On the Use of an Interpolation Approach for the Choice of Gaussian Polarization Functions

Fernando Ratuchne,a,# Ana C. Mora,b,# Ricardo Celestea and Albérico B. F. da Silva*,b

aDepartamento de Química, Universidade Estadual do Centro-Oeste, Campus Cedeteg, 85040-080 Guarapuava-PR, Brazil
bDepartamento de Química e Física molecular, Instituto de Química de São Carlos, Universidade de São Paulo, C.P. 780, 13560-970 São Carlos-SP, Brazil

In this work, we tested a linear interpolation approach in order to select polarization functions (exponents) to be used with Gaussian basis sets. The Gaussian primitive functions were generated here for Ga to Kr and also for Sc to Cu. The general contraction method was used for the construction of contracted Gaussian basis sets of 6Z and 7Z quality. Polarization functions were added to the contracted bases by explicit optimization and also by interpolation of exponents. The performance of the contracted basis sets, augmented with polarization functions obtained by interpolation, was tested with molecular configurations interaction single and double excitations (CISD) and density functional theory (DFT) calculations for the systems Se, Se2, Se6, Ge2, CrH and FeH. The outcomes obtained in this work with interpolated polarization functions agreed very well with the ones augmented with polarization functions obtained by explicit optimization. The interpolation methodology presented here is useful to generate polarization functions for any Gaussian basis set in different series of atoms of the periodic table.

Keywords: Gaussian exponent optimization, interpolation of Gaussian exponents, Gaussian polarization functions

Introduction

In 1986, Mohallem et al.1 introduced in the scientific literature the so-called generator coordinate Hartree-Fock (GCHF) method. In the years to follow, the GCHF method was employed to generate a set of universal Slater and Gaussian basis sets.2-4 In that time, we were interested in the generation of universal basis sets, believing that they could be useful to reduce the number of integrals to be calculated,5,6 as we had an only set of exponents to be used for a large number of atoms.

In 2003, Barbosa and da Silva7 modified the way of solving the integral equations of the GCHF method with the aim to generate more flexible Gaussian basis sets to be used in atomic and molecular calculation. In 2015, the first set of primitive (extended) Gaussian basis sets, for the atoms from hydrogen (H) to barium (Ba), that is from Z = 1 to 56, was presented in the literature by employing the new way to discretize the integral equations of the GCHF method by using a polynomial.8

Basically, there are three steps to follow in order to get a basis set ready to be used in molecular calculation: (i) the construction of the primitive set of exponents; (ii) the contraction of this primitive set; (iii) the addition of polarization functions.

In the generation of basis sets, usually the primitive set is constructed for the free atom in its ground state and the description of the outer part of the electron density distribution is generally poor, so the addition of higher angular momentum functions (exponents) in the body of the primitive set of exponents improves the performance of a basis set in the molecular environment.

Generally, the polarization exponents are assigned by optimizing the total energy for a small set of molecules, for example at the Hartree-Fock (HF) and density functional (DFT) levels of theory. Thus, the values of the polarization exponents for the 6+31G* basis sets were determined by the HF energy optimization for a set of small molecules at their equilibrium configuration.9 On the other hand, the
pc-n basis sets from Jensen used a BLYP\(^{10}\) (Becke-Lee-Yang-Parr) energy optimization.\(^{11,12}\)

Calculations including electronic correlation are also used to determine polarization exponents by the energy minimization of isolated atoms.\(^{13}\)

In the case of the consistent correlation (cc-pVXZ) basis sets, the polarization exponents were determined by minimization of the atomic energy\(^{14-16}\) by using the configurations interaction single and double excitation (CISD) method, whereas the Pople and co-workers\(^{17}\) 6-311G(d) basis sets employed Moller-Plesset second order perturbation theory (MP2) to optimization at the atomic level.\(^{17,18}\)

In this work, the polarization exponents have been determined by using a methodology based on an interpolation of Gaussian basis set exponents, combined with the generator coordinate (GC) method, with the aim to avoid the explicit optimization of the polarization exponents for each atom within a same period of the periodic table. As one example, we present the results obtained with Gaussian basis sets for Ga to Kr and the 3d series of the transition metals (Sc-Cu).

