Bridging the gap between quantum Monte Carlo and F12-methods

Sambasiva Rao Chinnamsetty, Hongjun Luo and Wolfgang Hackbusch
Max-Planck-Institut für Mathematik in den Naturwissenschaften, Inselstr. 22-26, D-04103 Leipzig, Germany

Heinz-Jürgen Flad and André Uschmajew
Institut für Mathematik, Technische Universität Berlin
Straße des 17. Juni 137, D-10623 Berlin, Germany

October 27, 2011

Abstract

Tensor product approximation of pair-correlation functions opens a new route from quantum Monte Carlo (QMC) to explicitly correlated F12 methods. Thereby one benefits from stochastic optimization techniques used in QMC to get optimal pair-correlation functions which typically recover more than 85 % of the total correlation energy. Our approach incorporates, in particular, core and core-valence correlation which are poorly described by homogeneous and isotropic ansatz functions usually applied in F12 calculations. We demonstrate the performance of the tensor product approximation by applications to atoms and small molecules. It turns out that the canonical tensor format is especially suitable for the efficient computation of two- and three-electron integrals required by explicitly correlated methods. The algorithm uses a decomposition of three-electron integrals, originally introduced by Boys and Handy and further elaborated by Ten-no in his 3d numerical quadrature scheme, which enables efficient computations in the tensor format. Furthermore, our method includes the adaptive wavelet approximation of tensor components where convergence rates are given in the framework of best N-term approximation theory.

1 Introduction

From the early days of quantum theory, it was a major goal to obtain highly accurate approximations for solutions of the many-particle Schrödinger equation. Focusing on the helium atom and isoelectronic ions, Hylleraas pioneering work [1] demonstrated the reachability of this goal. For technical reasons, however, his ansatz was not immediately applicable to more complicated atoms or molecules. In the subsequent development of quantum many-particle theory, two different lines of research have been followed. In quantum chemistry systematic expansions in terms of linear combinations of Slater determinants represented the favoured approach. Whereas research in nuclear and condensed matter physics also dealt with Jastrow-type wavefunctions [2] where correlation is described by a symmetric exponential ansatz function which explicitly depends on the inter-particle distances. Sophisticated many-particle theories [3, 4, 5] were developed to deal with the corresponding highly nonlinear optimization problem. It is interesting to see how both approaches merged to a certain extend in the framework of coupled cluster theory [6]. An alternative procedure represent stochastic optimization schemes for Jastrow-type wavefunctions [7, 8, 9, 10, 11, 12] which have been developed within the *quantum Monte Carlo* (QMC) approach. This highly successful method has been applied to atoms, molecules and solids [13, 14]. In the course of these studies, considerable knowledge concerning construction and performance of Jastrow factors [15, 16] has been accumulated.

The major bottleneck for highly accurate calculations using a basis of Slater determinants is the slow convergence with respect to the underlying single particle basis. To overcome this problem, additional

basis functions explicitly depending on inter-particle distances were also introduced in quantum chemistry. However, it was not until the work of Kutzelnigg and Klopper [17, 18] that this approach became manageable for larger molecules. Originally based on a simple linear ansatz in the inter-particle distance, the R12 method was recently extended to F12 methods [19, 20, 21, 22, 23, 24] by using more general ansatz functions with improved behaviour at large inter-particle distances. So far not much impact from QMC calculations on the construction of these correlation functions has been observed. This can be attributed to rather different underlying philosophies in both communities. Quantum chemists consider these additional correlation functions mainly as a completion of the original basis of Slater determinants which only contributes in the asymptotic region of the electron-electron cusp. In contrast to this, the Jastrow factor in QMC calculations typically acts on a Hartree-Fock (HF) wavefunction and carries over the whole burden of electron correlation in the short, intermediate and long-range regime. Recent studies using the F12 method [25], however, revealed the large potential of fairly simple optimized correlation functions also in this context.

It is the purpose of the present work to bridge the gap between QMC and F12 methods. Before we discuss the details of our approach let us briefly comment on similarities and differences of both methods. The common object is the pair-correlation function $\tau^{(2)}$ which enters into QMC via the Jastrow ansatz for the wavefunction, i.e., $\Psi = \mathcal{F}\Phi$, with Jastrow factor

$$\mathcal{F}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \exp\left(\sum_i \tau^{(1)}(\mathbf{x}_i) + \sum_{i<j} \tau^{(2)}(\mathbf{x}_i, \mathbf{x}_j) + \cdots\right), \tag{1.1}$$

and in F12 methods as additional basis functions for electron pairs

$$w_{ij}(\mathbf{x}_1, \mathbf{x}_2) := (1 - Q_1)(1 - Q_2)\tau^{(2)}(\mathbf{x}_1, \mathbf{x}_2)\phi_i(\mathbf{x}_1)\phi_j(\mathbf{x}_2), \tag{1.2}$$

with indices i, j running over all occupied orbitals. Here the projection operators Q_1 and Q_2 map into the space spanned by the occupied orbitals. The essential difference is that F12 methods restrict to pair-correlation functions which depend only on the inter-electron distance. The consequence of such a restriction can be well seen from QMC calculations, in particular the systematic study of Schmidt and Moskowitz [26] demonstrated the significance of additional degrees of freedom. Unfortunately, it is difficult to compare their results with recent F12 calculations based on optimized pair-correlation functions [25] because the latter consider only valence correlation whereas QMC inherently includes core and core-valence correlation. It is not surprising that often excellent results for valence correlation can be achieved with optimized F12 pair-correlation functions because the valence electron density is to a certain extent homogeneous and isotropic as well. Therefore, it seems that a real benefit of this more general approach can be expected in cases where core and core-valence correlation must be taken into account or when low lying valence orbitals, like for transition metals, are considered. This makes sense because such correlations require substantial extensions of standard basis sets.

2 Route from quantum Monte Carlo to F12 methods

Various explicit representations for Jastrow factors have been reported in the QMC literature. Within the present work we focus on the ansatz of Schmidt and Moskowitz [26] which was actually introduced earlier by Boys and Handy for their transcorrelated approach [27]. Presently, these Jastrow factors are the most popular for QMC calculations. The pair-correlation function is represented by a polynomial

$$\tau(\mathbf{x}_i, \mathbf{x}_j) = \sum_{I,J} \sum_{l,m,n} c_{lmn}^{I,J} (\bar{r}_{iI}^m \bar{r}_{jJ}^n + \bar{r}_{jI}^m \bar{r}_{iJ}^n) \bar{r}_{ij}^l, \tag{2.1}$$

in rational distance variables

$$\bar{r}_{iI} = \frac{|\mathbf{x}_i - \mathbf{A}_I|}{1 + |\mathbf{x}_i - \mathbf{A}_I|} \quad \text{and } \bar{r}_{ij} = \frac{|\mathbf{x}_i - \mathbf{x}_j|}{1 + |\mathbf{x}_i - \mathbf{x}_j|}. \tag{2.2}$$

Here and in the following, we have skipped the superscript in $\tau^{(2)}$ for notational simplicity. Powers of these variables can be further approximated by sums of Gaussians

$$\left[\frac{|\mathbf{x}|}{1+|\mathbf{x}|}\right]^m \approx \sum_{k=1}^{\kappa} b_{m,k} e^{-\alpha_{m,k}|\mathbf{x}|^2} \quad \text{for } m = 1, 2, \ldots, \tag{2.3}$$

leading to a separable approximation $\tilde{\tau}$ of the pair-correlation function, i.e.,

$$\tau(\mathbf{x},\mathbf{y}) \approx \tilde{\tau}(\mathbf{x},\mathbf{y}) := \sum_{I,J} \sum_{l,m,n} a_{lmn}^{I,J} \prod_{i=1}^{3} e^{-\alpha_l (x_i - A_{I,i})^2} e^{-\beta_m (x_i - y_i)^2} e^{-\gamma_n (y_i - A_{J,i})^2}, \tag{2.4}$$

with $\mathbf{x} := (x_1, x_2, x_3)$, $\mathbf{y} := (y_1, y_2, y_3)$, respectively. This separable approximation, however, has by construction a large separation rank and is of limited use for the computation of three-electron integrals. Therefore, a compression step is required which provides us with a low-rank approximation of the pair-correlation function. It is our goal to find the best possible separable approximation for a given separation rank κ in the general form

$$\tilde{\tau}(\mathbf{x},\mathbf{y}) \approx \sum_{k=1}^{\kappa} u_k^{(1)}(x_1, y_1) u_k^{(2)}(x_2, y_2) u_k^{(3)}(x_3, y_3), \tag{2.5}$$

