Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Tensor product approximation with optimal rank in quantum chemistry

by

Sambasiva Rao Chinnamsetty, Mike Espig, Boris N. Khoromskij, Wolfgang Hackbusch, and Heinz-Jürgen Flad

Preprint no.: 41 2007

Tensor product approximation with optimal rank in quantum chemistry

Sambasiva Rao Chinnamsetty, Mike Espig, Boris N. Khoromskij, and Wolfgang Hackbusch
Max-Planck-Institut für Mathematik in den Naturwissenschaften, Inselstr. 22-26, D-04103 Leipzig

Heinz-Jürgen Flad
Institut für Informatik, Christian-Albrechts-Universität zu Kiel, Christian-Albrechts-Platz 4, D-24098 Kiel

April 13, 2007

Abstract

Tensor product decompositions with optimal rank provide an interesting alternative to traditional Gaussian-type basis functions in electronic structure calculations. We discuss various applications for a new compression algorithm, based on the Newton method, which provides for a given tensor the optimal tensor product or so-called best separable approximation for fixed Kronecker rank. In combination with a stable quadrature scheme for the Coulomb interaction, tensor product formats enable an efficient evaluation of Coulomb integrals. This is demonstrated by means of best separable approximations for the electron density and Hartree potential of small molecules, where individual components of the tensor product can be efficiently represented in a wavelet basis. We present a fairly detailed numerical analysis, which provides the basis for further improvements of this novel approach. Our results suggest a broad range of applications within density fitting schemes, which have been recently successfully applied in quantum chemistry.

1 Introduction

1.1 Tensor products in electronic structure calculations

Tensor product approximations have a long and successful history in quantum chemistry. First of all we have to mention the representation of many-electron wavefunctions as antisymmetric tensor products of single-particle wavefunctions, usually termed as linear combination of Slater determinants. This ansatz dramatically reduces the computational complexity for the evaluation of matrix elements involving many-electron Hamiltonians and wavefunctions. The originally $3N$-dimensional integrals, where N is the number of electrons, decompose into the familiar one- and two-electron integrals. Recent work by Yserentant [1] suggests a special kind of regularity property of exact many-electron wavefunctions which provides a clue from the mathematical point of view why this ansatz turns out to be highly successful. Another prominent example for tensor products in quantum chemistry are *Gaussian-type orbitals* (GTO) basis sets for the approximation of single-particle wavefunctions, so-called orbitals, and related quantities like one- and two-electron reduced density matrices. Here the tensor product structure does not come to the fore like in the previous example. Nevertheless it turns out to be very useful for the computation of one- and two-electron integrals. More than ten years ago, Almlöf suggested in a visionary paper [2] a farther reaching approach, based on tensor product decompositions, for an efficient evaluation of the Fock matrix.

Within the present work, we want to address the question whether GTO basis sets are already optimal for tensor product approximations of orbitals and electron densities. Furthermore, we consider tensor product representations of the Coulomb potential and its convolution with the electron density leading to the Hartree potential. The latter represents a basic building block for the *Hartree-Fock* (HF) method and *density functional theory* (DFT).

1.2 Tensor product approximations with optimal rank

Customarily, tensor products appear in the approximation theory of multivariate functions via the construction of a multi-dimensional basis by taking tensor products of appropriate univariate basis functions. A typical example are plane wave basis sets frequently applied in solid state physics. Such kind of construction, however, suffers from an exponential growth of the tensor product bases with increasing dimension. For functions with sufficiently regular mixed derivatives and a hierarchical univariate basis, it is a priori possible to select a subset of the tensor product basis set, whose size is "almost" independent from dimension, without essential loss of approximation properties. This so-called *sparse grid* approach [3, 4, 5] can be further improved by considering a fully adaptive *best N-term approximation* [6, 7] where tensor products are selected according to the function to be approximated. For recent applications of best N-term approximation in the framework of quantum chemistry we refer to Refs. [8, 9]. In a certain sense, the modus operandi for GTO bases resembles to a best N-term approximation. However the GTOs are not part of an underlying systematic basis set, instead they have been optimized by themselves in atomic calculations. With respect to this, GTO bases have to be considered as some kind of generalized tensor product approximation. In recent years, general tensor product decompositions without predetermined univariate components attracted considerable interest in numerical mathematics. The interest in this subject was mainly driven by their pleasing algebraic properties and the possibility to obtain efficient algorithms for higher dimensional problems. We refer e.g. to Refs. [10, 11, 12, 13, 14, 15, 16, 17, 18] and the references therein. Besides quantum chemistry, numerous other applications of general tensor product decompositions in chemistry have been reported in the literature. Most of them are related to high dimensional problems in data analysis with moderate accuracy requirements. For further details and a comprehensive bibliography, we refer to the monograph [19]. Recently, Beylkin and Mohlenkamp [20, 21] initiated an extensive program to investigate low rank tensor product decompositions in electronic structure calculations. Related work has been also done by Ibraghimov [22].

Within the present work, we want to study *best separable approximations* with fixed Kronecker rank κ for functions in $\mathbb{R}^3$, like orbitals, electron densities and Hartree potentials, which means that we are looking for the best possible approximation of a function f in the format

$$f(\mathbf{x}) \approx \sum_{k=1}^{\kappa} h_k^{(1)}(x_1)\, h_k^{(2)}(x_2)\, h_k^{(3)}(x_3), \tag{1}$$

where the univariate functions $h_k^{(i)}(x_i)$ have to be chosen in an "optimal" way. In particular, there are no orthogonality constraints imposed. The notion of a Kronecker rank has been borrowed from linear algebra [19] where it denotes the least possible decomposition rank of a tensor. We refer to as separation rank in the case of tensor product decompositions for which the rank is supposed to be suboptimal. Strictly speaking, the error of best separable rank κ approximations is defined via the least-squares problem

$$\sigma_\kappa(f) := \inf_{h_k^{(i)} \in L^2(\mathbb{R})} \left\| f(\mathbf{x}) - \sum_{k=1}^{\kappa} h_k^{(1)}(x_1)\, h_k^{(2)}(x_2)\, h_k^{(3)}(x_3) \right\|_{L^2(\mathbb{R}^3)}, \tag{2}$$

where the L^2 norm has been chosen mainly for technical reasons. Such kind of variational problems are typically ill-posed, cf. Ref. [23], which means that a minimizer does not necessarily exist. This is due to the fact that the space of rank κ tensor products is not closed in $L^2(\mathbb{R}^3)$. Nevertheless it is possible to choose a set of functions $h_k^{(i)}(x_i)$ for which the error is arbitrarily close to σ_κ. However it might happen that this set is ill-conditioned in a sense to be discussed below. Another problem arises from a multitude of local minima into which a numerical algorithm for the solution of the variational problem (2) might get stuck. Despite these shortcomings, one should bear in mind that for our envisaged applications it is only of relevance to achieve a certain accuracy with a reasonably small Kronecker rank κ and it is completely irrelevant whether this representation actually corresponds to a global minimum. We want to mention at

the end of this paragraph, that an approximation similar to (2) has been already discussed by Dacre and Elder in the early seventies, cf. Ref. [24], for the computation of two-electron integrals. They required the univariate functions $h_k^{(i)}$ to be of Gaussian-type and optimized not only the coefficients but also their centers and exponents using a quasi-Newton algorithm.