This idea is not exactly new, that is, generating polarization Gaussian functions by interpolation\(^{19,20}\), but we did not know if it could be successful when using it with the GC method. Despite that, in this work we show that this idea is competitive with other methods (explicit optimization of Gaussian exponents) and it can be used with different set of Gaussian exponents.

**Methodology**

In order to generate our Gaussian basis sets for this work, we have employed the polynomial generator coordinate Hartree-Fock (PGCHF) method.\(^7\) The PGCHF method is the result of employing the GC ansatz\(^1\) in the independent particle model:

\[
\psi_n(1) = \int \phi_n(1, \alpha) f_n(\alpha) d\alpha, \quad k = 1, \ldots, n
\]  

where \(\phi_n\) is the generator function and it can be either Slater-type functions (STFs) or Gaussian-type functions (GTFs), \(f_n\) is the weight function, \(\alpha\) is the generator coordinate (exponents of the STFs or GTFs) and \(n\) is the number of particles. The application of the variational principle to the energy expectation value leads to the Griffing-Wheeler-Hartree-Fock (GWHF)\(^1\) equation:

\[
\int [F(\alpha, \beta) - \varepsilon_\alpha S(\alpha, \beta)] f_n(\beta) d\beta = 0, \quad k = 1, \ldots, n
\]  

where \(\varepsilon_\alpha\) are the orbital eigenvalues and \(F(\alpha, \beta)\) and \(S(\alpha, \beta)\) are the Fock and overlap kernels, respectively.\(^{21}\)

Initially, the GWHF equation was discretized by using the integral discretization technique through an equally spaced numerical mesh\(^{21}\) and used successfully in the generation of Slater- and Gaussian-type universal basis sets.\(^{22,23}\)

In the PGCHF method,\(^7\) the exponents (\(\alpha\)) of each atomic orbital symmetry \(w\) are determined using a polynomial expansion of \(q\) order:

\[
\ln A^{(w)} = \frac{\Omega_{\text{min}}^{(w)}}{A} + \frac{\Delta \Omega_{\text{min}}^{(w)}}{A} (k-1) + \frac{\Delta^2 \Omega_{\text{min}}^{(w)}}{A} (k-1)^2 + \ldots + \frac{\Delta^q \Omega_{\text{min}}^{(w)}}{A} (k-1)^q, \quad k = 1, \ldots, N
\]

where \(A\) is a scaling parameter determined numerically (\(A = 6.0\)), \(N\) is the number of discretization points and \(\Omega_{\text{min}}^{(w)}\) and \(\Delta \Omega_{\text{min}}^{(w)}\) are, respectively, the initial point of the mesh and the increment used to obtain the subsequent points of the mesh.

Alternatively to the original GCHF method,\(^{21}\) a version developed by Jorge and de Castro,\(^{24}\) named improved generator coordinate Hartree-Fock (IGCHF) method, in which the space \(\Omega\) is discretized for each orbital symmetry by using two or three independent arithmetic sequences, was used in the generation of Gaussian basis sets for the atoms of the periodic table.\(^{25-28}\)

In this work, we have used the PGCHF\(^7\) method to generate primitive Gaussian basis sets to the series from Ga to Kr and Sc to Cu. Afterwards, each generated Gaussian basis set was contracted through the general contraction scheme from Davidson.\(^{29}\) The best contraction was defined by considering the lowest loss in the atomic total HF energy when compared with the respective numerical Hartree-Fock (NHF) energy value.\(^{30}\) Having defined the quality of the Gaussian basis sets, different polarization functions were added to the contracted bases in order to produce polarized basis sets to be used in atomic and molecular calculations.

The choice of the polarization exponents was carried out in two forms: first, we have determined the polarization exponents by explicit optimization from atomic calculations at the CISD level of theory; second, we have obtained the polarization exponents by numerical interpolation from linearization (least squares fitting) of the exponents generated in the first form described above in function of the atomic number.