which is known in the literature as the canonical tensor format [28]. Like in our previous work [29, 30], we applied Newton's algorithm [31] to solve the least-squares problem

$$\sigma_\kappa(\tilde{\tau}) := \inf_{w_k^{(i)}} \left\| e^{-\mu(|\mathbf{x}|^2+|\mathbf{y}|^2)} \left(\tilde{\tau}(\mathbf{x},\mathbf{y}) - \sum_{k=1}^{\kappa} \prod_{i=1}^{3} w_k^{(i)}(x_i, y_i) \right) \right\|_{L_2(\mathbb{R}^6)} \tag{2.6}$$

for the pair-correlation function in the weighted $L_2(\mathbb{R}^6)$ norm. The weight $e^{-\mu(\mathbf{x}^2+\mathbf{y}^2)}$ has been introduced for computational convenience because the asymptotic behaviour of the pair-correlation function is not well defined for $|\mathbf{x}|$ or $|\mathbf{y}| \to \infty$. It has been shown by Espig [31] that the best approximation

$$\{u_k^{(i)}(x_i, y_i)\}_{k=1,\ldots,\kappa,\ i=1,2,3} := \arg\min \sigma_\kappa(\tilde{\tau}) \tag{2.7}$$

is contained in the $L_2(\mathbb{R}^2)$ subspaces, cf. (2.4),

$$U_i := \operatorname{span}\left\{ e^{-\alpha_l (x_i - A_{I,i})^2} e^{-\beta_m (x_i - y_i)^2} e^{-\gamma_n (y_i - A_{J,i})^2} \right\}, \tag{2.8}$$

which means that the bivariate components $u_k^{(i)}$ are given in the form

$$u_k^{(i)}(x_i, y_i) = \sum_{I,J} \sum_{l,m,n} c_{k,lmn}^{I,J} e^{-\alpha_l (x_i - A_{I,i})^2} e^{-\beta_m (x_i - y_i)^2} e^{-\gamma_n (y_i - A_{J,i})^2}. \tag{2.9}$$

The computational complexity of the compression algorithm for a target rank κ is given by

$$\mathcal{O}\left(\kappa(\kappa + K) + \kappa^3 + (\kappa + K) \sum_{i=1}^{3} \dim U_i \right), \tag{2.10}$$

where K denotes the initial separation rank of $\tilde{\tau}$. The basic assumption for our applications is $\kappa \ll K$, and therefore it is the last term which eventually determines the computational complexity. The last term actually corresponds to a precomputing step performed before the minimization of the functional (2.6) using Newton's algorithm. By construction, the number of 2d-Gaussian geminals whose span defines U_i equals K, however, among these geminals a multitude of almost linear dependencies have to be expected.

In order to achieve optimal performance in the compression step, it is important to use an appropriate norm for the least-squares fit. For the pair-correlation function, its contribution to the kinetic energy is decisive. The weak formulation of the kinetic energy requires the first partial derivatives $\partial_x \tau$ which are not part of the L_2 error. Therefore, the Sobolev space H^1 which also takes into account the L_2 error of first partial derivatives, is appropriate. It is not hard to see that the original pair-correlation function τ belongs to $H^1(\Omega)$ for any bounded domain $\Omega \subset \mathbb{R}^6$. Again the boundedness of Ω is required just because we want to avoid any explicit statement concerning the asymptotic behaviour of τ in the limit $|\mathbf{x}|$ or $|\mathbf{y}| \to \infty$. Recently, one of us proved that the best canonical tensor product approximation in the L_2 sense preserves the Sobolev regularity of the function to be approximated [32]. It is consistent with this regularity result to consider a H^1_{mix} norm for the approximation of $\tilde{\tau}$, already in a separable format, which explicitly depends on first partial derivatives $\partial_x \tau$. For a function given in a separable representation, the H^1_{mix} norm is defined as

$$\| \sum_{k=1}^{\kappa} \prod_{i=1}^{3} w_k^{(i)} \|^2_{H^1_{mix}} := \sum_{k,\tilde{k}} \prod_{i=1}^{3} \Big(\langle w_k^{(i)}, w_{\tilde{k}}^{(i)} \rangle + \lambda \langle \partial_{x_i} w_k^{(i)}, \partial_{x_i} w_{\tilde{k}}^{(i)} \rangle + \lambda \langle \partial_{y_i} w_k^{(i)}, \partial_{y_i} w_{\tilde{k}}^{(i)} \rangle \Big), \tag{2.11}$$

where

$$\langle w_k^{(i)}, w_{\tilde{k}}^{(i)} \rangle := \int_{\mathbb{R}^2} w_k^{(i)}(x_i, y_i) w_{\tilde{k}}^{(i)}(x_i, y_i)\, dx_i dy_i, \tag{2.12}$$

and

$$\langle \partial_{x_i} w_k^{(i)}, \partial_{x_i} w_{\tilde{k}}^{(i)} \rangle := \int_{\mathbb{R}^2} \partial_{x_i} w_k^{(i)}(x_i, y_i) \partial_{x_i} w_{\tilde{k}}^{(i)}(x_i, y_i)\, dx_i dy_i. \tag{2.13}$$

In our applications, we have chosen $\lambda = 1/2$. A major advantage of this definition is that it requires no modifications of the Newton algorithm which is not the case if we would insist on optimization in the H^1 norm.

In order to demonstrate the performance of the compression step and the accuracy of corresponding canonical tensor product approximations, we have studied QMC optimized pair-correlation functions for some simple atoms and molecules. In Fig. 2.1, compression errors at various ranks and corresponding *variational Monte Carlo* (VMC) energies are shown for He, Ne, CH_4 and H_2O. It can be seen that the VMC energy of the original pair-correlation functions are recovered at separation ranks between 40 and 50. The H^1_{mix} error converges exponentially with the separation rank. For orientation of the reader, fixed-node diffusion Monte Carlo (DMC) energies and exact ground state energies are shown. The DMC energies can be actually considered as the real benchmark for the Jastrow ansatz which is not exact because of the fixed-node approximation.

3 Tensor product approximation of three electron integrals

The tensor product approximation of stochastically optimized pair-correlation functions paves the way for their application in explicitly correlated methods. Because of the anisotropic and inhomogeneous character of the ansatz (2.1), a direct computation of the required two- and three-electron integrals would be hard to accomplish. Via the separable tensor product representation, however, the computational complexity of these integrals can be drastically reduced by appropriate modifications of already existing numerical intergration schemes. In this section we want to outline such an approach and present some preliminary numerical results.

For explicitly correlated methods, three-electron integrals represent a major bottleneck [24] and the ultimate success of these methods is closely linked to the development of efficient and accurate approximation schemes. In this section we want to discuss a tensor product based approach for the computation of three-electron integrals. The canonical tensor product approximation of the pair-correlation function