The variational problem (2) can be considerably simplified if we assume that the original function f is already given in a tensor product format. In the mathematical literature, the representation of a given tensor by a tensor product of minimal rank became known as *canonical decomposition* (CANDECOMP) or *parallel factors* (PARAFAC) model, cf. Ref. [11]. A typical example are functions represented in a GTO basis set

$$f(\mathbf{x}) = \sum_{k=1}^{K} a_k \, g_k^{(1)}(x_1 - A_1^{(k)}) g_k^{(2)}(x_2 - A_2^{(k)}) g_k^{(3)}(x_3 - A_3^{(k)}) \tag{3}$$

with

$$g_k^{(i)}(x_i) = C_{i,k} x_i^{l_{i,k}} e^{-\zeta_k x_i^2} \quad \text{for } i = 1, 2, 3. \tag{4}$$

For simplicity we consider here and in the following only primitive cartesian GTO basis sets. In a preparatory step, the individual GTOs are discretized in each direction on a uniform grid

$$g_k^{(i)}(x_i - A_i^{(k)}) \approx \sum_a b_{k,l,a}^{(i)} \, \varphi_{l,a}(x_i) \quad \text{where } b_{k,l,a}^{(i)} = 2^{-l/2} g_k^{(i)}(2^{-l} a - A_i), \tag{5}$$

with grid spacing $h = 2^{-l}$ bohr, using the interpolating scaling functions $\varphi_{l,a}(x) := 2^{l/2} \varphi(2^l x - a)$ of Deslauriers and Dubuc [25]. The discrete counterpart of the function (3) is represented by the tensor product decomposition

$$\mathbf{f}_l := \sum_{k=1}^{K} a_k \, \mathbf{b}_{k,l}^{(1)} \otimes \mathbf{b}_{k,l}^{(2)} \otimes \mathbf{b}_{k,l}^{(3)}, \tag{6}$$

with entries

$$(\mathbf{f}_l)_{a_1 a_2 a_3} := \sum_{k=1}^{K} a_k \, \left(\mathbf{b}_{k,l}^{(1)}\right)_{a_1} \left(\mathbf{b}_{k,l}^{(2)}\right)_{a_2} \left(\mathbf{b}_{k,l}^{(3)}\right)_{a_3} \tag{7}$$

where $\left(\mathbf{b}_{k,l}^{(i)}\right)_{a_i} := b_{k,l,a_i}^{(i)}$ for $i = 1, 2, 3$ refers to coefficient vectors in the expansion (5). Such kind of discretization procedure can be applied to any function f in the format (1) irrespective of whether or not the representation is optimal. In the following we will denote by K the separation rank of arbitrary tensor products and reserve κ for the Kronecker rank of best separable approximations. The subsequent compression step is performed in a purely algebraic manner using the discretized representation (6) and the corresponding least-squares functional

$$\sigma_\kappa(\mathbf{f}) := \inf_{\mathbf{h}_k^{(i)} \in \ell^2} \left\| \mathbf{f} - \sum_{k=1}^{\kappa} \mathbf{h}_k^{(1)} \otimes \mathbf{h}_k^{(2)} \otimes \mathbf{h}_k^{(3)} \right\|_{\ell^2}, \tag{8}$$

where the L^2 norm of the function space has been replaced by the ℓ^2 norm defined on the sequence space of coefficients $\mathbf{b}_{k,l}^{(i)}$. We want to mention, however, that for the (bi)orthogonal wavelet and scaling function bases discussed below, a norm equivalence exists between L^2 and ℓ^2. This means that minimizing the discrete functional (8) is virtually equivalent to minimizing the continuous functional (2). We refer to Appendix A for a further discussion of this subject. As already mentioned, it might happen that the optimization problem is ill-posed. In such a case it becomes necessary to apply a Tichonov regularization to the least-squares functional (8), cf. Ref. [21] for further details.

Two different compression algorithms were available to us from the outset of our project. The first algorithm uses an alternating least squares method [26] and the second one is based on the Newton method [27]. Both algorithms achieved comparable compression rates for small Kronecker ranks. However it turned

out that the Newton algorithm is superior at high accuracies where comparatively large Kronecker ranks are required. We want to emphasize, however, that both algorithms typically generate a sequence of local minima. In an initial step at Kronecker rank κ, the Newton algorithm tries to find the best rank one approximation with respect to the residuum of the best rank $\kappa - 1$ approximation. Although it is known to exist [27, 28], it might be hidden among a large number of local minima. Therefore, the algorithm computes several local minima and selects the most promising one as a starting guess for the actual Newton optimization. Within the minimization, the Newton direction has to be calculated in each step. Already at fairly small compression ranks > 10, the condition number of the Hessian matrix becomes rather poor resulting in a significant increase of computational costs. This problem can be circumvented by using a trust-region method [29] to modify the Hessian matrix in those cases where its condition number becomes critical. In our applications, we have used both versions of the Newton algorithm with and without the trust-region method. It turned out, that these algorithms end up in different local minima which may slightly differ with respect to their L^2 error for compression ranks $\kappa > 10$. Such small variations of the lowest rank, required in order to achieve a certain accuracy, are not really relevant from a practical point of view. At higher compression ranks $\kappa > 20$, the trust-region method becomes mandatory for certain applications and exhibits a stable convergence behaviour.

Starting from the discrete Kronecker tensor product (6), the Newton algorithm tries to solve the corresponding optimization problem (8) with $\kappa < K$. For this linear vector spaces

$$\mathbf{B}^{(i)} := \text{span}\left\{\mathbf{b}_{1,l}^{(i)}, \ldots, \mathbf{b}_{K,l}^{(i)}\right\} \quad \text{for } i = 1, 2, 3, \tag{9}$$

of dimension $r_i \leq K$ have to be considered. It can be shown that the optimized components $\mathbf{h}_k^{(i)}$ in each direction belong to the linear space $\mathbf{B}^{(i)}$, cf. Refs. [27, 30]. Concerning the computational complexity of the algorithm let us first assume $\kappa \ll K$. Typically, we observed that the dimensions r_i of the linear vector spaces $\mathbf{B}^{(i)}$ are considerably smaller than the initial separation rank K. Linear dependencies among the vectors $\mathbf{b}_{k,l}^{(i)}$ occur e.g. because certain univariate Gaussian functions may appear repeatedly within different GTOs in the expansion (3). The initial orthogonalization step to create orthogonal bases for the spaces $\mathbf{B}^{(i)}$ needs a QR decomposition which requires $O(n_l K^2)$ computational effort where n_l denotes the size of the vectors $\mathbf{b}_{k,l}^{(i)}$. The actual Newton algorithm has $O(\kappa K r_{max} + \kappa^3)$, with $r_{max} := \max\{r_i\}$, computational complexity and requires $O(\kappa r_{max})$ memory. It should be mentioned that the time consuming orthogonalization step can be improved by using a hierarchical wavelet basis instead of scaling functions.

Instead of discretizing the function (3) in a scaling function basis it is possible to work directly in the function spaces

$$G^{(i)} := \text{span}\left\{g_k^{(i)}(\cdot - A_i^{(k)})\right\} \quad \text{for } i = 1, 2, 3, \tag{10}$$

spanned by the univariate Gaussians. As already mentioned in the previous paragraph this becomes possible because the univariate components $h_k^{(i)}$ belong to the function spaces $G^{(i)}$. For a proof of this statement, we refer to Ref. [27]. The Newton algorithm only requires the construction of an orthogonal basis in $G^{(i)}$ which can be easily achieved e.g. by Löwdin orthogonalization after removing possible linear dependences among the $g_k^{(i)}(\cdot - A_i^{(k)})$. With such kind of ansatz, the $h_k^{(i)}$ are finally expressed in terms of Gaussians again, which might be of interest for certain applications, e.g. where best separable approximations have to be incorporated into an already existing code that operates with GTO basis functions. Within the present work, we do not pursue this ansatz further because it is only of limited use for our more general approach to the evaluation of Coulomb integrals outlined in Section 2. Our approach requires basis sets which provide a stable resolution of the identity. This is hard to achieve with GTO basis sets.

