupling constant relative to the uncontracted pcJ-4 result

Table 9 Average contraction errors for the recommended contractions of the pcJ-*n* ($n = 0, 1, 2, 3, 4$) basis sets for 51 unique coupling constants in a test set of 21 molecules (see text for details)

Basis set	Previous contraction [14]		New contraction	
	Average error	Average %error	Average error	Average %error
pcJ-0	14.8	32.7	14.1	31.2
pcJ-1	6.2	78.4	2.3	28.8
pcJ-2	2.7	78.9	1.2	35.4
pcJ-3	0.40	54.4	0.21	28.3
pcJ-4	0.15		0.08	

Average error is the average unsigned contraction error in Hz. Average %-error is the sum of contraction errors in percent of the unsigned inherent error relative to the uncontracted pcJ-4 results

Based on the contraction analysis for the H_2, F_2 and P_2 systems, we have constructed analogous contractions for the remaining first and second row elements, and tested the performance for a larger selection of systems (H_2, C_2, N_2, F_2, Si_2, P_2, Cl_2, HF, CO, SiS, H_2O, HCN, Cl_2S, CH_4, C_2H_2, C_2H_4, C_2H_6, C_2F_2, H_2CO, H_2CS, H_2PF) for a total of 51 unique coupling constants. As mentioned in the previous sections, this larger test set showed that it in some cases was necessary to slightly relax the contraction compared to the optimum contractions derived from the F_2 and P_2 molecules, and these relaxed contractions are the recommended contractions in Tables 1, 3, 5 and 7.

Table 9 shows the average (unsigned) contraction error in Hz and a percentage error relative to the (unsigned) inherent error of the uncontracted pcJ-*n* basis set relative to the uncontracted pcJ-4 results. As spin–spin coupling constants span a range from essentially zero to several thousand Hz, the percentage error has been calculated as a sum of contraction errors divided by a sum of inherent errors, rather than as a sum of percentage errors, to avoid the percentage error being dominated by contributions from systems with small values of the spin–spin coupling constant.

Comparing the results in Table 9 with the contractions in Tables 4, 7 and 8, it is seen that the new contractions lead to slightly smaller errors for the pcJ-0, -1 and -2 basis sets with essentially the same number of contracted basis functions. For the pcJ-3 and -4 basis sets, the new contractions lead to both fewer basis functions and lower contractions error, but it should be recognized that the computational cost for these basis sets will be dominated by the large number of polarization functions.

4 Summary

We have performed a systematic search for the optimum contraction schemes for the pcJ-*n* basis sets designed for calculating spin–spin coupling constants using density functional theory. Only modest improvements have been made compared to the previously proposed contraction, and given the exhaustive search in the present work, this shows that it is difficult to contract these basis sets to any significant degree without losing the inherent accuracy. A similar conclusion has been reached for calculating hyperfine coupling constants [38].

The present work provides guidelines for searching for optimum contraction schemes at theoretical levels, like CCSD, where a systematic search is impractical. We suggest that the maximum number of general contracted functions is the number of independent sets of contraction coefficients. A complete search over all possible general contractions is a combinatorial problem that rapidly becomes unmanageably large, and is furthermore prone to fortuitous error cancellations. A close to optimum general contraction can be found by starting from a fully contracted basis set, and sequentially uncontracting the inner contracted function from the outermost primitive functions, combined with an uncontraction of the innermost primitive functions in the outer contracted function, until the contraction error reaches a minimum. For the present case of spin–spin coupling constants, there is relatively little difference in performance between a segmented and general contraction, and the difference diminished as the basis sets become larger. For other properties, the use of general contractions may offer a definite advantage, but this will have to be settled on a case-by-case study.

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

References

1. Lutnaes OB, Ruden TA, Helgaker T (2004) Mag Res Chem 42:S117
2. Helgaker T, Lutnaes OB, Jaszunski M (2007) J Chem Theor Comp 3:86
3. Gauss J, Stanton JF (1995) J Chem Phys 103:3561
4. Gauss J, Stanton JF (1996) J Chem Phys 104:2574
5. Stanton JF, Gauss J, Siehl H-S (1996) Chem Phys Lett 262:183
6. Hansen MB, Kongsted J, Toffoli D, Christiansen O (2008) J Phys Chem A 112:8436
7. Peralta JE, Scuseria GE, Cheeseman JR, Frisch MJ (2003) Chem Phys Lett 375:452
8. Manninen P, Vaara J (2006) J Comp Chem 27:434
9. Ruden TA, Helgaker T, Jaszunski M (2003) Chem Phys 296:53
10. Ruden TA, Lutnæs OB, Helgaker T, Ruud K (2003) J Chem Phys 118:9572
11. Helgaker T, Jaszunski M, Pecul M (2008) Prog Nucl Res Spect 53:249
12. van Mourik T, Dingley AJ (2007) ChemPhysChem 8:288
13. Deng W, Cheeseman JR, Frisch MJ (2006) J Chem Theory Comput 2:1028
14. Jensen F (2006) J Chem Theory Comput 2:1360

15. Benedikt U, Auer AA, Jensen F (2008) J Chem Phys 129:064111
16. Jensen F (2001) J Chem Phys 115:9113
17. Jensen F (2002) J Chem Phys 116:3502
18. Jensen F (2002) J Chem Phys 116:7372
19. Jensen F, Helgaker T (2004) J Chem Phys 121:3462
20. Dunning TH Jr (1989) J Chem Phys 90:1007
21. Davidson ER, Feller D (1986) Chem Rev 86:681
22. Helgaker T, Taylor PR (1995) In: Yarkony D (ed) Modern electronic structure theory, Part II. World Scientific, New Jersey, pp 727–856
23. Almlöf J, Taylor PR (1987) J Chem Phys 86:4070
24. Enevoldsen T, Oddershede J, Sauer SPA (1988) Theor Chem Acc 100:275
25. Provasi PF, Aucar GA, Sauer SPA (2003) Int J Mol Sci 4:231
26. Provasi PF, Aucar GA, Sauer SPA (2004) J Phys Chem A 108:5393
27. Helgaker T, Watson M, Handy NC (2000) J Chem Phys 113:9402
28. Helgaker T, Jensen HJA, Jørgensen P, Olsen J, Ruud K, Ågren H, Auer AA, Bak KL, Bakken V, Christiansen O, Coriani S, Dahle P, Dalskov EK, Enevoldsen T, Fernez 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, Schimmelpfenning B, Sylvester-Hvid KO, Taylor PR, Vahtras O (2005) DALTON, a molecular electronic structure program, Release 20
29. Becke AD (1993) J Chem Phys 98:5648
30. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) J Phys Chem 98:11623
31. Maximoff SN, Peralta JE, Barone V, Scuseria GE (2005) J Chem Theory Comput 1:541
32. Keal TW, Helgaker T, Salek P, Tozer DJ (2006) Chem Phys Lett 425:163
33. London F (1937) J Phys Radium Paris 8:397
34. McWeeny R (1962) Phys Rev 126:1028
35. Wolinski K, Hilton JF, Pulay P (1990) J Am Chem Soc 112:8251
36. Ditchfield R (1974) Mol Phys 27:789
37. Davidson ER (1996) Chem Phys Lett 260:514
38. Kong J, Boyd RJ (1997) J Chem Phys 107:6270