CrossMark

REGULAR ARTICLE

Fast calculation of two-electron-repulsion integrals: a numerical approach

Pedro E. M. Lopes[1]

Abstract An alternative methodology to evaluate two-electron-repulsion integrals based on numerical approximation is proposed. Computational chemistry has branched into two major fields with methodologies based on quantum mechanics and classical force fields. However, there are significant shadowy areas not covered by any of the available methods. Many relevant systems are often too big for traditional quantum chemical methods while being chemically too complex for classical force fields. Examples include systems in nanomedicine, studies of metalloproteins, etc. There is an urgent need to develop fast quantum chemical methods able to study large and complex systems. This work is a proof-of-concept on the numerical techniques required to develop accurate and computationally efficient algorithms to compute electron-repulsion integrals, one of the most significant bottlenecks in computational quantum chemistry. All concepts and calculations were performed for the three-center integral $(p_{xA}p_{xB}|p_{xC}\text{-}p_{xC})$ with all atoms being carbon. Starting with the explicit analytical formulas, convenient decompositions were tested to provide smooth 2-dimensional surfaces that were easily fitted. The approximating algorithm consisted of a multilayered approach based on multiple fittings of 2-dimensional surfaces. An important aspect of the new method is its independence on the number of contracted Gaussian primitives. The basis set of choice was STO-6G. In future development of this work, larger basis set will be developed. This work is part of a large effort aimed at bringing simplified quantum mechanical methods to systems where accuracy can be sacrificed for speed. An initial application will be development of quantum mechanical techniques for molecular recognition.

1 Introduction

The field of computational quantum chemistry has experienced extraordinary progress to date due to advances in computing power and the development of new algorithms. While advances have been reached, still there are limitations in the size and/or complexity of the systems that can be studied. In the second decade of the twenty-first century, the words of Paul Dirac in 1929 [1] still echo: *the underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.* Today, Dirac's statement still remains true and many of the equations governing the chemical phenomena are still too complex to solve using today's computational resources.

To answer complex chemical phenomena, computational quantum chemistry has suffered multiple numerical approximations and simplifications. Some are

✉ Pedro E. M. Lopes
  plopesuk@yahoo.co.uk

1 Rua Almirante Reis, N° 28A, 2° Esq,
   2330-099 Entroncamento, Portugal

⚫ Springer

approximations to the fundamental equations such as the Born–Oppenheimer approximation that greatly simplifies the Schrödinger equation by considering that the much heavier nuclei remain stationary during the calculation. Other approximations, for example, empirical and semiempirical methods, consider simplified forms of the first-principles underlying equations that are typically faster to solve. Other classes of methods in computational chemistry have abandoned the quantum chemical principles altogether, and classical approximations to the potential energy surfaces based on empirical force fields were developed, as they are computationally less intensive than quantum chemical electronic structure calculations. Empirical force fields are currently the methods of choice for studies of large systems in biology and materials science, for example conformational studies of proteins or DNA and protein–ligand binding thermodynamics. However, empirical force fields have severe limitations: limited applicability, difficulty in describing complex chemistries, and inability to describe systems where formation and breakage of bonds occur. Empirical force fields are highly parameterized and typically include terms for bonds, angles, and torsions plus non-bonding terms [2]. Empirical force fields are limited to the systems used in developing the parameters (e.g., proteins, lipids, DNA/RNA, etc.) and the parameterizations usually cover sp, sp^2, and sp^3 hybridizations. It is extremely difficult to cover complex chemical spaces with force fields, for example when transition metals are involved. The harmonic nature of force fields does not typically allow for breaking and formation of chemical bonds. In contrast, quantum chemical methods can describe most systems, but are still limited to small models, at least when compared to typical systems studied by classical force fields. QM/MM mix quantum chemical methods with empirical force fields and, thus, are able to study large adequate to describe large systems. QM/MM methods work better when the quantum region is highly localized but are, in general, inadequate to describe the dynamics of large systems.

Currently, many systems in biophysics, biochemistry, materials science, nanomedicine, etc., cannot be described using existing methodologies. These systems have huge chemical spaces that are impossible to cover using existing empirical force field methods and are too big for current quantum mechanics (QM) techniques, even the best linear scaling methods. There is a clear "capabilities gap" in existing computational methodologies that need to be urgently addressed. Not only emerging fields such as nanomedicine or materials science would benefit from new computational methodologies based on QM. Traditional applications of classical force field methods would benefit as well. For example, it is estimated that half of all proteins are metalloproteins [3]. Simulations of metalloproteins

would greatly benefit from fast QM methods since existing classical force fields have problems describing such systems.

In summary, new computational methodologies are needed to bridge the "capabilities gap" between current quantum chemical methods and classical force fields. In the base Hartree–Fock method, the major contributors to the cost of the calculation are the computation of the two-electron-repulsion integrals (ERIs), with a formally quartically scaling $O(N^4)$, although for large systems the scaling asymptotically approaches $O(N^2)$, diagonalization of the Fock matrix with a cubically scaling $O(N^3)$, and the self-consistent field (SCF) procedure that typically adds more than ten iterations for small systems. Development of new computational methodologies based on QM will have to address each of the restrictions in order to achieve acceptable speeds. The aim of the current work is to develop an alternative technique, based on accurate numerical approximations, for the fast computation of ERIs.

It was already apparent in the 1950s that calculations of polyatomic systems based on Slater-type orbitals would be intractable. The breakthrough occurred when Boys proposed basis functions based on Cartesian Gaussian functions [4]. Importantly, it was found that linear combinations of Gaussians, designated as contracted Gaussians, could approximate atomic orbitals with great accuracy. Ever since, contracted Gaussians have been the basis set of choice, being used in all of the major program packages.

Computation of ERIs has a long history. Initially, all molecular integrals were calculated explicitly since closed formulas for integrals over Gaussians were easily derived. The explicit analytical formulas being specific to each integral do not allow the systematic calculation of integrals of higher angular momentum. Several recursive methodologies were then developed and gained acceptance in modern computational quantum chemistry programs. In this category are included the methods of Rys polynomials [5, 6], McMurchie and Davidson [7] and Obara and Saika [8]. More recently, approximated methodologies have been developed to speed up the computation of ERIs, for example, approaches using density fitting or the Cholesky decomposition. A very good and recent review of the calculation of ERIs has been published by Reine et al. [9].

The methodology to compute ERIs proposed in this work differs in concept and praxis relative to previous and current approaches. Existing methodologies need to be generic and applicable to any basis set. In contrast, the method being proposed approximates a predetermined basis set and is designed to be fast using simple linear algebra operations. The new computational methodology will use single-ζ (double-ζ for transition metals) basis sets.

This paper is a proof-of-concept on the development of accurate numerical approximations to the explicit analytical formulas for ERIs. The work will focus on the integral $(p_{xA}p_{xB}|p_{xC}p_{xC})$ with atoms A, B, and C being carbon. The choice of three-center integrals offers significant advantages. The numerical approximations are simpler in three-center ERIs than in four-center ERIs because fewer coordinates are required for the approximations. By keeping all elements the same, significant symmetry relationships are introduced and smaller domains for the coordinates can be considered (see Sect. 4; Fig. S1), thus reducing the number of target points that are required for the numerical approximations. Three-center ERIs are very important within density fitting approaches, in particular ERIs of the type $(\phi_A\phi_B|\eta_C)$, where η_C is used in the expansion of the product $\phi_C(r)\phi_D(r)$ [9]. The four-center integrals will be approximated using a different approach and will be the subject of a different publication.