1.3 "Optimal" tensor products versus GTO bases

In order to judge the applicability of best separable rank κ approximations in quantum chemistry, we have studied orbital compression rates for a few single-electron systems. First we solved the Schrödinger equation

$$\left(-\frac{1}{2}\Delta - \sum_{\mathbf{A}} \frac{Z_A}{|\mathbf{x}-\mathbf{A}|}\right)\phi(\mathbf{x}) = \varepsilon_0\,\phi(\mathbf{x}), \tag{11}$$

using large uncontracted GTO basis sets. In a second step we have generated best separable rank κ approximations of the wavefunctions using the algorithms outlined in Section 1.2. The simplest of these systems is the H atom. It is only for the H atom where rigorous results are available concerning approximation errors of GTO basis sets for single-electron Schrödinger equations. It has been shown in Refs. [31, 32] that GTO bases achieve almost exponential convergence rates $\sigma_\kappa(\phi_{\mathrm{H}}) \sim e^{-\mu\sqrt{\kappa}}$ for the H atom. Whether this result remains valid in the case of HF calculations for general molecules seems to be an open issue, although there exists some numerical evidence in favour of it [33]. Errors in energy for best separable rank κ approximations obtained with the Newton algorithm are shown in Fig. 1 a). These have been compared with best radial Gaussian approximations [34] of the same rank. The latter have been optimized with respect to a weighted $L^\infty(\mathbb{R}_+)$ norm. It can be seen that both approaches lead to similar approximation errors, where the best separable rank κ approximation actually performs slightly better. Our results indicate that Gaussians already provide almost optimal separable approximations for the H atom. However this does not seem to be the case for systems with several nuclei as can be seen from Fig. 1 b) where significant compression rates can be achieved with respect to standard GTO basis sets. Here we have chosen an uncontracted $(8s4p3d)$ GTO basis for each atom with exponents taken from V5Z basis sets [35]. Once the Schrödinger equation (11) has been solved in these basis sets, we compressed the wavefunctions to different Kronecker ranks $1 \le \kappa \le 15$ using the alternating least-squares algorithm. Thereafter, the set of components $\{h_k^{(1)}(x_1)\,h_k^{(2)}(x_2)\,h_k^{(3)}(x_3)\}$ of the tensor product expansion (1) was taken as a new basis in which the Schrödinger equation (11) has been solved again. We have performed such calculations for systems containing two or three nuclei. It should be mentioned that the approximation error of best separable rank κ approximations might depend on the orientation of the molecule with respect to the coordinate axes, e.g. a linear combination of two Gaussians with different centers and same exponents can be written as a rank one tensor if oriented along an axis. The dimers H_2^+, HHe^{2+}, and a linear trimer H_3^{2+} were therefore oriented along the diagonal in order to avoid such kind of rank reductions due to symmetry. Furthermore, we have considered a triangular structure for H_3^{2+} with each nucleus located on an axis. The resulting errors in energy are shown in Fig. 1 b). It can be seen that the overall convergence for these systems is rather similar and an error in energy $\le 10^{-4}$ hartree, which roughly corresponds to the basis set error of the GTO bases, can be achieved at Kronecker rank $\kappa \le 15$. For comparison, the separation ranks of the initial GTO bases, ranging from $K = 58$ for H_2^+ up to $K = 126$ for H_3^{2+} (triangular structure), are considerably larger.

Another important aspect of best separable rank κ approximations is the sparse representation of univariate components $h_k^{(i)}$ for $i = 1, 2, 3$. For the tensor product compression step, these components have been represented so far in a scaling function basis (5) on a uniform grid. This becomes rather inefficient for extended systems containing heavy nuclei because the resolution level has to be adjusted to the typical length-scales of atomic cores. Instead it seems to be more favourable to work in a wavelet representation where adaptive refinement becomes possible. Furthermore wavelets enable a second compression step on the univariate components $h_k^{(i)}$ (see Appendix A for further details). Such kind of approach has been discussed by Ford and Tyrtyshnikov, cf. Refs. [14, 15], for solving dense linear systems in higher dimensions. The transition from scaling functions into a wavelet basis can be achieved by means of a fast wavelet transformation which consists of a successive application of the two-scale decomposition of scaling

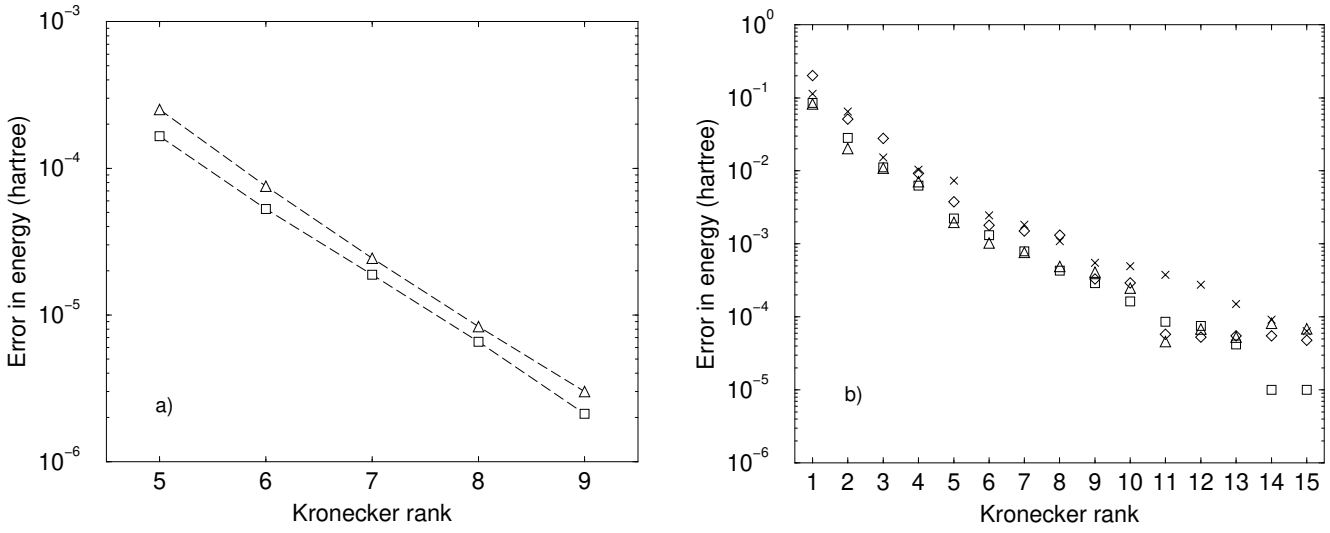

Figure 1: Error in total energy (hartree) versus Kronecker rank κ of best separable approximations for one-electron wavefunctions. a) Best separable $L^2(\mathbb{R}^3)$ ($\square$) and best radial Gaussian $L^\infty(\mathbb{R}_+)$ (Δ) approximations for the H atom. b) Best separable rank κ approximations for various positively charged dimers ($\square$) H_2^+ ($K = 58$), ($\diamond$) HHe^{2+} ($K = 58$), and trimers (Δ) H_3^{2+} ($K = 126, \text{triangle}$), ($\times$) H_3^{2+} ($K = 75, \text{linear}$). The reference wavefunctions have been calculated in an uncontracted ($8s4p3d$) GTO bases. Nuclei are located on the diagonal or on the axes to form a triangle.

functions into wavelets and scaling functions on the next coarser level, i.e.