In order to check out the performance of the basis sets augmented with the polarization functions generated by exponent interpolation, we have carried out a series of exploratory calculations of the total energy: (i) at the CISD level for the atoms under study and (ii) employing the B3LYP\(^{31}\) (Becke-three parameters-Lee-Yang-Parr) functional for the molecular systems studied. The results
are compared with those obtained from Gaussian basis sets augmented with polarization exponents obtained by explicit optimization. Both atomic and molecular calculated energies were acquired with the Gaussian 09.32

Results and Discussions

In this section, we are going to present initially the results obtained with our interpolation methodology for the series of atoms from Ga to Kr, but this methodology is also valid for any series of atoms of the periodic table, as we are going to comment ahead in this work. Also, we are going to present molecular property results for molecules containing atoms of the 3d series of the transition metals using polynomial generator coordinate Gaussian basis sets augmented with polarization functions obtained by interpolation.

Having defined the PGC basis set composition in terms of primitive functions consisting of $2s^2 1p^6 1d^6$ for Ga to Kr and with the aim to reach the accuracy of mHartree for the total energy, when compared to NHF results,29 we employed the general contraction scheme for each primitive set in order to obtain a contracted basis set of 6Z quality in the valence region. Finally, we determined the polarization functions corresponding to the symmetries f, g and h. Three sets of polarization exponents 2f1g, 3f1g and 3f2g1h were obtained by minimizing the total CISD energy for the atoms from Ga to Kr.

In fact, we observed that the values of the polarization exponents increase linearly (within a “period”) with the increasing of the atomic number. The tendency of the polarization exponents corresponding to the polarization sets 2f1g, 3f1g and 3f2g1h is presented in Figure 1 as a function of the atomic number. Curves of linearization of the same exponents with their respective $R^2$ (coefficient of determination) values are also shown in Figure 1. The range of $R^2$ varied from 0.9897 (1h exponent) to 0.9258 (2f exponent) for the polarization set 3f2g1h. The results for the linearization for other polarization functions are within this range including the polarization set 2f1g.

The curves presented in Figure 1 also show a deviation for Se ($Z = 34$) and can be observed that such deviation is common for all orbital symmetries of Se used as polarization functions. Although the curves obtained from the optimized exponents for the atomic series 4p of the p block (Figure 1) do not have a perfect linear behavior, it is possible, therefore, to approximate these curves to a straight line through the generation of exponents, for example, to only two atoms of the series (in this case Ga and Kr) and afterwards to obtain the polarization exponents for the remaining atoms of the series by interpolation.

The exponent interpolation strategy used here avoids the explicit (whole) optimization of polarization exponents (in Gaussian basis sets) for each atom within one period. Thus, we have made a linear adjustment of the explicitly optimized exponents 2f1g and 3f2g1h as a function of the atomic number to Ga and Kr, and then we recalculated the exponents from the linear equation obtained. The interpolated exponents were calculated by using the QtiPlot33 software with the precision of $10^{-4}$ provided by the software.

The values of the exponents both optimized and interpolated for Se are presented in Table 1. We have chosen Se since it presents the greatest deviation (Figure 1) from linearization. Nonrelativistic total energies calculated with

Figure 1. Optimized and interpolated polarization exponents for the PGC-6Z basis sets: 2f1g, 3f1g and 3f2g1h for the Ga-Kr series.
the PGC-6Z-2f1g and PGC-6Z-3f2g1h Gaussian basis sets are shown in Table 2 for the ground state of molecular systems containing the Se atom.