2 Theoretical background

2.1 Revisit the explicit analytical calculation of two-electron-repulsion integrals

In the early years of computational chemistry, ERIs were calculated using explicit analytical formulas [10]. The notation used for the explicit expressions of ERIs over Cartesian Gaussian functions is kept as close as possible to the one used by Clementi [11]. An important concept in molecular orbital theory is the expansion of the basis functions $\phi_i(A)$ as liner combinations of primitive Cartesian Gaussian type functions (GTFs):

$$\phi_i(A) = \sum_{a=1}^{N} c_{i,a}\eta_a(A) \tag{1}$$

Cartesian GTFs are composed of a radial Gaussian function multiplied by Cartesian coordinates x, y, and z with exponents l_i, m_i, and n_i.

$$\eta_i(A) = x_A^{l_i} y_A^{m_i} z_A^{n_i}(-\alpha_i r_A^2) \tag{2}$$

The basic steps required to derive the explicit analytical formula of $(p_{xA}p_{xB}|p_{xC}p_{xC})$ are briefly described. Testing the novel numerical algorithms on three-center ERIs is important because they are significantly simpler than the four-center counterparts, due to having fewer degrees of spatial freedom, while still requiring the same techniques to perform the approximation. ERIs over basis functions are themselves written as linear combinations of the primitive GTFs:

$$(\phi_A\phi_B|\phi_C\phi_D) = \sum_{a,b,c,d} c_a c_b c_c c_d (\eta_A\eta_B|\eta_C\eta_D) \tag{3}$$

The advantage of using GTFs stemming from the Gaussian product rule is that the product of two GTFs is another GTF. In Eq. (3), the product of the first pair centered at $\mathbf{A}$ and $\mathbf{B}$ results in the general formula:

$$\begin{aligned} \eta(\alpha_1, \mathbf{A}, l_1)\eta(\alpha_2, \mathbf{B}, l_2) &= \exp\left(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1\right) \\ &\times \sum_{i=0}^{l_1+l_2} f_i\left(l_1, l_2, \overline{PA}_x, \overline{PB}_x\right) x_P^i \exp\left(-\gamma_1 x_P^2\right) \\ &\times \sum_{j=0}^{m_1+m_2} f_j\left(m_1, m_2, \overline{PA}_y, \overline{PB}_y\right) y_P^j \exp\left(-\gamma_1 y_P^2\right) \\ &\times \sum_{k=0}^{n_1+n_2} f_k\left(n_1, n_2, \overline{PA}_z, \overline{PB}_z\right) z_P^k \exp\left(-\gamma_1 z_P^2\right) \end{aligned} \tag{4}$$

with

$$\gamma_1 = \alpha_1 + \alpha_2 \tag{5}$$

and

$$\mathbf{P} = \frac{\alpha_1\mathbf{A} + \alpha_2\mathbf{B}}{\gamma_1} \tag{6}$$

Similar equations can be derived for the second pair with:

$$\gamma_2 = \alpha_3 + \alpha_4 \tag{7}$$

and

$$\mathbf{Q} = \frac{\alpha_3\mathbf{C} + \alpha_4\mathbf{D}}{\gamma_2} \tag{8}$$

The functions $f_i, \ldots, f_{k'}$ appearing in Eq. (4) result from the application of the binomial theorem to the products of Gaussian functions. Their generic formula is:

$$f_i(l_1, l_2, A, B) = \sum_{j=\max(0,i-l_2)}^{\min(i,l_1)} \frac{l_1! l_2! A^{l_1-j} B^{l_2-i+j}}{j!(l_1-j)!(i-j)!(l_2-i+j)!}. \tag{9}$$

Explicit values of the function $f_i(l_1, l_2, A, B)$ are given in Table 1 up to $l_1 + l_2 = 4$. Substituting the pairs $\eta(\alpha_1, \mathbf{A}, l_1)\eta(\alpha_2, \mathbf{B}, l_2)$ and $\eta(\alpha_3, \mathbf{C}, l_3)\eta(\alpha_4, \mathbf{D}, l_4)$ into Eq. (3) results in the formal formula for the explicit analytical calculation of ERIs (Eq. 10) where the normalization factors are written as N_α.

$$(\phi_A\phi_B|\phi_C\phi_D) = \sum_{a,b,c,d} c_a c_b c_c c_d (\eta_A\eta_B|\eta_C\eta_D)$$

$$= \sum_{a,b,c,d} c_a c_b c_c c_d N_a N_b N_c N_d$$

$$\times \exp\left(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1\right) \exp\left(-\alpha_1\alpha_2\overline{CD}^2/\gamma_1\right)$$

$$\times \sum_{i=0}^{l_1+l_2} f_i\left(l_1,l_2,\overline{PA}_x,\overline{PB}_x\right) \sum_{j=0}^{m_1+m_2} f_j\left(m_1,m_2,\overline{PA}_y,\overline{PB}_y\right)$$

$$\times \sum_{k=0}^{n_1+n_2} f_k\left(n_1,n_2,\overline{PA}_z,\overline{PB}_z\right)$$

$$\times \sum_{i'=0}^{l_3+l_4} f_{i'}\left(l_3,l_4,\overline{QC}_x,\overline{QD}_x\right) \sum_{j'=0}^{m_3+m_4} f_{j'}\left(m_3,m_4,\overline{QC}_y,\overline{QD}_y\right)$$

$$\times \sum_{k'=0}^{n_3+n_4} f_{k'}\left(n_3,n_4,\overline{QC}_z,\overline{QD}_z\right)$$

$$\times \iint x_{P_1}^i y_{P_1}^j z_{P_1}^k x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'} \frac{1}{r_{12}} \exp\left(-\gamma_1 r_{P_1}^2 - \gamma_2 r_{Q_2}^2\right) dV_1 dV_2$$

$$(10)$$

A simplified notation, $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k | x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$, is introduced for the integral $\iint x_{P_1}^i y_{P_1}^j z_{P_1}^k x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'} \frac{1}{r_{12}}$ $\exp(-\gamma_1 r_{P_1}^2 - \gamma_2 r_{Q_2}^2)dV_1 dV_2$ in the remaining of the text.

Calculation of ERIs according to Eq. (10) requires repeated evaluations of $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k | x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$, "f" functions, normalization factors, and the two exponential functions $\exp(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1)$ and $\exp(-\alpha_3\alpha_4\overline{CD}^2/\gamma_2)$ over multiple loops. There are loops over the contraction coefficients, c_a, c_b, c_c, and c_d, and the indices i, j, etc. The indices i, j, etc., determine the "f" functions and the integrals $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k | x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$. The index i runs between zero and $l_1 + l_2$ and similarly for j, k, i', …, which depend on $m_1 + m_2$, $n_1 + n_2$, $l_3 + l_4$, etc. When the exponents i, j, … are zero, the corresponding $x_{P_1}^i$, $y_{P_1}^j$ terms are indicated as "1" in $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k | x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$.

Although the complexity of the integrals $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k | x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$ increases with larger values of l_1, l_2, …, each is a well-defined function of **P** and **Q**, through the distance $\overline{PQ}$ and the corresponding nonzero projections along the Cartesian axis $\overline{PQ}_x$, $\overline{PQ}_y$, and $\overline{PQ}_z$. Recalling the definitions of **P** and **Q** from Eqs. (6) to (8), respectively, the integrals $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k | x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$ are functions of the coordinates of A, B, C, and D (A, B, and C in the three-center case).

The main purpose of this work is to illustrate how the three-center integrals $(p_{xA}p_{xB}|p_{xC}p_{xC})$, when carbon atoms are placed at the centers A, B, and C, can be calculated accurately through numerical approximation. In the following Sections, the specific simplifications introduced by

Table 1 Possible values of the f function as a function of the quantum numbers l_1, l_2 and generic parameters A and B

$$f_i(l_1, l_2, A, B) = \sum_{j=\max(0,i-l_2)}^{\min(i,l_1)} \frac{l_1! l_2! A^{l_1-j} B^{l_2-i+j}}{j!(l_1-j)!(i-j)!(l_2-i+j)!}$$

	$A \neq 0, B \neq 0$	$A = 0, B = 0$
$l_1 = 2, l_2 = 2$		
$i = 0$	A^2B^2	0
$i = 1$	$2AB^2 + 2A^2B$	0
$i = 2$	$B^2 + 4AB + A^2$	0
$i = 3$	$2B + 2A$	0
$i = 4$	1	1
$l_1 = 2, l_2 = 1$		
$i = 0$	A^2B	0
$i = 1$	$2AB + A^2$	0
$i = 2$	$B + 2A$	0
$i = 3$	1	1
$l_1 = 2, l_2 = 0$		
$i = 0$	A^2	0
$i = 1$	$2A$	0
$i = 2$	1	1
$l_1 = 1, l_2 = 1$		
$i = 0$	AB	0
$i = 1$	$B + A$	0
$i = 2$	1	1
$l_1 = 1, l_2 = 0$		
$i = 0$	A	0
$i = 1$	1	1
$l_1 = 0, l_2 = 0$		
$i = 0$	1	1

considering a three-center ERI and the mathematical details of the numerical approximations are discussed.