$$\psi^{(0)}_{j,a}(x) = \sum_b \left(c_{ab}\, \psi^{(0)}_{j-1,b}(x) + d_{ab}\, \psi^{(1)}_{j-1,b}(x) \right). \tag{12}$$

Here and in the following we use the joint notation

$$\psi^{(0)}_{j,a}(x) := \varphi_{j,a}(x) \quad \text{and } \psi^{(1)}_{j,a}(x) := \psi_{j,a}(x). \tag{13}$$

for scaling functions φ and wavelets ψ, respectively. Herewith, we obtain the wavelet representation of the univariate components

$$\begin{aligned} h^{(i)}_k(x_i) &= \sum_a h^{(i)}_{k,0,l,a} \psi^{(0)}_{l,a}(x_i) \\ &= \sum_a h^{(i)}_{k,0,l_0,a} \psi^{(0)}_{l_0,a}(x_i) + \sum_{j=l_0}^{l-1} \sum_a h^{(i)}_{k,1,j,a} \psi^{(1)}_{j,a}(x_i). \end{aligned} \tag{14}$$

A typical result is shown in Fig. 2 for H_2^+ at $\kappa = 5$ for D10 wavelets with five vanishing moments. The univariate components $h^{(1)}_k(x_1)$, depicted in Fig. 2 a), appear to be rather smooth. It is shown in Fig. 2 b) that the fraction of wavelet coefficients with $|h^{(i)}_{k,1,j,a}| \geq 10^{-6}$ decreases exponentially with respect to the level j. This observation paves the way for a fast computation of various integrals involving orbitals and electron densities. Before we can delve into this subject it is necessary to have a separable approximation for the Coulomb potential at our disposal, which is adapted to the specific requirements of wavelet bases.

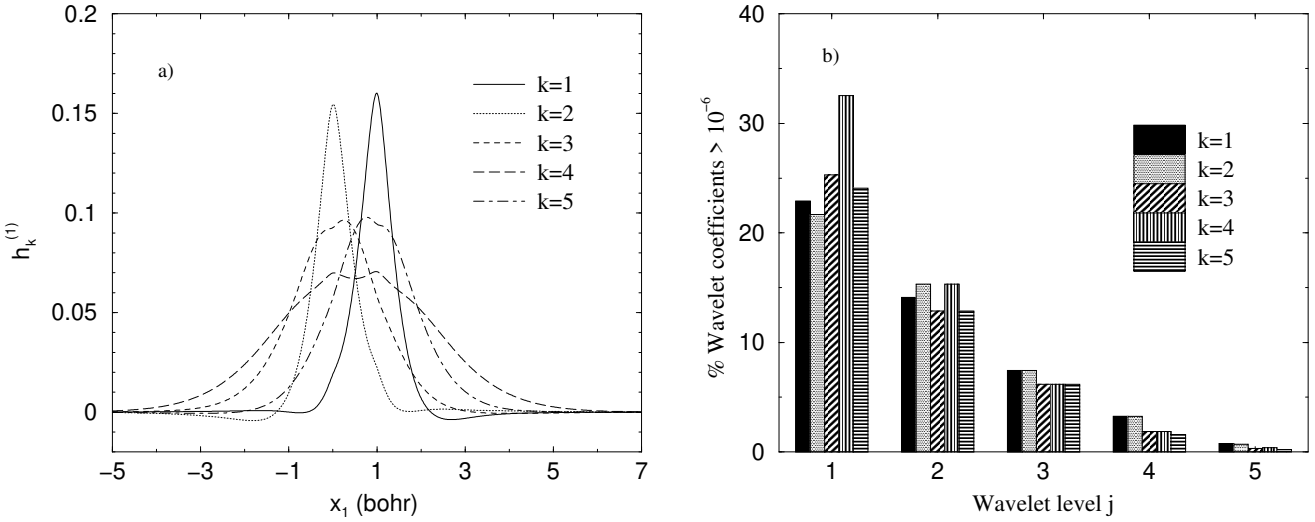

Figure 2: Best separable rank 5 approximation for H_2^+ oriented along the diagonal. a) Univariate components $h_k^{(1)}(x_1)$ of Kronecker tensor product. b) Wavelet compression of univariate components $h_k^{(1)}$ for D10 wavelets with five vanishing moments. The percentage of wavelet coefficients with absolute magnitude $\geq 10^{-6}$ is shown for different levels j.

2 Separable approximation of the Coulomb potential

The standard integration schemes for the evaluation of Coulomb integrals for GTO basis sets are based on the Gaussian transform of the Coulomb interaction

$$\frac{1}{\mid \mathbf{x}-\mathbf{y} \mid} = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-|\mathbf{x}-\mathbf{y}|^2 t^2}\, dt. \tag{15}$$

Such kind of representation suggests a formal ansatz for a separable approximation

$$\frac{1}{\mid \mathbf{x}-\mathbf{y} \mid} \approx \sum_i \omega_i\, e^{-|\mathbf{x}-\mathbf{y}|^2 t_i^2}, \tag{16}$$

which allows for different concrete realizations. One possibility is to consider a pointwise approximation in the finite radial interval $[1, R]$. This can be accomplished either by quadrature formulas for the integral (15), cf. [36, 37], or in the sense of a best N-term approximation studied in Refs. [38]. Using a simple scaling relation it can be adapted to approximate the Coulomb interaction in radial intervals $[\epsilon, \epsilon R]$ for any $\epsilon > 0$. Recently this method has been successfully applied in combination with multi wavelets by Harrison and coworkers [39, 40] for HF and KS calculations and for interpolating scaling functions and wavelets within the *BigDFT* program of Deutsch et al. [41, 42]. Another possible approach is to construct quadrature formulas which approximate the Coulomb integrals itself with respect to a given basis. For GTO basis sets this corresponds to the familiar Rys quadrature introduced into quantum chemistry thirty years ago by Dupuis, Rys and King [43].

In order to evaluate general Coulomb integrals for tensor product decompositions, where the univariate components (14) are represented in a wavelet basis, we essentially follow the nonstandard approach originally suggested by Beylkin et al. [44, 45]. It uses a completeness relation similar to the *resolution of the identity* (RI) methods commonly employed in quantum chemistry. We want to mention however, that contrary to GTOs, wavelets provide not only complete but also stable basis sets, cf. Appendix A, which enable a mathematically rigorous notion of the RI approach. An appropriate choice for this purpose are isotropic 3d tensor product wavelets

$$\gamma_{j,\mathbf{a}}^{(\mathbf{s})}(\mathbf{x}) = \psi_{j,a_1}^{(s_1)}(x_1)\, \psi_{j,a_2}^{(s_2)}(x_2) \psi_{j,a_3}^{(s_3)}(x_3), \tag{17}$$

cf. our notation (13), which consist of different combinations of (bi)orthogonal univariate scaling functions and wavelets. Accordingly, we denote by $\tilde{\gamma}^{(\mathbf{s})}_{j,\mathbf{a}}$ the corresponding dual wavelets. The special case $s_1 = s_2 = s_3 = 0$ corresponds to 3d scaling functions whereas the seven other possible combinations, with at least a single univariate wavelet in the tensor product, represent the 3d wavelet basis. These wavelets possess different properties depending on the total number of vanishing moments in the various directions. Given any two functions η_α, η_β of the form

$$\eta(\mathbf{x}) = \eta^{(1)}(x_1)\,\eta^{(2)}(x_2)\,\eta^{(3)}(x_3), \tag{18}$$