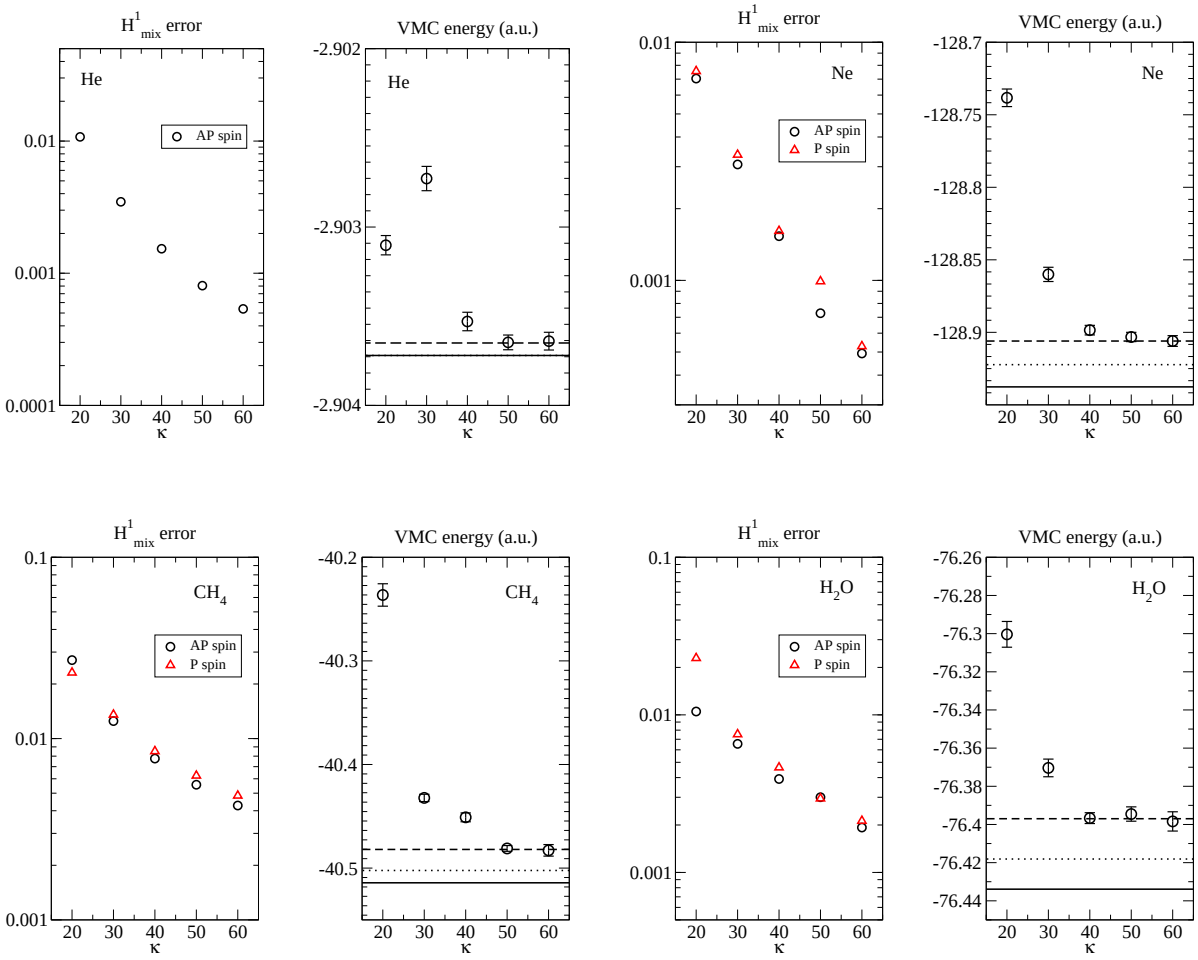

Figure 2.1: H^1_{mix} compression errors and corresponding VMC energies versus separation rank κ of canonical tensor product approximations of pair-correlation functions for He, Ne, CH_4 and H_2O. For reference, VMC energies before compression (dashed line), DMC energies (dotted line) and exact energies (solid line) are given.

(2.5) leads to bivariate components of the form (2.9) which can be used for the computation of two- and three-electron integrals. As an illustrative example, we want to discuss a specific type of three-electron integral

$$\langle ij||\mathbf{x}_1 - \mathbf{x}_2|^{-1} Q_2 \tau(\mathbf{x}_1, \mathbf{x}_2)|kl\rangle = \sum_m \langle ijm||\mathbf{x}_1 - \mathbf{x}_2|^{-1} \tau(\mathbf{x}_1, \mathbf{x}_3)|kml\rangle, \tag{3.1}$$

where the indices i, j, k, l, m run over all occupied orbitals. Such integrals are required, e.g., for the explicitly correlated MP2 method. Our tensor product based approach can be considered as a variant of Ten-no's numerical quadrature scheme [33] which was originally introduced by Boys and Handy in their transcorrelated approach [34]. We take the same decomposition of three-electron integrals, but instead of a 3d numerical quadrature, tensor product approximations are used. The basic idea is to obtain low-rank approximations for convolutions[1] of orbital products with the Coulomb potential and pair-correlation function from which three-electron integrals are easily computed. To illustrate our concept, let us consider

[1]Here it should be mentioned that the notion of convolution has been used in a slightly generalized sense in order to apply it to pair-correlation functions, not only depending on the inter-electron distance, as well.

the integral (3.1) more closely. This integral can be decomposed into convolutions according to

$$\langle ijm||\mathbf{x}_1 - \mathbf{x}_2|^{-1}\tau(\mathbf{x}_1,\mathbf{x}_3)|kml\rangle = \int_{\mathbb{R}^3} A_{jm}(\mathbf{x}_1)\phi_i(\mathbf{x}_1)\phi_k(\mathbf{x}_1)B_{ml}(\mathbf{x}_1)d^3x_1, \tag{3.2}$$

with

$$A_{jm}(\mathbf{x}_1) := \int_{\mathbb{R}^3} |\mathbf{x}_1 - \mathbf{x}_2|^{-1}\phi_j(\mathbf{x}_2)\phi_m(\mathbf{x}_2)d^3x_2, \tag{3.3}$$

$$B_{ml}(\mathbf{x}_1) := \int_{\mathbb{R}^3} \tau(\mathbf{x}_1,\mathbf{x}_3)\phi_m(\mathbf{x}_3)\phi_l(\mathbf{x}_3)d^3x_3. \tag{3.4}$$

Ten-no's approach consists of approximating the integral (3.2) by a 3d numerical quadrature. Instead of this, we are aiming at low-rank tensor product approximations for the intermediate quantities A_{jm} and B_{ml}, thereby Eq. (3.2) factorizes into sums of products of 1d-integrals. Our algorithm consists of a succession of product evaluations, convolutions and compressions. In detail, we consider the following steps:

(i) *Tensor product approximation of products of orbitals.* For each pair of occupied orbitals, we compute a low-rank approximation of their product

$$\phi_i(\mathbf{x})\phi_j(\mathbf{x}) \approx \sum_{k=1}^{\kappa'} \varphi^{(1)}_{ij,k}(x_1)\,\varphi^{(2)}_{ij,k}(x_2)\,\varphi^{(3)}_{ij,k}(x_3). \tag{3.5}$$

A detailed discussion of this topic with an application to HF exchange integrals has been given in [30, 35].

(ii) *Convolution with Coulomb potential and pair-correlation function.* In order to get the intermediate quantities A_{jm} and B_{ml}, convolutions of products of orbitals with the Coulomb potential and pair-correlation function must be computed. This can be done in the canonical tensor format in a very efficient manner, e.g.,

$$B_{ml}(\mathbf{x}) = \sum_{k,k'=1}^{\kappa,\kappa'} \prod_{i=1}^{3} \int_{\mathbb{R}} u^{(i)}_k(x_i,y_i)\varphi^{(i)}_{ml,k'}(y_i)dy_i. \tag{3.6}$$

(iii) *Compression of intermediate quantities A_{jm} and B_{ml}.* It is a typical feature of the tensor product approach that operations on a tensor product increase its rank and a subsequent compression step is mandatory in order to avoid an undue increase of the computational complexity. The convoluted quantities are therefore further approximated by low-rank tensor products, e.g.,

$$B_{ml}(\mathbf{x}) \approx \sum_{k=1}^{\kappa} b^{(1)}_{ml,k}(x_1)\,b^{(2)}_{ml,k}(x_2)\,b^{(3)}_{ml,k}(x_3), \tag{3.7}$$

which can be stored at low costs for each pair of occupied orbitals. A similar approach has been outlined in [29] for the computation of standard two-electron integrals.

(iv) *Intermediate sum and further compression.* In explicitly correlated methods it is only the sum of three-electron integrals (3.1) which enters into the calculation. With the low-rank approximations of the previous step at hand, the sum can be computed in the tensor format

$$\begin{aligned} D_{jl}(\mathbf{x}) &:= \sum_m A_{jm}(\mathbf{x})B_{ml}(\mathbf{x}) \\ &= \sum_m \sum_{k,k'=1}^{\kappa,\kappa'} \prod_{i=1}^{3} a^{(i)}_{jm,k}(x_i)b^{(i)}_{ml,k'}(x_i), \end{aligned} \tag{3.8}$$

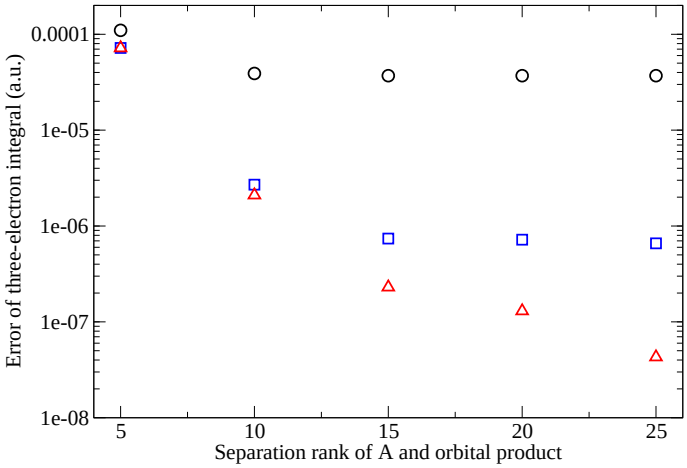

Figure 3.1: Error in the "three-electron" integral (3.1) of the helium atom for various separation ranks of intermediate quantities. The same separation rank, depicted at the abscissa, has been taken for the orbital product and its convolution with the Coulomb potential A. For the convolution with the pair-correlation function B, three different separation ranks $\kappa = 40$ (◯), 60 (□), 80 (△) have been chosen. A 8s GTO basis was taken for the 1s orbital.