2.2 Numerical fitting of three-center two-electron-repulsion integrals $(p_{xA}p_{xB}|p_{xC}p_{xC})$

Calculation of the three-center ERI $(p_{xA}p_{xB}|p_{xC}p_{xC})$ involves significant simplifications resulting from many of the "f" factors becoming null, according to Table 1. After the null terms are omitted, Eq. (10) can be rewritten as:

$$(p_{xA}p_{xB}|p_{xC}p_{xC}) = \sum_{a,b,c} c_a c_b c_c^2 N_a N_b N_c^2 \times \exp\left(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1\right)$$

$$\times \left[\overline{PA}_x\overline{PB}_x\left\{111|x_{Q_2}^2 11\right\} + \left(\overline{PA}_x + \overline{PB}_x\right)\right.$$

$$\times \left.\left\{x_{P_1}11|x_{Q_2}^2 11\right\} + \left\{x_{P_1}^2 11|x_{Q_2}^2 11\right\}\right]$$

$$(11)$$

The expressions of $\{x_{P_1}^2 11|x_{Q_2}^2 11\}$, $\{x_{P_1} 11|x_{Q_2}^2 11\}$, and $\{111|x_{Q_2}^2 11\}$ are, respectively:

$$\left\{x_{P_1}^2 11\big|x_{Q_2}^2 11\right\} = \frac{\pi^{5/2}}{2\beta(\gamma_1 + \gamma_2)^{7/2}}$$

$$\times \left\{ 4\beta^2 F_4(t)\overline{PQ}_x^4 - 12\beta F_3(t)\overline{PQ}_x^2 \right.$$

$$+ \left[2(\gamma_1 + \gamma_2)\overline{PQ}_x^2 + 3\right] F_2(t)$$

$$\left. - \frac{(\gamma_1 + \gamma_2)}{\beta} F_1(t) + \frac{(\gamma_1 + \gamma_2)}{\beta} F_0(t) \right\} \tag{12a}$$

$$\left\{x_{P_1} 11\big|x_{Q_2}^2 11\right\} = \frac{\pi^{5/2}}{\beta(\gamma_1 + \gamma_2)^{7/2}}$$

$$\times \left\{ 2\gamma_2 \beta F_3(t)\overline{PQ}_x^3 - [2\gamma_2 F_2(t) + (\gamma_1 + \gamma_2)F_1(t)]\overline{PQ}_x \right\} \tag{12b}$$

$$\left\{111\big|x_{Q_2}^2 11\right\} = \frac{\pi^{5/2}}{\beta(\gamma_1 + \gamma_2)^{7/2}}$$

$$\times \left[2\gamma_2^2 F_2(t)\overline{PQ}_x^2 - \frac{\gamma_2(\gamma_1 + \gamma_2)}{\gamma_1} F_1(t) + \frac{(\gamma_1 + \gamma_2)^2}{\gamma_1} F_0(t) \right] \tag{12c}$$

where β is defined as $(\gamma_1\gamma_2)/(\gamma_1 + \gamma_2)$ and the terms $F_n(t)$ are the Boys function:

$$F_n(t) = \int_0^1 x^{2n} \exp(-tx^2)dx \tag{13}$$

exponents and separation between points **P** and **Q**, being independent of the spatial orientation of the system. In contrast, the factor $\overline{PQ}_x$ (also y and z), which is the x component of the vector **PQ**, depends on the orientation of the system. The "f" functions also introduce terms that depend on the orientation of the system, $\overline{PA}_x\overline{PB}_x$ and $(\overline{PA}_x + \overline{PB}_x)$ (see Table 1 and Eq. 11).

The algorithm developed for the calculation of ERIs is based on the multivariate numerical approximation of all functions contributing to the integrals in the desired interval. The terms contributing to Eq. (11) consisting of products of Eqs. (12a)–(12c) and their respective "f" terms from Table 1 have complex spatial dependencies resulting from the Boys functions, $\overline{PA}_x\overline{PB}_x$, $(\overline{PA}_x + \overline{PB}_x)$, and $\overline{PQ}_x$ terms. The strategy used in this work consists in recasting the terms making the total ERI in terms of simpler functions which are products of the rotationally dependent functions $\overline{PQ}_x$, $\overline{PA}_x\overline{PB}_x$, and $(\overline{PA}_x + \overline{PB}_x)$, designated as g_n^{rot}, and a rotationally invariant term, G_n. The index n is the exponent of $\overline{PQ}_x$. In addition to the g_n^{rot} and G_n terms, there is an additional rotationally dependent term derived from $\overline{PA}_x\overline{PB}_x$, $g^{rot}(\overline{PA}_x\overline{PB}_x)$. The corresponding rotationally invariant term is designated as $G_{\overline{PA}_x\overline{PB}_x}$. Using the terms of $\overline{PQ}_x^4$ for illustration, the rotationally dependent functions g_4^{rot} and the corresponding rotationally independent term G_4 are calculated as:

$$g_4^{rot}\left(\overline{PQ}_x^4\right) = \frac{\sum_{a,b,c,d} c_a c_b c_c c_d N_a N_b N_c^2 \times \exp\left(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1\right) \times \frac{4\pi^{5/2}\beta}{2(\gamma_1+\gamma_2)^{7/2}} \times \left(\overline{PQ}_x^4\right) \times F_4(t)}{\sum_{a,b,c,d} c_a c_b c_c c_d N_a N_b N_c^2 \times \exp\left(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1\right) \times \frac{4\pi^{5/2}\beta}{2(\gamma_1+\gamma_2)^{7/2}} \times F_4(t)} \tag{14}$$

The evaluation of the Boys function had a recent renewed interest and was the subject of recent publications [12, 13]. A different algorithm was developed for this work and will be discussed in a forthcoming paper.

The integrals $\{x_{P_1}^i y_{P_1}^j z_{P_1}^k|x_{Q_2}^{i'} y_{Q_2}^{j'} z_{Q_2}^{k'}\}$ have important characteristics that can be explored to simplify the numerical approximations. The factors γ_1, γ_2, and β depend on the orbital exponents and are unaffected by geometrical changes. The Boys functions $F_n(t)$ depend on the orbital

$$G_4 = \sum_{a,b,c,d} c_a c_b c_c c_d N_a N_b N_c^2 \times \exp\left(-\alpha_1\alpha_2\overline{AB}^2/\gamma_1\right)$$

$$\times \frac{4\pi^{5/2}\beta}{2(\gamma_1 + \gamma_2)^{7/2}} \times F_4(t) \tag{15}$$

The important rotationally invariant term G_0, which makes a direct contribution to the total computed ERI, is

$$G_0 = \frac{\pi^{5/2}}{\beta(\gamma_1 + \gamma_2)^{7/2}} \left[6F_2(t) - \frac{2(\gamma_1 + \gamma_2)}{\beta} F_1(t) \right.$$

$$\left. + \frac{2(\gamma_1 + \gamma_2)}{\beta} F_0(t) - \frac{\gamma_2(\gamma_1 + \gamma_2)}{\gamma_1} F_1(t) + \frac{(\gamma_1 + \gamma_2)^2}{\gamma_1} F_0(t) \right] \tag{16}$$

3 Mathematical background

3.1 Multivariate approximation

In many applications, it is convenient to introduce approximate functions. An approximate function $g(\mathbf{x})$ is a function that given m data points $\mathbf{x}$ approximates the target values produced by the function $f(\mathbf{x})$ as closely as possible according to some metric. The approximant $g(\mathbf{x})$ is desired to be as smooth and compact as possible. The need to approximate often occurs when it is too costly or complex to use the true function, or even when the true function is unknown. The mathematical theory of approximation is well documented (see for example Ref. [14]).

This work explores the possibility of approximating the complex and computationally costly Eq. (11) with simpler, and faster to evaluate, functions. All approximants are based on polynomial expansions (Eq. 17), in which the coefficients a_i are scalars and the basis functions $H_i(\mathbf{x})$ can take different forms:

$$f(\mathbf{x}) = a_0 + a_1 H_1(\mathbf{x}) + \cdots + a_n H_n(\mathbf{x}) \tag{17}$$

The main criterion to determine the quality of an approximation is the measurement of the "distance" between target data points and the same set of points as obtained by the specified approximating function (approximant). It is important that the target and approximated points are as close as possible. A suitable metric to account for the global different between the set of true values and their respective approximations used extensively in this work is the l^2-norm.[1]

The multivariate scheme developed in this work to approximate ERIs consists of multiple levels of bivariate (or univariate) approximants, with the fitting variables of a given level being expressed in terms of the variables of the next immediate level. The methodology is illustrated with the help of a 3-dimensional model depicted in Fig. 1. To approximate the point $f(x_1, x_2, x_3)$, represented by the red sphere, a numerical approximant $g(x_1, x_2)$ of all points on the (x_1, x_2) plane is first developed. The function $g(x_1, x_2)$ is expanded in terms of primitive functions $H_i(x_1, x_2)$ according to Eq. (17). The dependency of x_3, which is illustrated in Fig. 1 by the vector originating at the blue sphere, is carried by fitting parameters a_i as functions of x_3. In mathematical terms, the dependency of the fitting parameters a_i is given by another expansion similar to Eq. (17). The basis functions are represented by $H_i'(x_3)$ and the expansion has adjustable coefficients a_i':
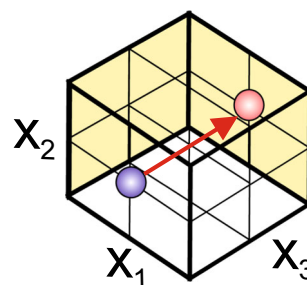

Fig. 1 Illustration of the spatial dependency of multilayered approximating functions

$$a(x_3) = a_0' + \sum_{i=1}^{n'} a_i' H_i'(x_3) \tag{18}$$

The process can be repeated multiple times, generating complex dependencies of multivariate functions. However, as the number of fitting parameters grows very fast with each additional layer of variable dependencies, in practice, the process is limited to a small number of layers.