and using telescopic expansions of isotropic 3d wavelets

$$\sum_{\mathbf{a}} |\tilde{\gamma}^{(\mathbf{0})}_{j+1,\mathbf{a}}\rangle\langle\gamma^{(\mathbf{0})}_{j+1,\mathbf{a}}| = \sum_{\mathbf{p}}\sum_{\mathbf{a}} |\tilde{\gamma}^{(\mathbf{p})}_{j,\mathbf{a}}\rangle\langle\gamma^{(\mathbf{p})}_{j,\mathbf{a}}|, \tag{19}$$

the corresponding approximate Coulomb integral

$$\begin{aligned}\int\int \eta_\alpha(\mathbf{x})\,\frac{1}{|\,\mathbf{x}-\mathbf{y}\,|}\,\eta_\beta(\mathbf{y})\,d\mathbf{x}d\mathbf{y} \\ \approx \sum_{j=j_0}^{j_{max}} {\sum_{\mathbf{p},\mathbf{q}}}' \sum_{\mathbf{a},\mathbf{b}} \langle\eta_\alpha|\tilde{\gamma}^{(\mathbf{p})}_{j,\mathbf{a}}\rangle\langle\tilde{\gamma}^{(\mathbf{q})}_{j,\mathbf{b}}|\eta_\beta\rangle \int\int \gamma^{(\mathbf{p})}_{j,\mathbf{a}}(\mathbf{x})\,\frac{1}{|\,\mathbf{x}-\mathbf{y}\,|}\,\gamma^{(\mathbf{q})}_{j,\mathbf{b}}(\mathbf{y})\,d\mathbf{x}d\mathbf{y},\end{aligned} \tag{20}$$

can be expressed through elementary Coulomb integrals

$$\int\int \gamma^{(\mathbf{p})}_{j,\mathbf{a}}(\mathbf{x})\,\frac{1}{|\,\mathbf{x}-\mathbf{y}\,|}\,\gamma^{(\mathbf{q})}_{j,\mathbf{b}}(\mathbf{y})\,d\mathbf{x}d\mathbf{y} = \frac{2^{-2j+1}}{\sqrt{\pi}}\int_0^\infty \mathcal{I}^{(\mathbf{p},\mathbf{q})}(t,\mathbf{a}-\mathbf{b})dt, \tag{21}$$

with

$$\mathcal{I}^{(\mathbf{p},\mathbf{q})}(t,\mathbf{a}) := G^{(p_1,q_1)}(a_1,t)\,G^{(p_2,q_2)}(a_2,t)\,G^{(p_3,q_3)}(a_3,t), \tag{22}$$

and

$$G^{(p_i,q_i)}(a_i,t) := \int\int \psi^{(p_i)}(x-a_i)\,e^{-(x-y)^2t^2}\,\psi^{(q_i)}(y)\,dxdy, \tag{23}$$

where we have used the Gaussian transform of the Coulomb potential (15) and a simple scaling relation. The prime at the sum over the wavelet types $\mathbf{p},\mathbf{q}$ indicates that the pure scaling function combination $\mathbf{p} = \mathbf{q} = \mathbf{0}$ appears only on the coarsest level j_0. A more detailed outline of the nonstandard approach, applied to Coulomb integrals (20), and a general algorithm for the computation of the integrals (23) have been given in Ref. [46]. Elementary Coulomb integrals (21) for wavelets with nonoverlapping supports, i.e. $\mathrm{dist}(\mathrm{supp}\,\gamma^{(\mathbf{p})}_{j,\mathbf{a}}, \mathrm{supp}\,\gamma^{(\mathbf{q})}_{j,\mathbf{b}}) > 0$, can be estimated according to, cf. Refs. [47, 48],

$$\left|\int\int \gamma^{(\mathbf{p})}_{j,\mathbf{a}}(\mathbf{x})\frac{1}{|\mathbf{x}-\mathbf{y}|}\gamma^{(\mathbf{q})}_{j,\mathbf{b}}(\mathbf{y})\,d\mathbf{x}d\mathbf{y}\right| \le C_{\mathbf{p},\mathbf{q}}\frac{2^{-j(3+m(|\mathbf{p}|+|\mathbf{q}|))}}{\mathrm{dist}(\mathrm{supp}\,\gamma^{(\mathbf{p})}_{j,\mathbf{a}}, \mathrm{supp}\,\gamma^{(\mathbf{q})}_{j,\mathbf{b}})^{1+m(|\mathbf{p}|+|\mathbf{q}|)}}, \tag{24}$$

with $|\mathbf{p}| := p_1 + p_2 + p_3$ and m denotes the number of vanishing moments of the underlying univariate wavelet. The estimate (24) applied to the right hand side of Eq. (20) enables an efficient evaluation of these integrals.

To take advantage of the tensor product structure of the functions η_α, η_β, we require an efficient quadrature formula for the integral on the right hand side of Eq. (21), which is robust with respect to the lattice distance $|\mathbf{a}-\mathbf{b}|$ of the wavelets. Such kind of quadrature formulas based on sinc function approximation, cf. Stenger's monograph [49], have been discussed by two of us in Ref. [37]. Recently it has been proven by Schwinger [50] that the so-called exponential quadrature formula

$$\int_0^\infty \mathcal{I}^{(\mathbf{p},\mathbf{q})}(t,\mathbf{a})dt \approx h\sum_{m=-M}^{M} e^{mh}\mathcal{I}^{(\mathbf{p},\mathbf{q})}(e^{mh},\mathbf{a}), \tag{25}$$

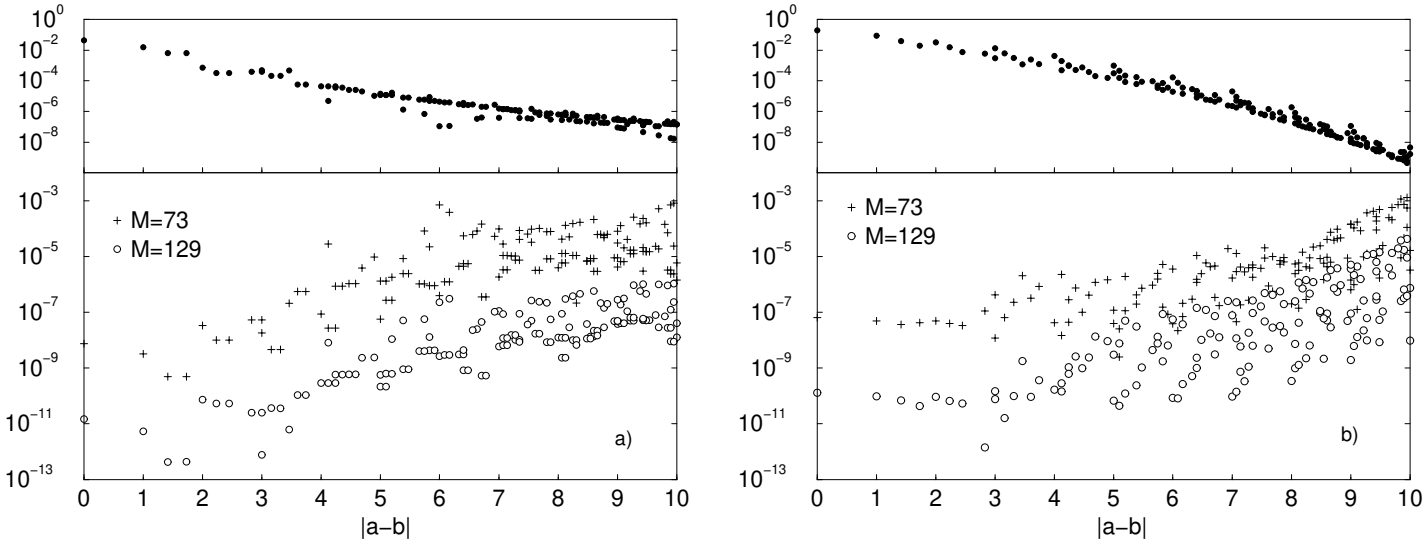

Figure 3: Relative error of elementary Coulomb integrals (25) for different number of quadrature points. The upper part shows the corresponding absolute values of the integrals. a) $\mathbf{p} = (0,0,0)$, $\mathbf{q} = (1,0,0)$. b) $\mathbf{p} = (1,1,1)$, $\mathbf{q} = (1,1,1)$. Isotropic 3d-wavelets were generated from the univariate SDD6 wavelet basis.