Table 1. Comparison between optimized and interpolated polarization exponents for the polarization sets 2f1g and 3f2g1h for the Se atom by employing the PGC-6Z basis set

<table>
<thead>
<tr>
<th>Symmetry</th>
<th>2f1g</th>
<th>2f1g</th>
<th>3f2g1h</th>
<th>3f2g1h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1f</td>
<td>0.70243581</td>
<td>0.7332</td>
<td>0.86523399</td>
<td>0.8705</td>
</tr>
<tr>
<td>2f</td>
<td>0.28185646</td>
<td>0.2990</td>
<td>0.47581691</td>
<td>0.5062</td>
</tr>
<tr>
<td>Se</td>
<td>0.575403655</td>
<td>0.5907</td>
<td>0.88804667</td>
<td>0.9090</td>
</tr>
<tr>
<td>1g</td>
<td>0.575403655</td>
<td>0.5907</td>
<td>0.88804667</td>
<td>0.9090</td>
</tr>
<tr>
<td>2g</td>
<td>0.39750892</td>
<td>0.4003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1h</td>
<td>0.70745011</td>
<td>0.7223</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Optimized exponents; †linearized exponents.

When comparing the results for the linearized exponents with the explicit optimized exponents, we can see that the exponent differences are in a range from 0.030764 (1f) for 2f1g to 0.002791(2g) for 3f2g1h.

Initially, we thought this difference was quite significant; however, the difference between the energies with explicit and interpolated exponents listed in Table 2 showed that the polarization exponents do not need to be numerically very similar, since the errors obtained are out of the required precision in atomic and molecular calculations.

We have also performed calculations for molecular properties (harmonic vibrational frequencies and equilibrium distance) at the B3LYP level of theory using the PGC-6Z basis sets with polarization functions 3f2g1h (optimized and linearized). The results are present in Table 3 and in both cases the B3LYP/PGC-6Z-3f2g1h (optimized and linearized exponents) Se2 harmonic vibrational frequencies are less than 2% of the experiment value 385 cm⁻¹ reported by Huber and Herzberg. 34 The harmonic vibrational frequency of Ge₂ compares very well with the experimental value reported by Li et al. 35 (308 cm⁻¹).

Also from Table 3, we can see that the CPU (central process unit) time is lower for some cases when we used the basis set with linearized polarization.

Similarly, in Table 3, we compare the calculated vibrational modes of Se₆ with the experimental spectra obtained for the rhombohedral structure: 103, 151 and 253 cm⁻¹, with symmetry assignment E_u, A₂u and E_u, respectively. 36

Table 2. CISD atomic energy values for the Se atom and B3LYP molecular energies for molecular systems containing the Se atom. Calculations were performed with the primitive PGC-6Z basis set plus the polarization functions for both interpolated and optimized 2f1g and 3f2g1h sets of polarization functions

<table>
<thead>
<tr>
<th>Polarization set</th>
<th>Se</th>
<th>Se₂</th>
<th>Se₆</th>
</tr>
</thead>
<tbody>
<tr>
<td>2f1g</td>
<td>-2399.990878</td>
<td>-4809.296058</td>
<td>-14409.978527</td>
</tr>
<tr>
<td>2f1g</td>
<td>-2399.990863</td>
<td>-4809.296052</td>
<td>-14409.978325</td>
</tr>
<tr>
<td>RE / %</td>
<td>6.38 × 10⁻⁷</td>
<td>1.39 × 10⁻⁷</td>
<td>1.40 × 10⁻⁷</td>
</tr>
<tr>
<td>3f2g1h</td>
<td>-2399.993435</td>
<td>-4809.29689</td>
<td>-14409.989</td>
</tr>
<tr>
<td>3f2g1h</td>
<td>-2399.993431</td>
<td>-4809.296079</td>
<td>-14409.989</td>
</tr>
<tr>
<td>RE / %</td>
<td>1.54 × 10⁻⁷</td>
<td>2.08 × 10⁻⁷</td>
<td>1.31 × 10⁻⁷</td>
</tr>
</tbody>
</table>

*Optimized exponents; †linearized exponents, RE: relative error.