In this work, the need for smooth functions arises because of the multiple dependencies of the variables. Since the fitting parameters carry additional dependencies themselves, it is important that they are as smooth as possible to avoid discontinuities that make the next level fittings more complex. Other important criteria in defining the fitting process are computational efficiency, simplicity of algorithm implementation, and future evolution of the method. When designing algorithms for numerical approximation, it is important to consider how fast and accurate the method is in the present and to have a clear plan for future development.

The fitting functions were chosen to be bivariate Chebyshev orthogonal polynomials. Chebyshev polynomials form an important class of functions in curve fitting [15]. A similar expansion can be developed for surfaces $f(x, y)$ where the polynomial is based on to Chebyshev series with $\bar{x}, \bar{y} \in [-1, 1] \times [-1, 1]$:

$$f(x, y) = \sum_{i=0}^{n_x} \sum_{j=0}^{n_y} a_{ij} T_i(\bar{x}) T_j(\bar{y}) \tag{19}$$

The 2-dimensional Chebyshev expansion was evaluated directly by computing the polynomials and summing all contributions according to Eq. (19).

3.2 Choice of coordinates

Each of the terms g_n^{rot}, $g^{\text{rot}}(\overline{PA}_x \overline{PB}_x)$, G_n, and $G_{\overline{PA}_x \overline{PB}_x}$ required to calculate ERIs according to the prescription of Eqs. (14)–(16) can be expressed in terms of a finite number of variables. Fitting of three-center ERIs requires six coordinates that are used to position the atomic centers
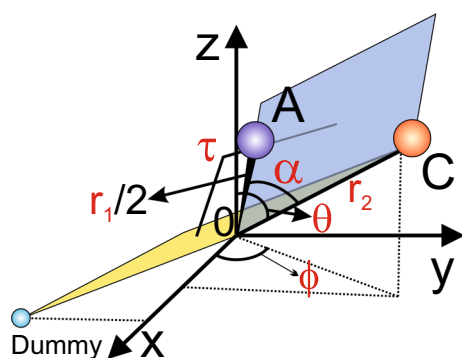
Fig. 2 Illustration of the coordinates used in the fitting of three-center electron-repulsion integrals

carrying the basis functions. Importantly, only the total number of variables has to be fulfilled and not the nature of the individual coordinates as long as they provide the spatial assignment of the atomic centers. It is, however, advisable to use combinations of variables that lead to simpler fitting expressions, in addition to having physical meanings that can be related to common geometrical transformations. In this respect, bond distances, angles, and torsions are prime candidates.

The protocol followed in this work consists in separating the rotationally dependent terms of $\overline{PQ}_x^n$ and $\overline{PA}_x\overline{PB}_x$ from the rotational invariant counterparts. The set of coordinates chosen for the fitting of the rotationally invariant part is two distances, r_1 and r_2, and the internal angle α. r_1 is the distance between the centers A and B and r_2 is the separation between **C** and the midpoint of **AB**, represented by **O**. α is the angle $\widehat{COA}$ (see Fig. 2). The projection $\overline{PQ}_x$ (also $\overline{PQ}_y$ and $\overline{PQ}_z$) and $\overline{PA}_x\overline{PB}_x$ require special attention since they affect the rotational invariance of the integrals. Its spatial dependency is significantly more complex, requiring an extra set of coordinates. The extra variables are the polar spherical coordinates θ and φ, which are used to position the atomic center C and the dihedral angle τ, which is used to determine the relative orientation of centers A (and B) relative to C (see Fig. 2).

4 Results and discussion

The following section is dedicated to evaluating the accuracy, and speed, of the numerical algorithm to approximate ERIs. Emphasis is placed on testing the ability of the method to accurately reproduce the integral $(p_{xA}p_{xB}|p_{xC}p_{xC})$ with carbon atoms at A, B, and C. All calculations were based on the STO-6G basis set. This basis set is sufficiently small to allow computation of the many target ERIs used in the parameterization in a

reasonable time. All calculations were done on an AMD Ryzen 1700 CPU and 24 GByte of RAM memory. All codes were compiled with Gfortran using the –O3 compiler flag. The approximation of the rotationally independent terms is discussed first, with G_4 and $G_{\overline{PA}_x\overline{PB}_x}$ being used as examples. Afterward, the fitting of the rotationally dependent terms is analyzed. The approximating methodologies are illustrated with the help of $g^{rot}(\overline{PA}_x\overline{PB}_x)$, since it is the largest contributor to the total ERI and is also representative of the other terms. The accuracy and speed of the multivariate methodology of approximation are discussed in Sects. 4.2 and 4.3.

The goodness-of-fit of the approximants was measured in terms of the root-mean-square error (RMSE). RMSE measures the total deviation of the computed from the target values, and a value closer to zero indicates the fit is good and is useful for prediction.

The domains of the variables influencing the rotation of the systems are $\alpha \in [0, 180°]$ and $\tau, \theta, \varphi \in [0, 90°]$. The domains of τ, θ, and φ are limited to 90° because of the symmetry relations resulting from having the same element at the positions A, B, and C. Figure S1 illustrates the symmetry effects for the dependencies of (α, τ) and (θ, φ) for the function $g_4^{rot}(\overline{PQ}_x^4)$.

## 4.1 Fitting the rotationally independent terms $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA}_x\overline{PB}_x}(\alpha, r_1, r_2)$

The protocol described in Sect. 3.1 for the multivariate fitting of the different terms making the explicit analytical expression of the ERIs starts with the initial fitting of rotationally invariant functions $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA}_x\overline{PB}_x}(\alpha, r_1, r_2)$. In most cases, these are auxiliary functions used to create smoother rotationally dependent surfaces that are easier to fit, although $G_0(\alpha, r_1, r_2)$ contributes directly to the final integral (Eq. 16). The dependencies of G_n and $G_{\overline{PA}_x\overline{PB}_x}$ are on the angle α and distances r_1 and r_2. The protocol followed in this work for the fitting of $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA}_x\overline{PB}_x}(\alpha, r_1, r_2)$ calls to the initial fitting of the α dependency. The plots of $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA}_x\overline{PB}_x}(\alpha, r_1, r_2)$ functions relative to α (with r_1 and r_2 fixed) show a similar symmetric sinusoidal curve. The dependency of $G_4(\alpha, r_1 = 2.6$ a.u.$, r_2 = 5.0$ a.u.$)$ is illustrated on Fig. S2. The function of choice for the fitting of the dependency of α, in radians, was a rational polynomial written of the form:

$$f'(\alpha) = \Delta(r_1, r_2)\left(\frac{a_1(r_1, r_2)\alpha^2 + a_2(r_1, r_2)\alpha^4}{1.0 + a_3(r_1, r_2)\alpha^2 + a_4(r_1, r_2)\alpha^4 + a_5(r_1, r_2)\alpha^6}\right)$$
$$+ a_0(r_1, r_2)$$

(20)

Fig. 3 Illustration of the spatial dependency of a_0 and a_5 of the rotationally independent term. $G_{\overline{PA_x}\overline{PB_x}}(\alpha, r_1, r_2)$. The surfaces are smooth and suitable for accurate approximation
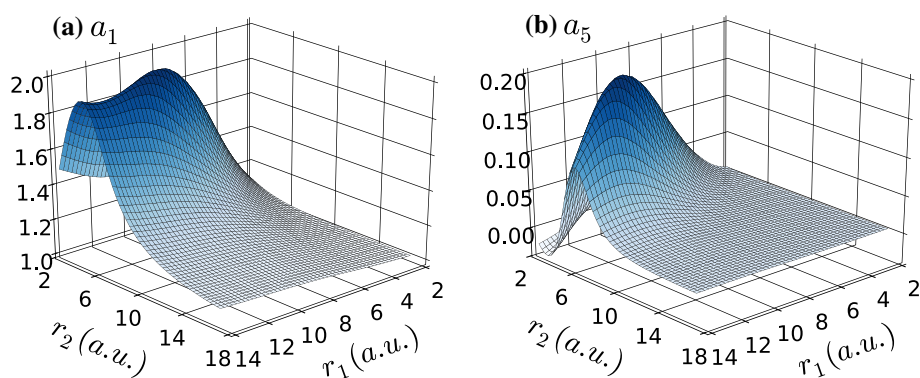
The dependency of each a_i term on (r_1, r_2) is highlighted in Eq. (20), in accordance with the multivariate fitting algorithm described in Sect. 3.1. The rational expansion was found to be acceptable in terms of computational cost and accuracy. Because of the dependence of the coefficients on the distances r_1 and r_2, it is important to keep the polynomial expansion above as compact as possible.