for $h = \sqrt{\frac{\pi\delta}{M}}$ ($\mathbf{p}$ or $\mathbf{q} \neq \mathbf{0}$) or $h = \sqrt{\frac{2\pi\delta}{M}}$ ($\mathbf{p} = \mathbf{q} = \mathbf{0}$), with $\delta < \frac{\pi}{4}$, provides a uniform error bound concerning the translation parameter $\mathbf{a}$. With this the integration error decreases almost exponentially with respect to the number of quadrature points

$$\left| \int_0^\infty \mathcal{I}^{(\mathbf{p},\mathbf{q})}(t,\mathbf{a})dt - h \sum_{m=-M}^{M} e^{mh} \mathcal{I}^{(\mathbf{p},\mathbf{q})}(e^{mh},\mathbf{a}) \right| \leq C_{\mathbf{p},\mathbf{q}} e^{-\alpha\sqrt{M}}, \tag{26}$$

where the constant α is given by $2\sqrt{\pi\delta}$ and $\sqrt{2\pi\delta}$ for $\mathbf{p}$ or $\mathbf{q} \neq \mathbf{0}$ and $\mathbf{p} = \mathbf{q} = \mathbf{0}$, respectively. The different behaviour is due to the vanishing moment property of wavelets which forces the function $\mathcal{I}^{(\mathbf{p},\mathbf{q})}(t,\mathbf{a})$ to vanish in the $\lim t \to 0$, whereas for pure scaling functions it converges to a finite nonzero value. Actually it turned out that the estimate (26) is valid for a large class of univariate wavelet basis $\psi^{(0)}$, $\psi^{(1)}$ for which the convolution product decays exponentially, i.e.

$$\left| \int \psi^{(p)}(x-y)\, \psi^{(q)}(y)\, dy \right| \leq C_{p,q} e^{-c|x|} \quad \text{for } c > 0. \tag{27}$$

This is trivially the case for any wavelet with compact support and is also satisfied e.g. for the Mexican hat wavelet which corresponds to the second derivative of a Gaussian function. In Fig. 3 we have shown relative errors of the integrals (25) for two different 3d-wavelet combinations $\mathbf{p} = (0,0,0)$, $\mathbf{q} = (1,0,0)$ and $\mathbf{p} = (1,1,1)$, $\mathbf{q} = (1,1,1)$ which correspond to two extreme cases with a univariate wavelet in one direction only in the first case and univariate wavelets in all directions in the second case. The underlying SDD6 wavelet (see Appendix A for further details) has six vanishing moments. In accordance with estimate (24), the absolute values of the second kind of integrals decay much faster with respect to their distance on the lattice because of the larger number of vanishing moments acting in all directions. Even so the constant in the estimate (26) might depend on $\mathbf{p}$, $\mathbf{q}$, it can be seen from Fig. 3 that the relative errors are rather similar for both cases.

With the uniformly stable quadrature formula (25) at our disposal, we can rearrange Eq. (20) according to

$$\int\int \eta_\alpha(\mathbf{x}) \frac{1}{\mid \mathbf{x} - \mathbf{y} \mid} \eta_\beta(\mathbf{y})\, d\mathbf{x} d\mathbf{y}$$

$$\approx \frac{2h}{\sqrt{\pi}} \sum_{m=-M}^{M} e^{mh} \sum_{j=j_0}^{j_{max}} 2^{-2j} \sum_{\mathbf{p},\mathbf{q}}{}' L^{(1)}_{j,p_1,q_1}(e^{mh})\, L^{(2)}_{j,p_2,q_2}(e^{mh})\, L^{(3)}_{j,p_3,q_3}(e^{mh}), \tag{28}$$

where the order of integration has been reversed to obtain a separable expression through the quantities

$$L^{(i)}_{j,p_i,q_i}(t) = \sum_{a \in \Lambda^{(i,\alpha)}_{j,p_i}} \sum_{b \in \Lambda^{(i,p_i,q_i)}_{j,a}} \langle \eta^{(i)}_{\alpha} | \tilde{\psi}^{(p_i)}_{j,a} \rangle \langle \tilde{\psi}^{(q_i)}_{j,b} | \eta^{(i)}_{\beta} \rangle\, G^{(p_i,q_i)}(a-b,t). \tag{29}$$

The index sets $\Lambda^{(i,\alpha)}_{j,p_i}$, $\Lambda^{(i,\beta)}_{j,q_i}$ result from thresholds C_j on the absolute values of the wavelet (scaling function) coefficients, e.g.

$$\Lambda^{(i,\alpha)}_{j,p_i} := \left\{ a : |\langle \eta^{(i)}_{\alpha} | \tilde{\psi}^{(p_i)}_{j,a} \rangle| \geq C_j \right\}. \tag{30}$$

In order to perform the second sum with respect to b in a more efficient way we have had to recourse to the estimate (24) from which further bounds $C^{(p_i,q_i)}_j$ for the index sets

$$\Lambda^{(i,p_i,q_i)}_{j,a} := \left\{ b : |a-b| \leq C^{(p_i,q_i)}_j \right\} \cap \Lambda^{(i,\beta)}_{j,p_i} \tag{31}$$

can be derived. These bounds do not depend on the parameter t and represent asymptotic properties of the Coulomb potential which are not necessarily reflected by individual components of the decomposition (28). Alternatively, it is possible to use individual bounds for the functions (23) at each quadrature point. Our decomposition of the integral (28) resembles to a recently proposed method of Sundholm, cf. Ref.[51], who used interpolating Lagrange functions instead of wavelets.

In our envisaged applications, the integrals (20) have to be calculated for a large set of functions $\{\eta_\nu\}$. In such a case it makes sense to introduce further compression steps. For this we consider the convolution

$$\int \frac{1}{\mid \mathbf{x}-\mathbf{y} \mid} \eta_\beta(\mathbf{y}) d\mathbf{y} \approx \frac{2h}{\sqrt{\pi}} \sum_{j=j_0}^{j_{max}} 2^{-2j} \sum_{m=-M}^{M} e^{mh} \sum_{\mathbf{p},\mathbf{q}}{}' \prod_{i=1}^{3} \left(\sum_a M^{(i,p_i,q_i)}_{j,a}(e^{mh})\, \tilde{\psi}^{(p_i)}_{j,a}(x_i) \right), \tag{32}$$

with

$$M^{(i,p_i,q_i)}_{j,a}(t) = \sum_{b \in \Lambda^{(i,p_i,q_i)}_{j,a}} \langle \tilde{\psi}^{(q_i)}_{j,b} | \eta^{(i)}_{\beta} \rangle\, G^{(p_i,q_i)}(a-b,t), \tag{33}$$

which represents an intermediary step in the calculation of the integrals (20). Using the fast wavelet transform (14), Eq. (32) can be rewritten in such a way that each component, originally expressed in terms of scaling functions, is represented in the wavelet basis, i.e.