and a subsequent compression step provides an efficient approximation of this intermediate quantity

$$D_{jl}(\mathbf{x}) \approx \sum_{k=1}^{\kappa} d_{jl,k}^{(1)}(x_1)\, d_{jl,k}^{(2)}(x_2)\, d_{jl,k}^{(3)}(x_3). \tag{3.9}$$

(v) *Computation of the integrals.* In the last step, three-electron integrals are easily computed in the tensor format, i.e.,

$$\sum_m \langle ijm||\mathbf{x}_1 - \mathbf{x}_2|^{-1}\tau(\mathbf{x}_1,\mathbf{x}_3)|kml\rangle = \sum_{k',k''=1}^{\kappa,\kappa'} \prod_{p=1}^{3} \int_{\mathbb{R}} d_{jl,k'}^{(p)}(x_p)\varphi_{ik,k''}^{(p)}(x_p)dx_p. \tag{3.10}$$

We want to emphasize that all intermediate quantities and corresponding compression steps refer to two index quantities with respect occupied orbitals.

In order to demonstrate the feasibility of our approach, we computed the "three-electron" integral (3.1) for the helium atom. The error at various separation ranks of intermediate quantities is shown in Fig. 3.1, where it can be seen that the separation rank of the convolution of the pair-correlation function is the most sensible quantity. This simple example indicates that with moderate separation ranks of the intermediate quantities, it is possible to achieve sufficiently high accuracies. For more complicated atoms and molecules we expect no dramatic increase of the separation ranks. Our expectation is based on our previous work [29, 30] where we have studied tensor product approximations of orbital products and Hartree potentials for some small molecules.

4 Best N-term approximation theory for tensor components

In Sections 2 and 3, we discussed our approach to obtain a low-rank tensor product approximation for the pair-correlation function and how to use it for the computation of three-electron integrals. So far our discussion, however, leaves an important point open, namely how to represent the uni- and bivariate

tensor components. This can be done e.g. by equidistant or adaptive grids as well as by some kind of basis functions. The tensor components reflect electron-nuclear and inter-electron cusps and one cannot expect a higher regularity beyond the result in [32]. Consequently, uniform representations, like equidistant grids, become highly inefficient. Instead adaptive schemes are required which enable data sparse approximations after an additional compression step. For this purpose, we have already studied wavelet approximations of tensor components for the electron density and the Hartree potential [29] where good compression rates have been achieved. It is the purpose of this section to put these heuristic observations on solid mathematical ground.

The following discussion is of practical significance with respect to the construction of adaptive wavelet algorithms for the computation of intermediate quantities in the tensor format, cf. [29]. In order to obtain an algorithm with optimal computational complexity it is essential to achieve the best possible benefit to cost ratio with respect to approximation order and size of support of a wavelet. We provide the order of approximation for different wavelet bases in an adaptive setting, where it turns out that uni- and bivariate tensor components behave differently.

The concept of best N-term approximation belongs to the realm of nonlinear approximation theory. For a detailed exposition of this subject, we refer to Ref. [36]. Loosely speaking, one considers for a given basis $\{\zeta_i : i \in \Lambda\}$ the best possible approximation of a function f in the nonlinear subset Σ_N which consists of all possible linear combinations of at most N basis functions, i.e.,

$$\Sigma_N := \left\{ \sum_{i\in\Delta} c_i \zeta_i : \Delta \subset \Lambda, \#\Delta \leq N,\ c_i \in \mathbb{R} \right\}. \tag{4.1}$$

Here, the approximation error

$$\sigma_N(f) := \inf_{f_N \in \Sigma_N} \|f - f_N\|_H \tag{4.2}$$

is given with respect to the norm of an appropriate Hilbert space H. Best N-term approximation spaces $A_q^\alpha(H)$ for a Hilbert space H can be defined according to

$$A_q^\alpha(H) := \{f \in H : |f|_{A_q^\alpha(H)} < \infty\} \quad \text{with } |f|_{A_q^\alpha(H)} := \left(\sum_{N\in\mathbb{N}} \left(N^\alpha \sigma_N(f) \right)^q N^{-1} \right)^{\frac{1}{q}}. \tag{4.3}$$

It follows from the definition that the error $\sigma_N(f)$ decreases asymptotically like $\mathcal{O}(N^{-\alpha})$ with respect to the number of basis functions. In fact $\sigma_N(f)$ is the best error bound which can be obtained with N degrees of freedom. For a complete orthonormal basis, the best N-term approximation in L_2 is simply given by the N largest coefficients.

In our applications we consider best N-term approximation for wavelet bases. For the ease of the reader, some basic facts about wavelets are given in Appendix A. In the following only orthogonal wavelet bases are considered, however, our considerations can be easily generalized to biorthogonal wavelets. Furthermore, in our notation, we do not distinguish between scaling functions and wavelets on the coarsest level. Such a distinction is not relevant in the following and should cause no confusion. The multiscale best N-term approximation of a univariate function in a wavelet basis is given by

$$f(x) \approx f_N(x) = \sum_{(j,a)\in\Delta} \langle \psi_{j,a} | f \rangle \psi_{j,a}(x), \quad \text{with } \#\Delta \leq N. \tag{4.4}$$

where the index set Δ has been chosen in order to minimize the error $\sigma_N(f)$. Since we are mainly interested in the energy of a molecule it is appropriate to restrict to a finite domain. Therefore, we consider in the following the Sobolev space $H^1(\mathcal{I})$ for a bounded interval $\mathcal{I} \subset \mathbb{R}$. The corresponding approximation spaces $A_q^\alpha(H^1(\mathcal{I}))$ have been identified as Besov spaces [36] which have an equivalent norm for $\frac{1}{q} = \alpha + \frac{1}{2}$ in terms of weighted ℓ_q spaces of wavelet coefficients, i.e.,

$$\| f \|^q_{\mathcal{A}_q^\alpha(\mathcal{I})} \equiv \sum_{j\geq j_0} 2^{jq} \left(\sum_a |\langle \psi_{j,a} | f \rangle|^q \right). \tag{4.5}$$

This norm equivalence is satisfied for all wavelet bases with $p > \alpha + 1$ vanishing moments and for which ψ belongs to an appropriate Besov space. By estimating wavelet coefficients for a given function f, one can try to show that the sum (4.5) is finite which proves that f belongs to $A_q^\alpha(H^1(\mathcal{I}))$.

Best N-term approximation can be also applied to bivariate tensor components of pair-correlation functions. For this we consider anisotropic tensor product wavelets

$$\psi_{j_1,a_1}(x)\,\psi_{j_2,a_2}(y), \tag{4.6}$$

and corresponding best N-term approximation for bivariate functions

$$f(x,y) \approx f_N(x,y) = \sum_{(j_1,j_2,a_1,a_2)\in\Delta} \langle \psi_{j_1,a_1}\,\psi_{j_2,a_2}|f\rangle \psi_{j_1,a_1}(x)\,\psi_{j_2,a_2}(y) \quad \text{with } \#\Delta \le N. \tag{4.7}$$

Similar to the univariate case, we consider approximation spaces $A_q^\alpha(H^1(\Omega))$ for a square $\Omega := \mathcal{I}\times\mathcal{I} \subset \mathbb{R}^2$. The necessary generalization of N-term approximation theory to anisotropic tensor product wavelets has been given by Nitsche [37]. Here we just want to mention that the equivalent weighted ℓ_q norm for $\frac{1}{q} = \alpha + \frac{1}{2}$ becomes

$$\| f \|^q_{\mathcal{A}_q^\alpha(\Omega)} \equiv \sum_{j_1,j_2\ge j_0} 2^{\max\{j_1,j_2\}q} \left(\sum_{a_1,a_2} |\langle \psi_{j_1,a_1}\,\psi_{j_2,a_2}|f\rangle|^q \right). \tag{4.8}$$

4.1 Univariate components of tensor products

In the course of our tensor algorithm for the computation of three-electron integrals outlined in Section 3, we encountered several intermediate quantities (3.5), (3.7) and (3.9) which are represented by tensor product approximations. Any computation using these tensor products greatly benefits from sparse approximations of the corresponding univariate tensor components.