Table 3. B3LYP/PGC-6Z-3f2g1h harmonic vibrational frequencies (ωₑ), equilibrium distances (Rₑ) for the lowest-lying states of Se₂, Se₆ and Ge₂. Results are shown for both optimized (Opt) and interpolated (Inter) sets of polarization functions

<table>
<thead>
<tr>
<th></th>
<th>Rₑ / Å</th>
<th>ωₑ / cm⁻¹</th>
<th>CPU time / (h:min:s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Se₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opt</td>
<td>2.19</td>
<td>318.168</td>
<td>3:31:54</td>
</tr>
<tr>
<td>Inter</td>
<td>2.19</td>
<td>379.672</td>
<td>2:55:59</td>
</tr>
<tr>
<td>Exp²</td>
<td>2.16</td>
<td>385</td>
<td></td>
</tr>
<tr>
<td>Ge₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opt</td>
<td>2.29</td>
<td>306.209</td>
<td>4:25:54</td>
</tr>
<tr>
<td>Inter</td>
<td>306.200</td>
<td>306.200</td>
<td>4:28:19</td>
</tr>
<tr>
<td>Exp²</td>
<td>2.447</td>
<td>308</td>
<td></td>
</tr>
<tr>
<td>Se₆</td>
<td>2.37</td>
<td>79.513(E_u), 151.914(A₂u), 257.912(E_u)</td>
<td>9:11:50</td>
</tr>
<tr>
<td>Opt</td>
<td>2.37</td>
<td>79.472(E_u), 151.914(A₂u), 257.912(E_u)</td>
<td>9:01:21</td>
</tr>
<tr>
<td>Inter</td>
<td>2.37</td>
<td>79.472(E_u), 151.914(A₂u), 257.912(E_u)</td>
<td>9:01:21</td>
</tr>
<tr>
<td>Exp²</td>
<td>2.30</td>
<td>103(E_u), 151(A₂u), 253(E_u)</td>
<td>9:11:50</td>
</tr>
</tbody>
</table>

*Reference 34; †reference 35; ‡reference 36; Se and Ge ground state for Se₂ and Ge₂, respectively; CPU time (h:min:s): process time in hour:minutes:seconds; E_u, A₂u and E_u; symmetry assignment to the vibrational modes to Se₆.
As additional calculation, we present results for the atoms of the 3d series of the transition metals using PGC-7Z basis sets. In this case, the optimized polarization exponents were generated by explicit optimization at the CISD level of theory from Sc to Cu. For this series, both the tendency of the polarization exponents corresponding to the polarization sets 2f, 3f, and 3f1g, as a function of atomic number, and the curves of linearization with their respective $R^2$ values are shown in Figure 2.

The range of $R^2$ varied from 0.9935 (1f exponent) to 0.9414 (2g exponent) for the polarization set 3f1g. The results for the linearization for other polarization functions are within this range including the polarization set 2f1g. From Figure 2, it is observed that the polarization exponents for Cr are more deviated from the linearity in all cases.

For the linearization process, we have used optimized polarization exponents obtained for Sc and Cu. The linearization equation attained has been used to recalculate the polarization exponents. The comparison between optimized and interpolated polarization exponents for the polarization sets 2f1g and 3f2g1h for the Cr atom is presented in Table 4.

The molecular property calculations were performed using the PGC-7Z basis sets with 3f2g1h polarization functions. Our results are compared with those obtained from basis sets augmented with polarization exponents obtained by explicit optimization. Results for CrH and FeH are presented in Table 5.

### Table 4. Comparison between optimized and interpolated polarization exponents for the polarization sets 2f1g and 3f2g1h for the Cr atom by employing the PGC-7Z basis set

<table>
<thead>
<tr>
<th>Symmetry</th>
<th>2f1g$^a$</th>
<th>2f1g$^b$</th>
<th>3f2g1h$^a$</th>
<th>3f2g1h$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1f</td>
<td>2.762645</td>
<td>2.6968</td>
<td>3.914512</td>
<td>3.7858</td>
</tr>
<tr>
<td>2f</td>
<td>0.850861</td>
<td>0.8309</td>
<td>1.403771</td>
<td>1.3960</td>
</tr>
<tr>
<td>3f</td>
<td>–</td>
<td>–</td>
<td>0.346343</td>
<td>0.3152</td>
</tr>
<tr>
<td>1g</td>
<td>1.709404</td>
<td>1.6759</td>
<td>2.880982</td>
<td>2.8756</td>
</tr>
<tr>
<td>2g</td>
<td>–</td>
<td>–</td>
<td>0.832711</td>
<td>0.8269</td>
</tr>
<tr>
<td>1h</td>
<td>–</td>
<td>–</td>
<td>1.74975</td>
<td>1.7363</td>
</tr>
</tbody>
</table>

$^a$Optimized exponents; $^b$ linearized exponents.