The dependency of the coefficients a_0 and a_5 in Eq. (20) on (r_1, r_2) is illustrated in Fig. 3. The key assumption of this work that the fitting coefficients of a polynomial approximant at a certain level have smooth spatial dependencies of the variables of the next level, and thus, are able to carry that spatial dependency, is fully fulfilled. Although no rigorous mathematical proof is presented, the surfaces of a_i and Δ are smooth and can be approximated using the multivariate techniques presented before. Each coefficient $a_i(r_1, r_2)$ and Δ were fitted with bivariate Chebyshev polynomials with arguments $\bar{r}_1$ and $\bar{r}_2$ (see Eq. 1). The order of the expansion was truncated at order 20. It is the largest order of the Chebyshev polynomials used in this work. The computational cost of using such a long expansion is not prohibitive for two reasons. First, each increase in the order of the Chebyshev polynomial only contributes its associated number of parameters times the number of parameters of the univariate polynomial expansion of α. Second, the rotationally invariant terms only need to be calculated once and can be stored. The same rotationally invariant term, $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA_x}\overline{PB_x}}(\alpha, r_1, r_2)$, can be used in the fitting of all ERIs regardless of involving p_x, p_y, or p_z functions. The explicit expression for fitting all spatial dependencies of $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA_x}\overline{PB_x}}(\alpha, r_1, r_2)$, combining Eq. (2) above with the bivariate (r_1, r_2) Chebyshev polynomial for each of the coefficients a_i is

$$
\begin{aligned}
G_n&(\alpha, r_1, r_2) \quad \text{or} \quad G_{\overline{PA_x}\overline{PB_x}}(\alpha, r_1, r_2) \\
&= \left(\sum_{i=0}^{n} \sum_{j=0}^{n} a_{ij}^{(\Delta)} T_i(\bar{r}_1) T_j(\bar{r}_2) \right) \\
&\quad \times \left[\frac{\left(\sum_{i=0}^{n} \sum_{j=0}^{n} a_{ij}^{(1)} T_i(\bar{r}_1) T_j(\bar{r}_2) \right) \alpha^2 + \cdots}{1.0 + \left(\sum_{i=0}^{n} \sum_{j=0}^{n} a_{ij}^{(3)} T_i(\bar{r}_1) T_j(\bar{r}_2) \right) \alpha^2 + \cdots} \right] \\
&\quad + \left(\sum_{i=0}^{n} \sum_{j=0}^{n} a_{ij}^{(0)} T_i(\bar{r}_1) T_j(\bar{r}_2) \right)
\end{aligned}
\tag{21}
$$

where $T_i(\bar{r}_1)$ is the Chebyshev polynomial of the first kind of degree i with argument $\bar{r}_1$, and $T_j(\bar{r}_2)$ is similarly defined for j and $\bar{r}_2$. The approximations are extremely accurate with overall RMSEs lower than 5.0E−09. The residuals for $G_n(\alpha, r_1, r_2)$ and $G_{\overline{PA_x}\overline{PB_x}}(\alpha, r_1, r_2)$ are plotted in Figs. S3 and S4, respectively, for $\alpha = 49.1°$ and $\tau = 66.7°$.

4.2 Fitting the rotationally dependent terms $g_n^{rot}(\overline{PQ_x^n})$ and $g^{rot}(\overline{PA_x}\overline{PB_x})$

The rotationally dependent functions hold the effects of the $\overline{PQ_x^n}$, $\overline{PA_x}\overline{PB_x}$, and $(\overline{PA_x} + \overline{PB_x})$ terms in the three-center ERI of the kind $(p_{xA}p_{xB}|p_{xC}p_{xC})$. These are considerably more challenging to approximate, since they depend on six variables instead of the three variables in the rotationally independent terms.

Two quantities were defined to measure the contribution of each term to the total ERI: the average absolute percentage contribution (AAPC) and maximum absolute percentage contribution (MAPC). The absolute value of each term was chosen because each can be positive or negative. The AAPC and MAPC quantities are calculated for the term f_a as, respectively, $\text{AAPC} = 100\times$

$[\sum_i^n (|f_{a,i}|/n(|f_{1,i}| + \cdots + |f_{7,i}|))]$ and $100 \times \max[|f_{a,i}|/(|f_{1,i}| + \cdots + |f_{7,i}|)]$. Values closer to 100% indicate a stronger contribution to the integral and likewise values close to zero mean smaller contributions. The most significant contributors are the rotationally independent term G_0 and the rotationally dependent function of $\overline{PA}_x\overline{PB}_x$. Interestingly, all the terms containing the projections $\overline{PQ}_x$ make considerably smaller contributions (see Table 2). According to the relevance of each term, different expansions can be defined without imparting significantly the accuracy of the approximation. The G_0 term is already fitted with the highest order of any Chebyshev polynomial, and the accuracy of the approximations can be hardly improved.

The fitting protocol for $g^{rot}_n(\overline{PQ}_x^n)$ and $g^{rot}(\overline{PA}_x\overline{PB}_x)$ requires three layers of fittings using bivariate Chebyshev polynomials. The pairing of variables is: $(\theta, \varphi) \to (\alpha, \tau) \to (r_1, r_2)$. In the first step, fitting functions $f(\theta, \varphi)$ are determined for each of the target points (α, τ, r_1, r_2) (see Eq. 22a). The coefficients a_{ij} carry the dependency of the remaining variables α, τ, r_1, and r_2. In the second level of optimization, each of the coefficients a_{ij} is fitted similarly with a bivariate Chebyshev polynomial (Eq. 22b). The fitting coefficients b_{kl}^{ij} carry the dependency of (r_1, r_2) and each is fitted in the third level of fittings (Eq. 22c).

$$g^{rot} \approx f(\theta, \varphi)_{|\alpha^0,\tau^0,r_1^0,r_2^0}$$

$$= \left(\sum_{i=0}^n \sum_{j=0}^n a_{ij}(\alpha, \tau, r_1, r_2) T_i(\bar{\theta}) T_j(\bar{\varphi}) \right) \quad (22a)$$

$$a_{ij}(\alpha, \tau)_{|r_1^0,r_2^0} = \left(\sum_{k=0}^n \sum_{l=0}^n b_{kl}^{ij}(r_1, r_2) T_k(\bar{\alpha}) T_l(\bar{\tau}) \right) \quad (22b)$$

$$b_{kl}^{ij}(r_1, r_2) = \left(\sum_{i=0}^n \sum_{j=0}^n c_{mn}^{ij,kl} T_m(\bar{r}_1) T_n(\bar{r}_2) \right) \quad (22c)$$

In Eq. (22), the superscript "0" means that the corresponding variable assumes a fixed value.

Figure 4 illustrates selected surfaces $f(\theta, \varphi)$ for specific values of (r_1, r_2), and (α, τ). All surfaces have similar Gaussian-like shapes. It is noteworthy that to facilitate the numerical approximations, the surfaces were symmetrized through the change in coordinate $\theta^* = 180° - \theta$. The functions of $\overline{PQ}_x^n$ were approximated with Chebyshev polynomials of order 14. The rotationally dependent function of $\overline{PA}_x\overline{PB}_x$ was approximated with Chebyshev polynomials of different orders, from 8 to 20. The overall results are discussed in Sect. 4.3.