$$\sum_a M^{(i,0,q_i)}_{j,a}(t)\, \tilde{\psi}^{(0)}_{j,a}(x_i) = \sum_{l=j_0}^{j} \sum_s{}' \sum_a N^{(i,0,q_i)}_{j,s,l,a}(t) \tilde{\psi}^{(s)}_{l,a}(x_i). \tag{34}$$

Again, a prime indicates that scaling functions appear only on the coarsest level j_0 of the expansion. Actually the sum over l runs up to $j-1$, however, for the sake of a consistent notation we have augmented the sum up to j with $N^{(i,0,q_i)}_{j,s,j,a}(t) := 0$. The wavelet components remain unchanged in the new representation, i.e.

$$N^{(i,1,q_i)}_{j,s,l,a}(t) := \begin{cases} M^{(i,1,q_i)}_{j,a}(t) & \text{for } l=j \text{ and } s=1 \\ 0 & \text{otherwise} \end{cases}. \tag{35}$$

With this representation it is possible to compress the tensor product on each level j separately

$$\frac{2h}{\sqrt{\pi}} \sum_{m=-M}^{M} e^{mh} \sum_{\mathbf{p},\mathbf{q}}{}' \prod_{i=1}^{3} \left(\sum_{l=j_0}^{j} \sum_s{}' \sum_a N^{(i,p_i,q_i)}_{j,s,l,a}(e^{mh}) \tilde{\psi}^{(s)}_{l,a}(x_i) \right) \approx\!\!\approx \sum_{k=1}^{\kappa_j} \prod_{i=1}^{3} \left(\sum_{l=j_0}^{j} \sum_s{}' \sum_a P^{(i,s)}_{k,j,l,a}\, \tilde{\psi}^{(s)}_{l,a}(x_i) \right). \tag{36}$$

Here and in the following $\approx$ denotes a best separable approximation of a tensor product. Following this intermediate compression step, the different levels can be compressed together into an even more compact expression

$$\int \frac{1}{\mid \mathbf{x}-\mathbf{y} \mid} \eta_\beta(\mathbf{y}) d\mathbf{y} \approx \sum_{j=j_0}^{j_{max}} 2^{-2j} \sum_{k=1}^{\kappa_j} \prod_{i=1}^{3} \left(\sum_{l=j_0}^{j} {\sum_{s}}' \sum_{a} P_{k,j,l,a}^{(i,s)} \tilde{\psi}_{l,a}^{(s)}(x_i) \right) \tag{37}$$

$$\approx \sum_{k=1}^{\kappa} \prod_{i=1}^{3} \left(\sum_{l=j_0}^{j_{max}} {\sum_{s}}' \sum_{a} Q_{k,l,a}^{(i,s)} \tilde{\psi}_{l,a}^{(s)}(x_i) \right). \tag{38}$$

Whether it is worthwhile to perform the additional compression step (38) depends on the specific application. If the finest resolution level of the functions η_α vary considerably, as it is typically the case for all-electron calculations, it might be favourable to use the multiscale representation (37) where the resolution levels can be chosen adaptively. Whereas in the case of pseudopotential calculations the total number of resolution levels $j_{max} - j_0 + 1$ remains small and representation (38) becomes more efficient. With the compressed coefficients $P_{k,j,l,a}^{(i,s)}$ or $Q_{k,l,a}^{(i,s)}$ at hand we can now efficiently evaluate Coulomb integrals (20) in the tensor product format, e.g.

$$\int\int \eta_\alpha(\mathbf{x}) \frac{1}{\mid \mathbf{x}-\mathbf{y} \mid} \eta_\beta(\mathbf{y}) d\mathbf{x} d\mathbf{y} \approx \sum_{k=1}^{\kappa} \prod_{i=1}^{3} \left(\sum_{l=j_0}^{j_{max}} {\sum_{s}}' \sum_{a} Q_{k,l,a}^{(i,s)} \langle \eta_\alpha^{(i)} | \tilde{\psi}_{l,a}^{(s)} \rangle \right), \tag{39}$$

where further usage can be made of sparsity features of the wavelet coefficients $\langle \eta_\alpha^{(i)} | \tilde{\psi}_{l,a_i}^{(s_i)} \rangle$.

3 Density fitting scheme with best separable approximations

A major bottleneck for the numerical solution of Kohn-Sham equations is the evaluation of the Hartree potential for a given electron density ρ. Originally introduced by Baerends, Ellis and Ros, cf. Ref. [52], it became a widespread approach within DFT to approximate the electron density in an auxiliary basis set in order to facilitate this part of the computation. Following Sambe and Felton, cf. Ref. [53], the most popular choice are GTO auxiliary basis sets [54, 55, 56, 57, 58, 59, 60, 61, 62] for which the Hartree potential can be evaluated analytically. Such kind of approach can be considered as a separable approximation of the electron density in the sense of Eq. (3), where the separation rank K corresponds to the total number of primitive GTOs in the auxiliary basis set. These basis sets have been carefully optimized for individual elements [56, 58, 61], and therefore it is interesting to compare their separation ranks with those obtained from best separable approximations of comparable accuracy. Most of the data reported in the literature is related to errors in energy resulting from this approximation. Therefore such a comparison is postponed to Section 4, where we discuss our results for the Hartree potential. Within this section, we focus instead on local errors of best separable approximations for the electron density

$$\rho(\mathbf{x}) \approx \sum_{k=1}^{\kappa} \varrho_k^{(1)}(x_1)\, \varrho_k^{(2)}(x_2)\, \varrho_k^{(3)}(x_3). \tag{40}$$

According to Eq. (2), the approximation error has been minimized with respect to the L^2 norm. The same measure has been chosen in the ADF program of Baerends and coworkers [52, 65]. However most of the existing DFT codes follow Dunlap et al., cf. Ref. [54], by minimizing the Coulomb repulsion of the residual. It has been argued in Refs. [54, 66] that this results in a quadratic error for the energy. Different types of interaction potentials have been discussed in Refs. [67, 68, 69]. From a computational point of view, the Coulomb repulsion does not provide a convenient measure for tensor product expansions. Nevertheless there exist several possible alternatives, e.g. weighted L^2 norms, which bound the Coulomb

repulsion from above, or different types of Sobolev norms. Among the Sobolev spaces, H^{-1} seems to be the most promising because it can be considered as a modification of the Coulomb repulsion where the Coulomb potential has been replaced by a Yukawa potential. This resembles closely to the attenuated Coulomb potential considered in Ref. [68]. However in order to be compatible with the tensor product format, the H^{-1} norm has to be taken in momentum space which requires the Fourier transform $\tilde{\rho}$ of the electron density. Because of the tensor product structure this can be accomplished with moderate effort by performing fast Fourier transformations for the univariate components. In momentum space $\mathbf{p}$, the transition from the Coulomb repulsion to a H^{-1} norm simply corresponds to replacing the weakly singular $|\mathbf{p}|^{-2}$ by a finite weight factor with positive parameter α, i.e.

$$\sigma_\kappa^{H^{-1}}(\rho) := \inf_{\varrho_k^{(i)} \in H^{-1}(\mathbb{R})} \int d^3p \, \frac{\left| \tilde{\rho}(\mathbf{p}) - \sum_{k=1}^{\kappa} \tilde{\varrho}_k^{(1)}(p_1)\, \tilde{\varrho}_k^{(2)}(p_2)\, \tilde{\varrho}_k^{(3)}(p_3) \right|^2}{\alpha + |\mathbf{p}|^2}. \tag{41}$$

Given a short tensor product approximation of the additional weight factor $(\alpha + |\mathbf{p}|^2)^{-1}$, e.g. in terms of Gaussians, it can be easily incorporated into the compression algorithm. For simplicity, we restricted ourselves within the present work to the L^2 norm. It will be the subject of our future work to perform comparative studies for different norms.