### Table 5. B3LYP/PGC-7Z-3f2g1h harmonic vibrational frequencies ($\omega_e$), equilibrium distances ($R_e$) for the lowest-lying states of CrH and FeH. Results are shown for both optimized (Opt) and interpolated (Inter) sets of polarization functions

<table>
<thead>
<tr>
<th></th>
<th>R$_e$ / Å</th>
<th>$\omega_e$ / cm$^{-1}$</th>
<th>CPU time / (h:min:s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Opt</td>
<td>Inter</td>
<td>Opt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CrH ($^6\Sigma$)</td>
<td>1.667</td>
<td>1.665</td>
<td>1637</td>
</tr>
<tr>
<td>Exp$^a$</td>
<td>1.662</td>
<td>1587</td>
<td></td>
</tr>
<tr>
<td>FeH ($^4\Delta$)</td>
<td>1.563</td>
<td>1.561</td>
<td>1829</td>
</tr>
<tr>
<td>Exp$^a$</td>
<td>1.589</td>
<td>1827</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Reference 37. $^6\Sigma$ and $^4\Delta$: lowest-lying states for CrH and FeH, respectively; CPU time (h:min:s): process time in hour:minutes:seconds.
Comparing the results for the linearized exponents with the explicit optimized exponents, we can see that the exponent differences are in a range from 0.06584 (1f) for 2f1g to 0.00538 (1g) for 3f2g1h.

The results presented in this work showed that suitable polarization exponents can be obtained by interpolation of any Gaussian exponents, taking advantage of the variation of the Gaussian exponents with the atomic number. Here we would like to point out that such tendency is always observed for a sequence of atoms in a row with the same structure valence, for example the series 2p B (Z = 5) to Ne (Z = 10); 3p Al (Z = 13) to Ar (Z = 18); 4p Ga (Z = 31) to Kr (Z = 36), etc., and for the series 3d Sc (Z = 21) to Zn (Z = 30); 4d Y (Z = 39) to Cd (Z = 48), etc., for the case of the transition metals.

Thus, our methodology becomes important for series with a larger number of atoms, for example the p, d and f blocks of atoms of periodic table and unnecessary for the atoms of the s block of the periodic table, since these have only two atoms (series 2s Li (Z = 3) and Be (Z = 4); 3s Na (Z = 11) and Mg (Z = 12), etc.).

Conclusions

In this work, we have determined polarization exponents by linearization using the periodic dependence of the polarization exponents with the atomic number for sequences of atoms within the same period of the periodic table. The PGC-6Z and PGC-7Z Gaussian basis sets for the atoms from Ga through Kr and Sc through Cu were generated employing the polynomial generator coordinate Hartree-Fock (PGCHF) method. The PGC-6Z basis sets were augmented with the polarization function sets 2f1g and 3f2g1h obtained both by explicit optimization and interpolation. The PGC-7Z basis sets were augmented with polarization functions set 3f2g1h, also using explicit optimization and interpolation. The atomic and molecular calculations performed here (with the CISD and B3LYP methods) show that the calculated total energy obtained with the PGC-6Z basis sets (augmented with interpolated polarization functions) are as accurate as the ones obtained with explicit optimized polarization functions (with a relative error of 10⁻⁷ Hartree).

The results presented in this work, employing an interpolation methodology, show that this strategy is a good alternative for the generation of polarized Gaussian basis sets without compromising the quality of the basis sets in atomic and molecular calculations when using any Gaussian basis set.

Acknowledgments

The authors would like to thank CNPq and CAPES (Brazilian Agencies) for the financial support.

References

33. Vasilef, I.; QtiPlot, Data Analysis and Scientific Visualization; Universiteit Utrecht, Utrecht, NLD, 2013.

Submitted: March 21, 2018
Published online: August 30, 2018