The next step in the fitting process consists in fitting the $a_{ij}(\alpha, \tau)$ surfaces. Figure 5 illustrates the coefficients a_1 and a_{231} of the $g^{rot}(\overline{PA}_x\overline{PB}_x)$ term for $r_1 = 2.5$ a.u. and $r_2 = 5.0$ a.u. Similarly to the approximation of the

Table 2 Average absolute percentage contribution (AAPC) and maximum absolute percentage contribution (AAPC) of each term contributing to $(p_{xA}p_{xB}|p_{xC}p_{xC})$

	Term						
	$\overline{PQ}_x^4$	$\overline{PQ}_x^3$	$\overline{PQ}_x^2$ $l_1 + l_2 = 0$	$\overline{PQ}_x$	$\overline{PQ}_x^2$ $l_1 + l_2 = 2$	$\overline{PA}_x\overline{PB}_x$	G_0
AAPC	0.4	0.1	0.8	0.4	1.0	45.7	51.6
MAPC	12.7	2.2	10.1	2.5	11.0	91.2	100.0

(a) $r_1 = 2.5$ a.u. $r_2 = 5.0$ a.u. $\alpha = 49.1°$ $\tau = 66.7°$

(b) $r_1 = 2.5$ a.u. $r_2 = 5.0$ a.u. $\alpha = 113.3°$ $\tau = 33.8°$

(c) $r_1 = 4.0$ a.u. $r_2 = 5.0$ a.u. $\alpha = 113.3°$ $\tau = 33.8°$
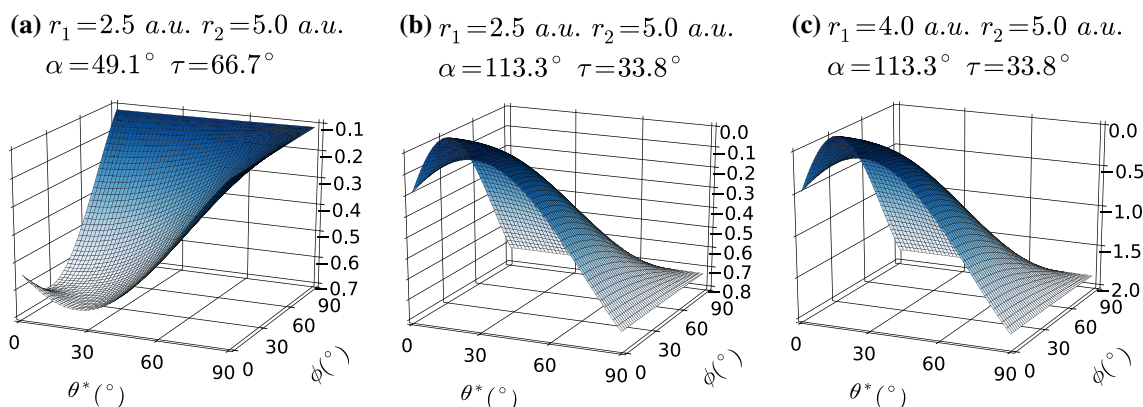
Fig. 4 Illustration of the dependency of θ and φ for fixed values of r_1, r_2, α, and τ for $g^{rot}(\overline{PA}_x\overline{PB}_x)$. **a, b** show the effect of varying α and τ, whereas **b** and **c** illustrate the effect of varying r_1, r_2. Fitting of each surface will require calculation of the coefficients a_{ij} of Eq. (22a)
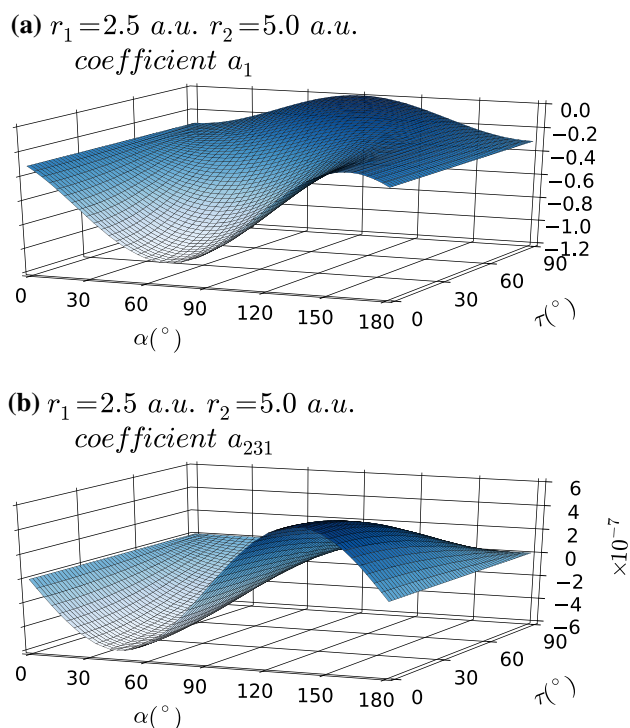
(a) $r_1 = 2.5$ a.u. $r_2 = 5.0$ a.u.
coefficient a_1

(b) $r_1 = 2.5$ a.u. $r_2 = 5.0$ a.u.
coefficient a_{231}

Fig. 5 Illustration of dependency of the coefficients a_{ij} of Eq. (22a) on the angles α and τ for fixed values of $r_1 = 2.5$ a.u. and $r_2 = 5.0$ a.u for $g^{\text{rot}}(\overline{PA}_x\overline{PB}_x)$. The surfaces are smooth and have distinctive magnitudes that can be explored to reduce the order of the polynomials used in the fittings

rotationally independent terms, the bivariate surfaces $a_{ij}(\alpha, \tau)$ are smooth and were easily approximated with bivariate Chebyshev polynomials. It was opted to approximate the $a_{ij}(\alpha, \tau)$ and $f(\theta, \varphi)$ surfaces with the Chebyshev polynomials of the same order for simplicity.

The final step in the approximation of the rotationally dependent functions is the fitting of the coefficients $b^{ij}_{kl}(r_1, r_2)$. In Fig. 6, the surfaces corresponding to b^1_1 and
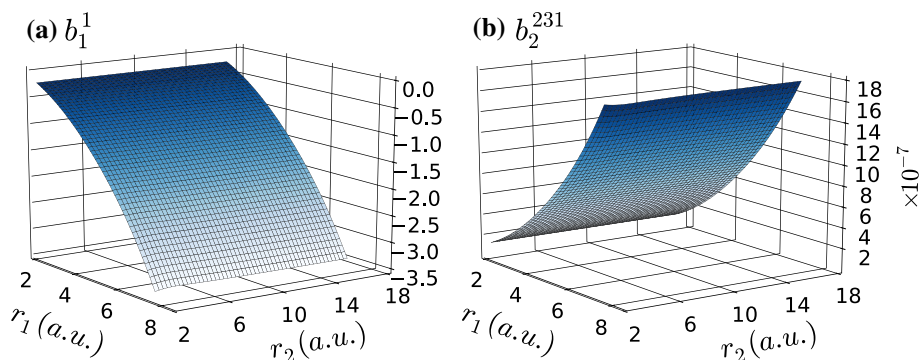
b^{231}_2 for $g^{\text{rot}}(\overline{PA}_x\overline{PB}_x)$ are shown as a function of the remaining coordinates r_1 and r_2. The dependency of the b^{ij}_{kl} coefficients is considerably simpler with the functions being monotonically increasing or decreasing in r_1 (i.e., for fixed values of r_2). Important simplifications can be introduced in the approximation of the b^{ij}_{kl} coefficients since many can be eliminated due to their small contribution to the $a_{ij}(\alpha, \tau)$ terms. While the terms derived from $g^n_n(\overline{PQ}^n_x)$ were approximated by Chebyshev polynomials of order 14, the effect of the length of the Chebyshev polynomials on the accuracy was tested for $g^{\text{rot}}(\overline{PA}_x\overline{PB}_x)$. $g^{\text{rot}}(\overline{PA}_x\overline{PB}_x)$ was chosen because of all rotationally dependent terms is single largest contributor to the total ERI. The quality of the different approximants of (θ, φ), $a_{ij}(\alpha, \tau)$, and $b^{ij}_{kl}(r_1, r_2)$ is gauged in Table 3, where goodness-of-fit results are presented and compared for the rotationally dependent terms. The total number of nonzero b^{ij}_{kl} terms is also shown. They determine the speed and accuracy of the proposed methodology. Starting with the $g^{\text{rot}}(\overline{PA}_x\overline{PB}_x)$ term, the RMSE values are good for all Chebyshev expansions and are systematically improved for higher orders. Very important is the slow growth of the number of nonzero b^{ij}_{kl} terms, meaning improvements of accuracy come with small additional computational costs. For the $g^{\text{rot}}_n(\overline{PQ}^n_x)$ terms, accuracy varies significantly, ranging from RMSE values of $\sim 10^{-7}$ for $g^{\text{rot}}_3(\overline{PQ}^3_x)$ and $\sim 10^{-5}$ for $g^{\text{rot}}_2(\overline{PQ}^2_x)(l_1 + l_2 = 2)$. Importantly, the highest RMSE occurs for terms with the lowest number of nonzero terms, opening the possibility of improving the accuracy without significant additional computational costs by using longer expansions for terms with fewer nonzero elements. The improvement of the accuracy with higher-order expansions is shown pictorially in Fig. S5 where histograms of residuals are plotted for the four expansions used for