Theorem 1. *Let the function η correspond to a solution ϕ_i of the Hartree-Fock equation, a pointwise product $\phi_i\phi_j$, its convolution with the Coulomb potential A_{lm} or pair-correlation function B_{lm}. The directional univariate components $\eta_k^{(i)}$, for $i = 1,2,3$ and $k = 1,\dots,\kappa$, of their canonical best separable tensor product approximations, i.e., local minimizers*[2]

$$\sum_{k=1}^{\kappa} \eta_k^{(1)} \otimes \eta_k^{(2)} \otimes \eta_k^{(3)} = \arg\min_{\{h_k^{(i)}\}} \left\| \eta - \sum_{k=1}^{\kappa} h_k^{(1)} \otimes h_k^{(2)} \otimes h_k^{(3)} \right\|_{L_2(\mathbb{R}^3)}, \tag{4.9}$$

belong to best N-term approximation spaces $A_q^\alpha(H^1(\mathcal{I}))$ for all $\alpha > 0$ and $\frac{1}{q} = \alpha + \frac{1}{2}$.

Proof. In the following, we always assume that $\operatorname{rank}\eta > \kappa$, i.e., the actual tensor rank of the function η to be approximated is larger than the separation rank κ of its approximation. This is generally satisfied in our applications.

According to a regularity result for univariate components of tensor products [32], the univariate components $\eta_k^{(i)}$, $i = 1,2,3$, $k = 1,\dots,\kappa$ of a local minimizer inherit the regularity of the function η, that is, if $\eta \in H^s(\mathbb{R}^3)$ then $\eta_k^{(i)}$ belongs to $H^s(\mathbb{R})$. It is known that HF orbitals and their products belong to Sobolev spaces $H^s(\mathbb{R}^3)$ for $s < 5/2$ [38]. A convolution with the Coulomb potential or pair-correlation function increases the Sobolev regularity further. Therefore, we assume in the following $\eta \in H^s(\mathbb{R}^3)$ for $s < 5/2$.

It can be seen from the proof in [32] that the univariate components satisfy an integral equation

$$\eta_k^{(1)}(x_1) = \int_{\mathbb{R}^2} \eta(\mathbf{x}) w_k(x_2,x_3)\,dx_2dx_3 \tag{4.10}$$

[2]For convolutions the L_2 norm is taken over a large finite cube $\mathcal{Q}$ containing $\mathcal{I}\times\mathcal{I}\times\mathcal{I}$ as a subcube, instead of $\mathbb{R}^3$.

with $w_k \in L_2(\mathbb{R}^2)$ and $\eta_k^{(1)} \in H^s(\mathbb{R})$ for $s < 5/2$. For the following estimates of the wavelet coefficients, let us consider w.l.o.g. the component in the x_1 direction and assume a single nucleus located at the origin. According to (4.10), a coefficient is given by

$$\langle \psi_{j,a}, \eta_k^{(1)} \rangle := \int_{\mathbb{R}} \psi_{j,a}(x_1) \eta_k^{(1)}(x_1) dx_1 = \int_{\mathbb{R}^3} \psi_{j,a}(x_1) \eta(\mathbf{x}) w_k(x_2, x_3) d^3x. \tag{4.11}$$

For wavelets with support close to the origin, i.e., $\operatorname{dist}(\operatorname{supp} \psi_{j,a}, 0) \lesssim 2^{-j}$, we get from (A.7) the estimate

$$\left| \langle \psi_{j,a}, \eta_k^{(1)} \rangle \right| \lesssim 2^{-\frac{3}{2}j} \|\partial_{x_1} \eta_k^{(1)}\|_{L_\infty(\operatorname{supp} \psi_{j,a})} \tag{4.12}$$

because of the embedding $\eta_k^{(1)} \in H^{\frac{5}{2}-\epsilon}(\mathbb{R}) \hookrightarrow C_B^1(\mathbb{R})$, for $0 < \epsilon < \frac{1}{2}$, into the space of functions with bounded weak derivative. The number of such wavelet coefficients is $O(1)$. Here and in the following $a \lesssim b$ means that $a \leq Cb$ for some constant C which is independent of variables or parameters on which a, b may depend on. Similarly $a \simeq b$ means that the quantities can be bounded by some multiple of each other.

Next let us consider $\operatorname{dist}(\operatorname{supp} \psi_{j,a}, 0) \gtrsim 2^{-j}$, in particular $\operatorname{dist}(\operatorname{supp} \psi_{j,a}, 0) \simeq 2^{-j}|a|$ can be assumed. In a neighbourhood of the nucleus, we can apply the asymptotic estimate

$$\left| \partial_{\mathbf{x}}^\beta \eta(\mathbf{x}) \right| \lesssim |\mathbf{x}|^{1-|\beta|} \qquad \text{for } \mathbf{x} \neq 0 \text{ and } |\beta| \geq 1, \tag{4.13}$$

which provides an upper bound on the divergence of partial derivatives. This estimate applies to orbitals, their products and convolutions, cf. [38] for further details. Here, we used the standard short-hand notation for mixed partial derivatives

$$\partial_{\mathbf{x}}^\beta := \frac{\partial^{\beta_1}}{\partial x_1^{\beta_1}} \frac{\partial^{\beta_2}}{\partial x_2^{\beta_2}} \frac{\partial^{\beta_3}}{\partial x_3^{\beta_3}}, \tag{4.14}$$

with $\beta_1, \beta_2, \beta_3 \in \mathbb{N}$ and absolute value of the multi-index $|\beta| := \beta_1 + \beta_2 + \beta_3$. Using (A.7) and Schwarz's inequality, corresponding wavelet coefficients can be estimated according to

$$\begin{aligned}
\left| \int_{\mathbb{R}} \psi_{j,a}(x_1) \eta_k^{(1)}(x_1) dx_1 \right| &\lesssim 2^{-(p+\frac{1}{2})j} \left\| \int_{\mathbb{R}^2} \partial_{x_1}^p \eta(\mathbf{x}) w_k(x_2, x_3) dx_2 dx_3 \right\|_{L_\infty(\operatorname{supp} \psi_{j,a})} \\
&\lesssim 2^{-(p+\frac{1}{2})j} \left\| \int_{\mathbb{R}^2} |\mathbf{x}|^{1-p} w_k(x_2, x_3) dx_2 dx_3 \right\|_{L_\infty(\operatorname{supp} \psi_{j,a})} \\
&\lesssim 2^{-(p+\frac{1}{2})j} \left(\int_0^\infty |2^{-2j} a^2 + r^2|^{1-p} r dr \right)^{\frac{1}{2}} \|w\|_{L_2} \\
&\lesssim 2^{-\frac{5}{2}j} |a|^{2-p},
\end{aligned} \tag{4.15}$$

where the estimate

$$\int_0^\infty |2^{-2j} a^2 + r^2|^{1-p} r dr \lesssim \left(2^{-2j} a^2\right)^{2-p}, \tag{4.16}$$

for $p > 2$ has been applied in the second last line.

The estimates for the wavelet coefficients (4.12) and (4.15) can be used to estimate the weighted ℓ_q norm (4.5), i.e.,

$$\begin{aligned}
\sum_{j \geq j_0} 2^{qj} \sum_a \left| \langle \psi_{j,a}, \eta_k^{(1)} \rangle \right|^q &\lesssim \sum_{j \geq j_0} 2^{qj} \left(2^{-\frac{3}{2}qj} + 2^{-\frac{5}{2}qj} \sum_{a \in \mathbb{Z} \setminus \{0\}} |a|^{(2-p)q} \right) \\
&\lesssim \sum_{j \geq j_0} 2^{-\frac{1}{2}qj} < \infty,
\end{aligned} \tag{4.17}$$

which applies for wavelets with $p > 2 + \frac{1}{q}$ vanishing moments. It should be mentioned that the norm equivalence (4.5) requires wavelets with more than $\frac{1}{2} + \frac{1}{q}$ vanishing moments. It follows that $\eta_k^{(1)}$ belongs to $A_q^\alpha(H^1(\mathcal{I}))$ or all $\alpha > 0$ and $\frac{1}{q} = \alpha + \frac{1}{2}$. □

As already mentioned before the optimization problem (4.9) is often ill-posed and global minimizers do not exist. It is however possible to generalize the Sobolev regularity theorem [32] to a well-posed optimization problem using Tikhonov regularization, which is frequently used in numerical algorithms.