(a) b^1_1

(b) b^{231}_2

Fig. 6 Illustration of the dependency of the coefficients $b^{ij}_{kl}(r_1, r_2)$ of Eq. (22b) on the distances r_1 and r_2 for $g^{\text{rot}}(\overline{PA}_x\overline{PB}_x)$. The surface on the left (**a**) is for $b^1_1(r_1, r_2)$ and the surface on the right (**b**) is for $b^{231}_2(r_1, r_2)$. The surfaces are monotonically increasing or decreasing

in the r_1 direction, i.e., for fixed r_2. These surfaces can be approximated with more compact Chebyshev polynomials, thus reducing the overall computational cost

Table 3 Goodness-of-fit estimates and number of nonzero elements b_{kl}^{ij} for terms $g_n^{rot}(\overline{PQ_x^n}), n = 1, 2, 3, 4$ and $g^{rot}(\overline{PA_x}\overline{PB_x})$

	$g_n^{rot}(\overline{PQ_x^n}), n = 1, 2, 3, 4$ Chebyshev order 14			$g^{rot}(\overline{PA_x}\overline{PB_x})$		
	RMSE	# nonzero b_{kl}^{ij}		Order of fitting	RMSE	# nonzero b_{kl}^{ij}
$g_4^{rot}(\overline{PQ_x^4})$	2.16E−06	5921		8	4.53E−05	1059
$g_3^{rot}(\overline{PQ_x^3})$	1.13E−07	6440		10	1.56E−05	1796
$g_2^{rot}(\overline{PQ_x^2})$ $l_1 + l_2 = 0$	7.42E−06	6536		14	4.55E−06	2157
$g_2^{rot}(\overline{PQ_x^2})$ $l_1 + l_2 = 2$	1.41E−05	3254		20	1.69E−06	4526
$g_1^{rot}(\overline{PQ_x})$	9.85E−06	2891				

$g^{rot}(\overline{PA_x}\overline{PB_x})$. The spread of the residuals becomes increasingly narrower and with the expansion of order 20 most errors are in the interval [−5.0E−06, 5.0E−06]. It is important to stretch that approximation with bivariate Chebyshev polynomials provides a way to systematically improve the quality of the approximation.

4.3 Adding all together: assembly of the computed $(p_{xA}p_{xB}|p_{xC}p_{xC})$ two-electron-repulsion integral

The culmination of this work on the numerical approximation of ERIs is the assembly of the calculated $(p_{xA}p_{xB}|p_{xC}p_{xC})$ ERIs from the different terms discussed above and the comparison with the real explicit analytical values. The contributors to the calculated ERI are the rotationally invariant term G_0, the rotationally dependent terms $g_n^{rot}(\overline{PQ_x^n}) \cdot G_n$ and $g^{rot}(\overline{PA_x}\overline{PB_x}) \cdot G_{\overline{PA_x}\overline{PB_x}}$. The approximation of G_0 was unique, using the highest order expansion of this work (Chebyshev polynomials of order 20 for the dependency of r_1 and r_2). Goodness-of-fit results for the total ERI are shown in Table 4. The results do not represent the full potential of the methodology since the contributions from both $g_2^{rot}(\overline{PQ_x^2}) \cdot G_2$ terms dominate. For this reason, there is no improvement when using Chebyshev polynomials of order 20 for $g^{rot}(\overline{PA_x}\overline{PB_x}) \cdot G_{\overline{PA_x}\overline{PB_x}}$ relative to order 14. However, the RMSEs for some of the $g_n^{rot}(\overline{PQ_x^n}) \cdot G_n$ terms can be improved approximately an order of magnitude by using longer Chebyshev polynomials as in $g^{rot}(\overline{PA_x}\overline{PB_x})$ terms (see Table 3). It is important to stretch not only the magnitude, but also the distribution of the residuals. In Fig. 7, the residuals for $r_1 = 2.5$ a.u., $r_2 = 5.0$ a.u., $\alpha = 49.1°$, 165.5°, and $\tau = 66.7°$, 61.6° are plotted as a function of θ and φ for two Chebyshev expansions of $g^{rot}(\overline{PA_x}\overline{PB_x})$ (order 20 and order 14). The dominant effect of both $g_2^{rot}(\overline{PQ_x^2}) \cdot G_2$ terms
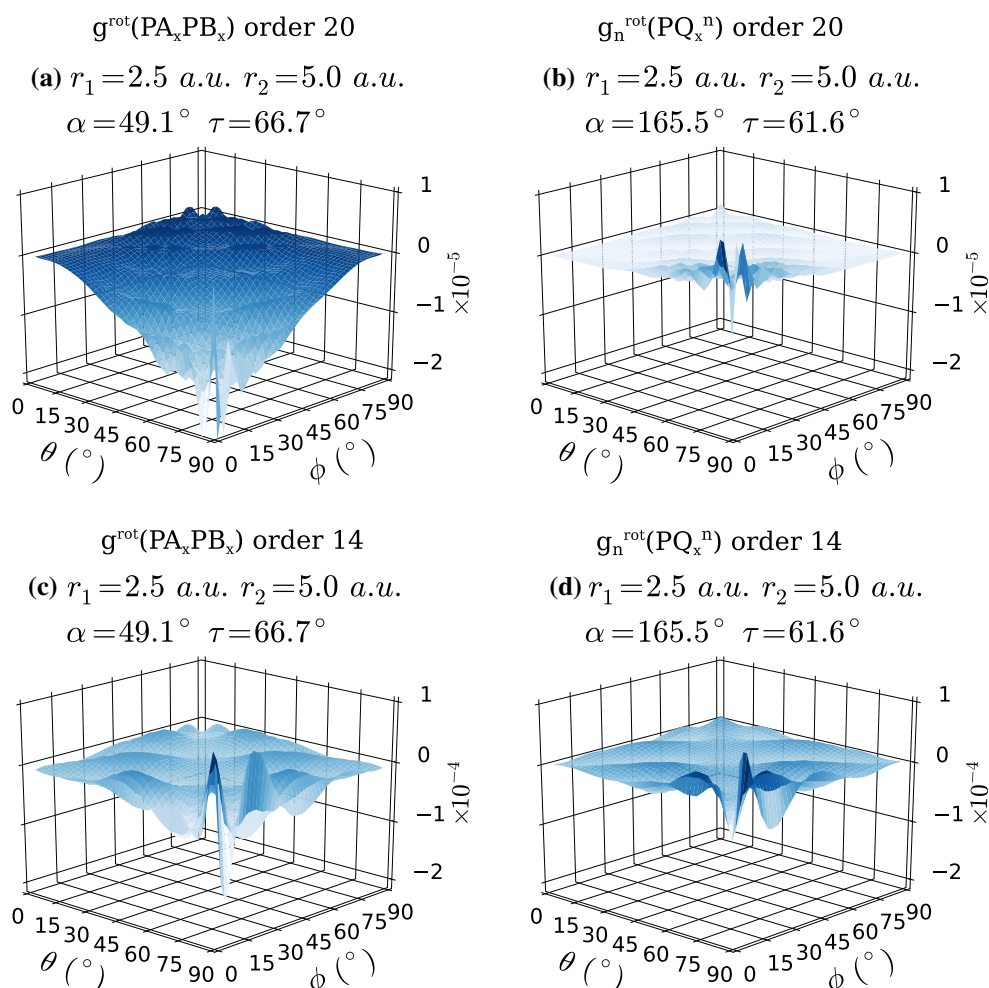
Table 4 Goodness-of-fit estimates and total execution timings for the approximated $(p_{xA}p_{xB}|p_{xC}p_{xD})$ ERI

Order of fitting of $g^{rot}(\overline{PA_x}\overline{PB_x})$		RMSE
Goodness-of-fit estimates		
8		6.07E−04
10		6.06E−04
14		1.59E−05
20		1.59E−05
	Numerical approximation (this work)	Uquantchem
Total execution timings		
Execution time (s)	80.5 ± 0.5	1017.5 ± 27.0