4.2 Bivariate tensor components of pair-correlation functions

In order to study the best N-term approximation of bivariate tensor components of a pair-correlation function, it is necessary to estimate its singular behaviour along the diagonal. No rigorous general results are presently available, however, guided by Kato's cusp condition and the ansatz (2.1) used in QMC calculations we make the following assumption

$$|\partial_{x_i}^{\beta}\tau(\mathbf{x},\mathbf{y})| \lesssim |\mathbf{x}-\mathbf{y}|^{1-|\beta|} f(\mathbf{x},\mathbf{y}), \quad \text{for } |\beta| \geq 1 \text{ and } \mathbf{x} \neq \mathbf{y}, \tag{4.18}$$

concerning the singular behaviour of the pair-correlation function near the diagonal, where f belongs to the Schwartz space $\mathcal{S}(\mathbb{R}^3 \times \mathbb{R}^3)$ of smooth rapidly decreasing functions, e.g., $f(\mathbf{x},\mathbf{y}) = e^{-\mu(|\mathbf{x}|^2+|\mathbf{y}|^2)}$ is fine. Here, f is introduced for technical reasons only, to make the pair-correlation function well-behaved at infinity. For example, one can choose a smooth cut-off function which is adapted to the size of the molecule. In the ansatz (2.1), we assume that only even powers with respect to electron-nuclear distance variables (2.2) are taken into account which means that the pair-correlation function has no off-diagonal cusps at the nuclei. This assumption is actually satisfied in most applications, cf. [26], and a comparative study [16] showed that inclusion of odd powers gives only a minor improvement. An extension of our result which takes odd powers into account is possible, however, due to additional singularities at the nuclei a considerably more elaborate proof is required, cf. [39] for further details. It is for the sake of simplicity that we refrain from such complications in the present work.

After these preliminary remarks, the best N-term approximation spaces for bivariate tensor components can be stated.

Theorem 2. *Suppose a two-particle correlation function τ satisfies the estimate (4.18). The bivariate components $u_k^{(i)}(x_i,y_i)$, $i=1,2,3$, $k=1,\dots,\kappa$, of a canonical tensor product approximation, i.e., a local minimizer*

$$\sum_{k=1}^{\kappa} u_k^{(1)} \otimes u_k^{(2)} \otimes u_k^{(3)} = \arg\min_{\{w_k^{(i)}\}} \left\| \tau - \sum_{k=1}^{\kappa} w_k^{(1)} \otimes w_k^{(2)} \otimes w_k^{(3)} \right\|_{L^2(\mathbb{R}^6)},$$

belong to best N-term approximation spaces $A_q^{\alpha}(H^1(\Omega))$ with $0<\alpha<\frac{3}{2}$ and $\frac{1}{q}=\alpha+\frac{1}{2}$ for anisotropic tensor product wavelets (4.6).

In contrast to the univariate case, the bivariate tensor components do not belong to all best N-term approximation spaces. This is because of the singular behaviour along the diagonal which cannot be fully compensated by adaptivity. In our previous work [39], the best N-term approximation of pair-correlation functions has been studied directly without intermediate tensor product approximation. It was shown that the pair-correlation function τ in this case belongs to best N-term approximation spaces $A_q^{\alpha}(H^1)$ for all $0<\alpha<\frac{1}{2}$ and $\frac{1}{q}=\alpha+\frac{1}{2}$. Provided sufficiently accurate tensor product approximations can be obtained at low ranks, the gain is significant in particular for the energy which converges quadratically with respect to the H^1 error.

Proof. Similarly to the univariate case, the bivariate tensor components satisfy an integral equation, e.g.,

$$u_k^{(1)}(x_1,y_1) = \int_{\mathbb{R}^2\times\mathbb{R}^2} \tau(\mathbf{x},\mathbf{y}) w_k(\mathbf{x}_{23},\mathbf{y}_{23}) d^2x_{23} d^2y_{23}, \tag{4.19}$$

with $\mathbf{x}_{23} := (x_2,x_3)$, $\mathbf{y}_{23} := (y_2,y_3)$, where $w_k \in L_2(\mathbb{R}^2 \times \mathbb{R}^2)$ and $u_k^{(1)} \in H^s(\mathbb{R}\times\mathbb{R})$ for $s<5/2$. The Sobolev regularity follows from [32] and the fact that τ belongs to $H^s(\mathbb{R}^3\times\mathbb{R}^3)$ for $s<\frac{5}{2}$, cf. [40]. With

it, the wavelet coefficients are given by

$$\begin{aligned}\langle \psi_{j_1,a_1}\psi_{j_2,a_2}, u_k^{(1)}\rangle &:= \int_{\mathbb{R}\times\mathbb{R}} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)u_k^{(1)}(x_1,y_1)dx_1dy_1 \\ &= \int_{\mathbb{R}^3\times\mathbb{R}^3} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^3xd^3y. \end{aligned} \tag{4.20}$$

In the following estimates, we take $\mathcal{U}_{j,a} := \operatorname{supp}\psi_{j,a}$, $L := \operatorname{diam}\operatorname{supp}\psi$ and assume w.l.o.g. $j_1 \geq j_2$.

The first case we have to consider are tensor product wavelets with support close to the diagonal, i.e., $\operatorname{dist}\{\mathcal{U}_{j_1,a_1}\times\mathcal{U}_{j_2,a_2},\operatorname{diag}\} \leq 2^{-j_2}L$. Let us subdivide the support $\mathcal{U}_{j_2,a_2}$ of the wavelet ψ_{j_2,a_2} into subintervals $\mathcal{I}_i$, $i=1,\ldots,2^{j_1-j_2}$, of length $2^{-(j_1-j_2)}L$. Then, let us split up the integral according to

$$\langle \psi_{j_1,a_1}\psi_{j_2,a_2}, u_k^{(1)}\rangle = \sum_{i=1}^{2^{j_1-j_2}} \int_{\mathcal{I}_i}\int_{\mathcal{U}_{j_1,a_1}} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)u_k^{(1)}(x_1,y_1)dx_1dy_1. \tag{4.21}$$

Let us estimate the case $\operatorname{dist}\{\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i,\operatorname{diag}\}\leq 2^{-j_1}L$, i.e.,

$$\begin{aligned}&\int_{\mathcal{I}_i}\int_{\mathcal{U}_{j_1,a_1}} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)u_k^{(1)}(x_1,y_1)dx_1dy_1 \\ &\quad= \int_{\mathcal{I}_i}\int_{\mathcal{U}_{j_1,a_1}}\int_{\mathbb{R}^2\times\mathbb{R}^2} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^3xd^3y, \end{aligned} \tag{4.22}$$

after splitting up the inner integral

$$\int_{\mathbb{R}^2\times\mathbb{R}^2} = \int_{\mathcal{D}} + \int_{\mathbb{R}^2\times\mathbb{R}^2\setminus\mathcal{D}}, \tag{4.23}$$

with

$$\mathcal{D} := \big\{(\mathbf{x}_{23},\mathbf{y}_{23}) : |\mathbf{x}_{23}-\mathbf{y}_{23}|\leq 2^{-j_1}L\big\}, \tag{4.24}$$

estimates for the two parts can be obtained separately

$$\begin{aligned}&\left|\int_{\mathcal{I}_i}\int_{\mathcal{U}_{j_1,a_1}}\int_{\mathcal{D}} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^3xd^3y\right| \\ &\quad\lesssim 2^{-\frac{5}{2}j_1}2^{\frac{1}{2}j_2}\left\|\int_{\mathcal{D}}\partial_{x_1}\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^2x_{23}d^2y_{23}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)} \\ &\quad\lesssim 2^{-\frac{5}{2}j_1}2^{\frac{1}{2}j_2}\left\|\left(\int_{\mathcal{D}}|\partial_{x_1}\tau(\mathbf{x},\mathbf{y})|^2d^2x_{23}d^2y_{23}\right)^{\frac{1}{2}}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)}\|w\|_{L_2(\mathcal{D})} \\ &\quad\lesssim 2^{-\frac{7}{2}j_1}2^{\frac{1}{2}j_2}, \end{aligned} \tag{4.25}$$

and

$$\begin{aligned}&\left|\int_{\mathcal{I}_i}\int_{\mathcal{U}_{j_1,a_1}}\int_{\mathbb{R}^2\times\mathbb{R}^2\setminus\mathcal{D}} \psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^3xd^3y\right| \\ &\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\left\|\int_{\mathbb{R}^2\times\mathbb{R}^2\setminus\mathcal{D}}\partial_{x_1}^p\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^2x_{23}d^2y_{23}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)} \\ &\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\left\|\left(\int_{\mathbb{R}^2\times\mathbb{R}^2\setminus\mathcal{D}}|\partial_{x_1}^p\tau(\mathbf{x},\mathbf{y})|^2d^2x_{23}d^2y_{23}\right)^{\frac{1}{2}}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)}\|w\|_{L_2(\mathbb{R}^2\times\mathbb{R}^2\setminus\mathcal{D})} \\ &\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\left\|\left(\int_{\mathbb{R}^2\times\mathbb{R}^2\setminus\mathcal{D}}|\mathbf{x}_{23}-\mathbf{y}_{23}|^{2(1-p)}|f(\mathbf{x},\mathbf{y})|^2d^2x_{23}d^2y_{23}\right)^{\frac{1}{2}}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)} \\ &\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\left(\int_{2^{-j_1}}^{\infty}s^{2(1-p)}sds\right)^{\frac{1}{2}} \\ &\quad\lesssim 2^{-\frac{7}{2}j_1}2^{\frac{1}{2}j_2}. \end{aligned} \tag{4.26}$$