The terms $g_n^{rot}(\overline{PQ_x^n}), n = 1, 2, 3, 4$ were fitted with Chebyshev polynomials of order 14, and the term G_0 was fitted with the rational function of Eq. (20) and Chebyshev polynomials of order 20 (for the radial dependencies). The radial terms of the rotationally dependent elements were fitted with Chebyshev polynomials of order 14. The term $g^{rot}(\overline{PA_x}\overline{PB_x}$ was fitted with Chebushev polynomials of order 20

is evident on Fig. 7a where the total ERI surface follows roughly the contributions from $g_2^{rot}(\overline{PQ_x^2}) \cdot G_2 (l_1 + l_2 = 0)$ and $g_2^{rot}(\overline{PQ_x^2}) \cdot G_2 (l_1 + l_2 = 2)$ (see Fig S3 for plots of $g_n^{rot}(\overline{PQ_x^n}) \cdot G_n$ and Fig S4 for plots of $g^{rot}(\overline{PA_x}\overline{PB_x}) \cdot G_{\overline{PA_x}\overline{PB_x}}$). When $g^{rot}(\overline{PA_x}\overline{PB_x}) \cdot G_{\overline{PA_x}\overline{PB_x}}$ is also fitted with Chebyshev polynomials of order 14, the resulting ERI does not show any dominant contributions since all terms have a similar effect. The biggest residual is approximately one order of magnitude smaller when $g^{rot}(\overline{PA_x}\overline{PB_x}) \cdot G_{\overline{PA_x}\overline{PB_x}}$ is fitted with polynomials of order 20. It is important to stretch that accuracy can be systematically improved by using longer expansions for the fitting of selected $g_n^{rot}(\overline{PQ_x^n}) \cdot G_n$ terms. The same pattern was

Fig. 7 Illustration of the residuals of the total computed ERI for two fittings of the term $g^{rot}(\overline{PA_x PB_x})$ (Chebyshev polynomials of order 20 and 14). The rotationally dependent terms $g_n^{rot}(\overline{PQ_x^n})$ were only fitted with Chebyshev polynomials of order 14. Fitting $g^{rot}(\overline{PA_x PB_x})$ with order 20 polynomials results in residuals one order of magnitude smaller than when fitting with order 14 polynomials. Importantly, the biggest absolute residuals are typically localized at $\theta = 90°$ and $\varphi = 0°$

$g^{rot}(PA_x PB_x)$ order 20

(a) $r_1 = 2.5$ *a.u.* $r_2 = 5.0$ *a.u.*
$\alpha = 49.1°$ $\tau = 66.7°$

$g_n^{rot}(PQ_x^n)$ order 20

(b) $r_1 = 2.5$ *a.u.* $r_2 = 5.0$ *a.u.*
$\alpha = 165.5°$ $\tau = 61.6°$

$g^{rot}(PA_x PB_x)$ order 14

(c) $r_1 = 2.5$ *a.u.* $r_2 = 5.0$ *a.u.*
$\alpha = 49.1°$ $\tau = 66.7°$

$g_n^{rot}(PQ_x^n)$ order 14

(d) $r_1 = 2.5$ *a.u.* $r_2 = 5.0$ *a.u.*
$\alpha = 165.5°$ $\tau = 61.6°$

found for other points, and in the future an exhaustive statistical study will be performed to better understand the spatial dependencies of the total fitted ERIs.

The timings of the calculation of 505,000 ERIs using the numerical approximated developed in this work and a reference method are given in Table 4. The reference results are from Uquantchem [16] which computes ERIs using Rys polynomials [5, 6]. The total execution time of the present method is mostly dependent on the number of nonzero b_{kl}^{ij} terms. Importantly, longer Chebyshev expansions result in significant improvements of accuracy, while adding comparatively fewer b_{kl}^{ij} terms. It is safe to say that the methodology developed in this work can be systematically improved with only a small penalty on the total execution time. Despite the significant results, with a speedup of 12.6, consideration of the timings required for the numerical approximation of ERIs is secondary in this work. The calculations were performed monolithically using fixed polynomial expansions for most fittings, and no optimizations of the fitting polynomials were attempted. In the future, specific optimizations will be introduced and

Chebyshev polynomials of different orders will be tested to provide the best cost/benefit result.

5 Conclusions and future prospects

This work is the first step of a large effort to develop novel tight-binding computational methodologies to study large and complex systems. In the path to faster and more generic computational quantum methods, three aspects are the most significant: (1) computation of ERIs, theoretically an $O(N^4)$ process, although for large systems, the asymptotic scaling approaches $O(N^2)$, (2) diagonalization, itself an $O(N^3)$ process, and (3) the SCF iterations. The focus of this work was on the efficient and accurate computation of ERIs. The approach consisted in using multivariate approximation techniques to reproduce pre-computed target data. It is a proof-of-concept work aimed at demonstrating the feasibility of such approximations. To my best knowledge, this was the first time that such techniques have been published.

The test system was the three-center ERI $(p_{xA}p_{xB}|p_{xC}p_{xC})$ with all atoms being carbons. Having all atoms the same, introduced important symmetry relations that reduced the amount of data required to develop the approximations. In the initial phase of development, when multiple calculations were needed to generate adequate target data, it was important to keep the number of calculations to a minimum. The same holds for the target basis set. The small STO-6G basis set was used because it allowed efficient calculation of the many ERIs required as target data.

The methodology for the numerical approximation consisted in decomposing a six-variable problem, and a three-variable problem, into three-bivariate problems, and one-univariate plus one-bivariate problem, respectively. The chosen approximating functions were bivariate Chebyshev polynomials and a univariate rational polynomial. The assumption was that for each sequential variable reduction, the approximating coefficients yield a continuous function that can be approximated by another set of polynomial approximants. If the approximating coefficients of a certain layer span a continuous surface, they can be fitted in the next layer of approximations using the same techniques. Although no mathematical justification was attempted, it was indeed verified that all surfaces and the single curve are continuous and could be accurately approximated. It is important to remember that the novel methodology to approximate ERIs is not general and is not intended to replace existing methodologies.

The results are excellent with reasonably small errors. In plots of residuals for two specific points, the best approximant gives absolute errors significantly less than $1.0E-05$, for most of the approximating domains. This means that the approximating methodology is able to maintain the rotational invariance of the computed integrals. Importantly, the new approach does not depend on the size of the contractions of the basis set. Although it was not a priority of this work, and no special attempts were made to optimize the speed of the numerical approximations, the methodology is very fast. Without specific optimization of the different Chebyshev expansions, the numerical approximation of the ERIs is 12.6 times faster than the calculation of the same ERIs using Rys polynomials as implemented in Uquantchem [16]. Considerable speed gains are expected from using optimized numerical libraries and GPU computing.

The first major development in the future will be the creation of a library of approximated ERIs, starting with the most common elements in Biology: H, C, N, O, S, Na, K, Cl, Fe, Zn, Cu, and Ni. Appropriate single- and double-ζ basis sets will be developed for these elements, and a new tight-binding approach will be implemented to take advantage of this work.

In conclusion, this work opens new perspectives to the future of computational chemistry, for example in molecular simulations of large and complex systems. The efficient computation of ERIs eliminates a significant barrier to the generalization of computational quantum methods to large systems and will allow the development of specialized quantum-based methodologies that will be simultaneously fast and accurate.

References

1. Dirac PAM (1929) Quantum mechanics of many-electron systems. Proc R Soc Lond A 123:714–733
2. Lopes PEM, Huang J, Shim J, Luo Y, Li H et al (2013) Polarizable force field for peptides and proteins based on the classical drude oscillator. J Chem Theory Comput 9:5430–5449
3. Thomson AJ, Gray HB (1998) Bio-inorganic chemistry. Curr Opin Chem Biol 2:155–158
4. Boys SF (1950) Electronic wave functions. I. A general method of calculation for the stationary states of any molecular system. Proc R Soc Lond A 200:542
5. Dupuis M, Rys J, King HF (1976) Evaluation of molecular integrals over Gaussian basis functions. J Chem Phys 65:111–116
6. King HF, Dupuis M (1976) Numerical integration using Rys polynomials. J Comput Phys 21:144–165
7. McMurchie LE, Davidson ER (1978) One- and two-electron integrals over Cartesian Gaussian functions. J Comput Phys 26:218–231
8. Obara S, Saika A (1986) Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. J Chem Phys 84:3963–3974
9. Reine S, Helgaker T, Lindh R (2012) Multi-electron integrals. WIREs Comput Mol Sci 2:290–303
10. Clementi E, Davis DR (1966) Electronic structure of large molecular systems. J Comput Phys 1:223–244
11. Clementi E (1991) Modern techniques in computational chemistry: MOTECC-91. ESCOM, Leiden
12. Guseinov II, Mamedov BA (2006) Evaluation of the Boys function using analytical relations. J Math Chem 40:179–183
13. Weiss AKH, Ochsenfeld C (2015) A rigorous and optimized strategy for the evaluation of the Boys function kernel in molecular electronic structure theory. J Comput Chem 36:1390–1398
14. Powell MJD (1981) Approximation theory and methods. Cambridge University Press, Cambridge
15. Mason JC, Handscomb DC (2003) Chebyshev polynomials. Chapman & Hall, Boca Raton, FL
16. Souvatzis P (2014) Uquantchem: a versatile and easy to use quantum chemistry computational software. Comput Phys Commun 185:415–421