Next we define $d_i := \mathrm{dist}\{\mathcal{U}_{j_1,a_1} \times \mathcal{I}_i, \mathrm{diag}\}$, and consider those subintervals $\mathcal{I}_i$ which refer to the index set $\Lambda := \{i : d_i > 2^{-j_1}L\}$, i.e.,

$$\begin{aligned}
&\sum_{i\in\Lambda}\left|\int_{\mathcal{I}_i}\int_{\mathcal{U}_{j_1,a_1}}\int_{\mathbb{R}^2\times\mathbb{R}^2}\psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^3xd^3y\right| \\
&\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\sum_{i\in\Lambda}\left\|\int_{\mathbb{R}^2\times\mathbb{R}^2}\partial_{x_1}^p\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^2x_{23}d^2y_{23}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)} \\
&\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\sum_{i\in\Lambda}\left\|\int_{\mathbb{R}^2\times\mathbb{R}^2}\left|d_i^2+|\mathbf{x}_{23}-\mathbf{y}_{23}|^2\right|^{\frac{1-p}{2}}|f(\mathbf{x},\mathbf{y})||w_k(\mathbf{x}_{23},\mathbf{y}_{23})|d^2x_{23}d^2y_{23}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{I}_i)} \\
&\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\sum_{i\in\Lambda}\left(\int_0^\infty\left|d_i^2+s^2\right|^{1-p}sds\right)^{\frac{1}{2}}\|w\|_{L_2(\mathbb{R}^2\times\mathbb{R}^2)} \\
&\quad\lesssim 2^{-(p+\frac{3}{2})j_1}2^{\frac{1}{2}j_2}\sum_{i\in\Lambda}d_i^{2-p} \\
&\quad\lesssim 2^{-\frac{7}{2}j_1}2^{\frac{1}{2}j_2},
\end{aligned}\tag{4.27}$$

where we used $\sum_{i\in\Lambda}d_i^{2-p}\lesssim 2^{-(2-p)j_1}$ for $p>3$ vanishing moments. Summing, we get the estimate

$$\left|\langle\psi_{j_1,a_1}\psi_{j_2,a_2},u_k^{(1)}\rangle\right|\lesssim 2^{-\frac{7}{2}j_1}2^{\frac{1}{2}j_2},\tag{4.28}$$

for coefficients of bivariate wavelets located on or next to the diagonal.

The second case we have to consider are wavelets with $d_{j_2,a_2} := \mathrm{dist}\{\mathcal{U}_{j_1,a_1}\times\mathcal{U}_{j_2,a_2},\mathrm{diag}\} > 2^{-j_2}L$. In this case one gets the following estimate

$$\begin{aligned}
&\left|\int_{\mathcal{U}_{j_2,a_2}}\int_{\mathcal{U}_{j_1,a_1}}\int_{\mathbb{R}^2\times\mathbb{R}^2}\psi_{j_1,a_1}(x_1)\psi_{j_2,a_2}(y_1)\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^3xd^3y\right| \\
&\quad\lesssim 2^{-(p+\frac{1}{2})(j_1+j_2)}\left\|\int_{\mathbb{R}^2\times\mathbb{R}^2}\partial_{x_1}^p\partial_{y_1}^p\tau(\mathbf{x},\mathbf{y})w_k(\mathbf{x}_{23},\mathbf{y}_{23})d^2x_{23}d^2y_{23}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{U}_{j_2,a_2})} \\
&\quad\lesssim 2^{-(p+\frac{1}{2})(j_1+j_2)}\left\|\int_{\mathbb{R}^2\times\mathbb{R}^2}\left|\mathbf{x}-\mathbf{y}\right|^{1-2p}|f(\mathbf{x},\mathbf{y})||w_k(\mathbf{x}_{23},\mathbf{y}_{23})|d^2x_{23}d^2y_{23}\right\|_{L_\infty(\mathcal{U}_{j_1,a_1}\times\mathcal{U}_{j_2,a_2})} \\
&\quad\lesssim 2^{-(p+\frac{1}{2})(j_1+j_2)}\left(\int_0^\infty\left|d_{j_2,a_2}^2+s^2\right|^{1-2p}sds\right)^{\frac{1}{2}}\|w\|_{L_2(\mathbb{R}^2\times\mathbb{R}^2)} \\
&\quad\lesssim 2^{-(p+\frac{1}{2})(j_1+j_2)}d_{j_2,a_2}^{2-2p}.
\end{aligned}\tag{4.29}$$

These wavelet coefficients are specified by the index sets $\Gamma_{j_1,a_1}^{j_2} := \{a_2 : d_{j_2,a_2} > 2^{-j_2}L\}$.

It remains to estimate the weighted ℓ_q norm (4.8) on the square $\Omega := \mathcal{I}\times\mathcal{I}$ using (4.28) and (4.29) which gives

$$\begin{aligned}
&\sum_{j_1\ge j_2\ge j_0}\sum_{a_1,a_2}2^{qj_1}\left|\langle\psi_{j_1,a_1}\psi_{j_2,a_2},u_k^{(1)}\rangle\right|^q \\
&\quad\lesssim\sum_{j_1\ge j_2\ge j_0}2^{qj_1}2^{j_1}\left[2^{-\frac{7}{2}qj_1}2^{\frac{1}{2}qj_2}+2^{-(p+\frac{1}{2})q(j_1+j_2)}\sum_{a_2\in\Gamma_{j_1,a_1}^{j_2}}d_{j_2,a_2}^{(2-2p)q}\right] \\
&\quad\lesssim\sum_{j_1\ge j_2\ge j_0}2^{qj_1}2^{j_1}\left[2^{-\frac{7}{2}qj_1}2^{\frac{1}{2}qj_2}+2^{-(p+\frac{1}{2})qj_1}2^{(p-\frac{5}{2})qj_2}\right] \\
&\quad\lesssim\sum_{j_1\ge j_0}2^{-2qj_1}2^{j_1},\quad(p>2),
\end{aligned}\tag{4.30}$$

where we used $\sum_{a_2 \in \Gamma^{j_2}_{j_1,a_1}} d^{(2-2p)q}_{j_2,a_2} \lesssim 2^{-(2-2p)qj_2}$ for $p > \frac{1}{2} + \frac{1}{q}$ vanishing moments. Therefore, the sum (4.30) converges for $q > \frac{1}{2}$ and $u^{(1)}_k$ belongs to $A^\alpha_q(H^1(\Omega))$ for all $0 < \alpha < \frac{3}{2}$ with $\frac{1}{q} = \alpha + \frac{1}{2}$. □

5 Conclusions

We discussed a general approach, based on tensor product approximation, for transfering optimized pair-correlation functions from QMC Jastrow factors into F12 methods, the latter are presently considered to represent the most accurate computational scheme in quantum chemistry. For most applications QMC Jastrow factors recover between 85 and 90 % of the correlation energy. In particular, core and core-valence correlations are taken into account. The canonical tensor product approximation of pair-correlation functions leads to moderate separation ranks which enables an efficient computation of three-electron integrals in the tensor format. This requires the computation of various intermediate two-index quantities which themselves are approximated in the canonical tensor format. In order to perform all the necessary computations in the tensor format in an efficient manner, it is beneficial to have sparse representations of tensor components available. In the second half of the paper we discussed adaptive approximation of tensor components in wavelet bases and provide convergence rates for the best N-term approximation.

6 Acknowledgement

The authors gratefully acknowledge Professor W. Klopper (KIT Karlsruhe) for useful discussions.

A Basic notions of multiresolution analysis

The purpose of this appendix is to provide some basic facts about wavelets which are required in Section 4. For a detailed exposition of this subject, we refer to the monographs [41, 42, 43].

In one dimension, multiresolution analysis provides a partition of the Hilbert space $L_2(\mathbb{R})$ into an infinite sequence of ascending subspaces $\cdots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset \cdots$, where the index j runs over all integers. The union of these subspaces $\bigcup_j V_j$ is dense in $L_2(\mathbb{R})$. On each subspace V_j the scaling function $\varphi(x)$ generates a basis

$$\varphi_{j,a}(x) := 2^{j/2}\varphi(2^j x - a), \quad a \in \mathbb{Z}, \tag{A.1}$$

via the operations of dilation and translation. The dilation factor 2^j scales the size of the basis functions, which means that with increasing j, the $\varphi_{j,a}$ provide a finer resolution in $L_2(\mathbb{R})$. An explicit embedding of V_j into the larger space V_{j+1} is given by the refinement relation

$$\varphi(x) = 2\sum_a h_a\ \varphi(2x - a),$$

where the number of nonzero filter coefficients h_a is finite for the scaling functions considered in our application. Wavelet spaces W_j are defined as complements of V_j in V_{j+1}. The corresponding wavelet basis is generated from a mother wavelet $\psi(x)$ analogous to Eq. (A.1)

$$\psi_{j,a}(x) := 2^{j/2}\psi(2^j x - a), \quad a \in \mathbb{Z}.$$

This construction leads to a hierarchical decomposition

$$L_2(\mathbb{R}) = V_{j_0} \oplus \bigoplus_{j \geq j_0} W_j \tag{A.2}$$

into wavelet subspaces W_j [41]. In an orthogonal wavelet basis, i.e.,

$$\langle \varphi_{j_0,a} | \varphi_{j_0,b} \rangle = \delta_{a,b}, \qquad \langle \psi_{j,a} | \psi_{k,b} \rangle = \delta_{j,k}\,\delta_{a,b}, \qquad \langle \varphi_{j_0,a} | \psi_{j,b} \rangle = 0, \tag{A.3}$$

with $j_0 \leq j, k$, any function in $L_2(\mathbb{R})$ can be represented as

$$f(x) = \sum_a \langle \varphi_{j_0,a} | f \rangle \, \varphi_{j_0,a}(x) \; + \; \sum_{j=j_0}^{\infty} \sum_a \langle \psi_{j,a} | f \rangle \, \psi_{j,a}(x), \tag{A.4}$$

where the scaling function and wavelet coefficients are given by the corresponding scalar products. The multiscale approximation of piecewise smooth functions reveals an important sparsity feature due to the vanishing moments property of wavelets. Depending on the specific choice of the wavelet, a certain number of moments vanish, i.e.,

$$\int dx \, x^k \, \psi(x) = 0, \;\; \text{for} \;\; k = 0, \ldots, p-1. \tag{A.5}$$

This property has a significant effect on the magnitude of wavelet coefficients, as can be seen from a local Taylor series expansion

$$f(x) = c_0 + \cdots + c_{n-1}(x - 2^{-j}a)^{n-1} + R_{n-1}(x)(x - 2^{-j}a)^n, \tag{A.6}$$

with $n \leq p$, at the center of a wavelet $\psi_{j,a}(x)$. Inserting the Taylor series (A.6) into a scalar product yields the following estimate for the wavelet coefficient

$$|\langle \psi_{j,a} | f \rangle| \lesssim 2^{-j(n+1/2)} \|f^{(n)}\|_{L_\infty(\operatorname{supp} \psi_{j,a})}, \tag{A.7}$$

which represents the basic estimate used in Section 4.

References

[1] E. A. Hylleraas, Z. Phys. **54** (1929) 347.

[2] J. W. Clark, in Progress in Nuclear and Particle Physics, Vol. 2, Ed. D. H. Wilkinson, Pergamon, Oxford, 1979, p. 89.

[3] E. Krotscheck, Phys. Rev. B **31** (1985) 4267.

[4] E. Krotscheck, W. Kohn and G.-X. Qian, Phys. Rev. B **32** (1985) 5693.

[5] C. E. Campbell, E. Krotscheck and T. Pang, Physics Reports **223** (1992) 1.

[6] R. F. Bishop, Theor. Chim. Acta **80** (1991) 95.

[7] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60** (1988) 1719.

[8] X. Lin, H. Zhang and A. M. Rappe, J. Chem. Phys. **112** (2000) 2650.

[9] M. W. Lee, M. Mella and A. M. Rappe, J. Chem. Phys. **122** (2005) 244103.

[10] C. J. Umrigar and C. Filippi, Phy. Rev. Lett. **94** (2005) 150201.

[11] J. Toulouse and C. J. Umrigar, J. Chem. Phys. **126** (2007) 084102.

[12] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella and R. G. Hennig, Phy. Rev. Lett. **98** (2007) 110201.

[13] A. Lüchow and J. B. Anderson, Annu. Rev. Phys. Chem. **51** (2000) 501.

[14] W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, Rev. Mod. Phys. **73** (2001) 33.

[15] A. J. Williamson, S. D. Kenny, G. Rajagopal, A. J. James, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, and P. Maccallum, Phys. Rev. B **53** (1996) 9640.

[16] N. D. Drummond, M. D. Towler and R. J. Needs, Phys. Rev. **B 70** (2004) 235119.

[17] W. Klopper and W. Kutzelnigg, Chem. Phys. Lett. **134** (1987) 17.

[18] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94** (1991) 1985.

[19] C. C. M. Samson, W. Klopper and T. Helgaker, Comput. Phys. Commun. **149** (2002) 1.

[20] S. Ten-no, Chem. Phys. Lett. **398** (2004) 56.

[21] A. J. May and F. R. Manby, J. Chem. Phys. **121** (2004) 4479.

[22] A. J. May, E. Valeev, R. Polly and F. R. Manby, Phys. Chem. Chem. Phys. **7** (2005) 2710.

[23] D. P. Tew and W. Klopper, J. Chem. Phys. **123** (2005) 074101.

[24] W. Klopper, F. R. Manby, S. Ten-No, E. F. Valeev, Int. Rev. Phys. Chem. **25** (2006) 427.

[25] S. Höfener, D. P. Tew, W. Klopper and T. Helgaker, Chem. Phys. **356** (2009) 25.

[26] K. E. Schmidt, and J. W. Moskowitz, J. Chem. Phys. **93** (1990) 4172.

[27] S. F. Boys and N. C. Handy, Proc. R. Soc. London Ser. A **310** (1969) 63.

[28] T.G. Kolda and B. W. Bader, SIAM Review **51** (2009) 455.

[29] S. R. Chinnamsetty, M. Espig, B. N. Khoromskij, W. Hackbusch and H.-J. Flad, J. Chem. Phys. **127** (2007) 084110.

[30] S. R. Chinnamsetty, M. Espig, H.-J. Flad and W. Hackbusch, Z. Phys. Chem. **224** (2010) 681.

[31] M. Espig and W. Hackbusch, Preprint Nr. 78 MPI-MIS 2010.

[32] A. Uschmajew, Constr. Approx., published online DOI 10.1007/s00365-010-9125-4

[33] S. Ten-no, J. Chem. Phys. **121** (2004) 117.

[34] S. F. Boys and N. C. Handy, Proc. R. Soc. London Ser. A **310** (1969) 43.

[35] S. R. Chinnamsetty, W. Hackbusch and H.-J. Flad, Comput. Vis. Sci. **13** (2010) 397.

[36] R. A. DeVore, Acta Numerica **7** (1998) 51.

[37] P.-A. Nitsche, Constr. Approx. **24** (2006) 49.

[38] H.-J. Flad, R. Schneider and B.-W. Schulze, Math. Methods Appl. Sci. **31** (2008) 2172.

[39] H.-J. Flad, W. Hackbusch and R. Schneider, ESAIM: M2AN **41** (2007) 261.

[40] H.-J. Flad, W. Hackbusch, D. Kolb and R. Schneider, J. Chem. Phys. **116** (2002) 9641.

[41] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics **61** (1992).

[42] Y. Meyer, Wavelets and Operators, Cambridge University Press, Cambridge, 1992.

[43] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.