In our numerical studies we focused on the simple molecules CH_4, C_2H_2, and C_2H_6. The original densities have been obtained from all-electron HF calculations with standard VDZ basis sets. Using the closedness of GTO basis functions under multiplication, we obtained a representation for the electron density in the form of Eq. (3), which has been handled according to the procedure outlined in Section 1.2. In order to achieve a sufficient resolution of the electron-nuclear cusp at the C atoms, we have chosen in the preparatory step (5), scaling functions at level $j = 7$ which corresponds to a grid with $h = 2^{-7} \approx 0.008$ bohr. This has to be compared with the standard deviation of 0.0087 bohr for the GTO with the largest exponent in the VDZ basis set. It is obvious that the electron density obtained from a VDZ basis set is smooth everywhere. Nevertheless we use the notion of a cusp to discuss the behaviour of electron densities near nuclei. This can be justified in view of the fact that we consider a discretized space which is too coarse to resolve the difference between the exact and approximate VDZ electron density.

Best separable approximations have been obtained using the Newton algorithm with and without application of the trust-region method. Convergence of L^2 errors (2) with the Kronecker rank κ for the molecules mentioned above are shown in Fig. 4. It can be seen that for moderate compression ranks ≤ 40, the trust-region method leads to local minima with slightly larger L^2 errors. However there appears a rather sharp bend around $\kappa \approx 20$ without the trust-region method. Beyond that bend, at certain ranks, the Newton algorithm without trust-region method fails to converge to a minima at all. The global character of the L^2 error does not reveal local differences in the quality of separable approximations. For this we have plotted in Fig. 5 the local error of best separable approximations at various ranks of the electron density for C_2H_6. It can be seen that the error converges fast except at the nuclear cusp of C atoms. We want to point out that the scale of our plots changes by an order of magnitude between $\kappa = 15$ and $\kappa = 20$. An apparent asymmetry of the peaks at the C nuclei is due to the rather coarse resolution of the lattice ($h = 2^{-4}$ bohr) used for our plots. The C atom in the foreground is precisely located on a lattice site whereas the C atom in the background is somewhere in between. To get a better idea of the behaviour of best separable approximations near nuclear cusps, we have shown in Fig. 6 the actual electron density at the lattice sites near a C nucleus of C_2H_6, where $h = 2^{-7}$ bohr corresponds to the actual lattice distance used for the compression. It can be seen that there is obviously fast convergence at low ranks. However it turned out that this convergence got stuck around $\kappa \approx 20$ where the relative error at the nucleus became $4.2 \cdot 10^{-4}$. In the following we want to examine how errors in the electron density propagate into the Hartree potential and finally into the HF energy.

Various modifications of our approach seem to be possible in order to improve the approximation quality in the immediate neighbourhood of nuclei. The application of pseudopotentials is an obvious way to improve the efficiency. This possibility has been already studied for simple molecules but will not be

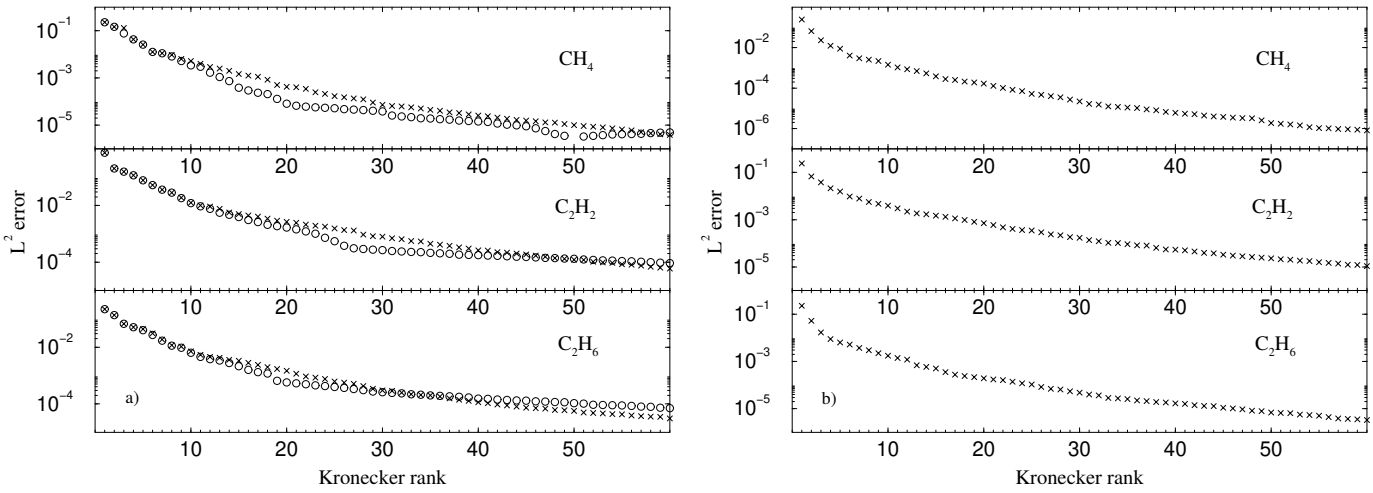

Figure 4: Relative L^2 error of best separable approximations of the electron density (a) and Hartree potential (b) for CH_4, C_2H_2, and C_2H_6. Results are shown for the Newton algorithm with ($\times$) and without ($\circ$) application of the trust-region method. The errors in (b) refer to initial Hartree potentials obtained from compressed electron densities at ranks $\kappa = 20$, 40, and 40 respectively. These densities were generated without application of the trust-region method.

pursued further in this paper, c.f. Ref. [63]. In the all-electron case, the most promising approach seems to be a combination of few precontracted GTO basis functions and best separable approximations, where GTOs can be used to approximate the electron density in the immediate neighbourhood of a nucleus. The contracted GTOs can be optimized for atomic densities either with respect to the total energy or any other norm which seems to be appropriate.

4 Approximation and compression of Hartree potential

4.1 Density fitting of Hartree potential

The Hartree potential for best separable approximations of the electron density

$$V_H(\mathbf{x}) = \sum_{k=1}^{\kappa} \int \frac{1}{\mid \mathbf{x} - \mathbf{y} \mid} \varrho_k^{(1)}(y_1)\, \varrho_k^{(2)}(y_2)\, \varrho_k^{(3)}(y_3) d\mathbf{y}, \tag{42}$$

has been calculated according to Eq. (32), where $M = 81$ has been chosen for the quadrature formula. Concerning the accuracy of this quadrature rule, we refer to our remarks below. From the preceding discussion it is clear that the leading errors of the electron density are located within the immediate neighbourhood of nuclei. Such kind of errors give rise to a corresponding error of the Hartree potential which is $O(1/r)$ with respect to the distance from the nuclei. Because of an exponential decay of the electron density, the Coulomb energy probes the Hartree potential mainly within short distances with respect to the nuclei. On this length scale, fluctuations of the error of the Hartree potential are small, as can be seen from Figs. 7 a) and c), where the error in a neighbourhood of a C atom is shown for C_2H_6. More interesting from a practical point of view is the error distribution with respect to the Coulomb matrix elements

$$J_{ab} = \int g_a(\mathbf{x}) g_b(\mathbf{x})\, V_H(\mathbf{x}) d\mathbf{x}, \tag{43}$$

calculated according to Eq. (28), in a GTO basis set. We have chosen for C and H uncontracted $9s4p1d$ and $4s1p$ GTO basis sets, respectively, with exponents taken from standard VDZ bases. The error of the

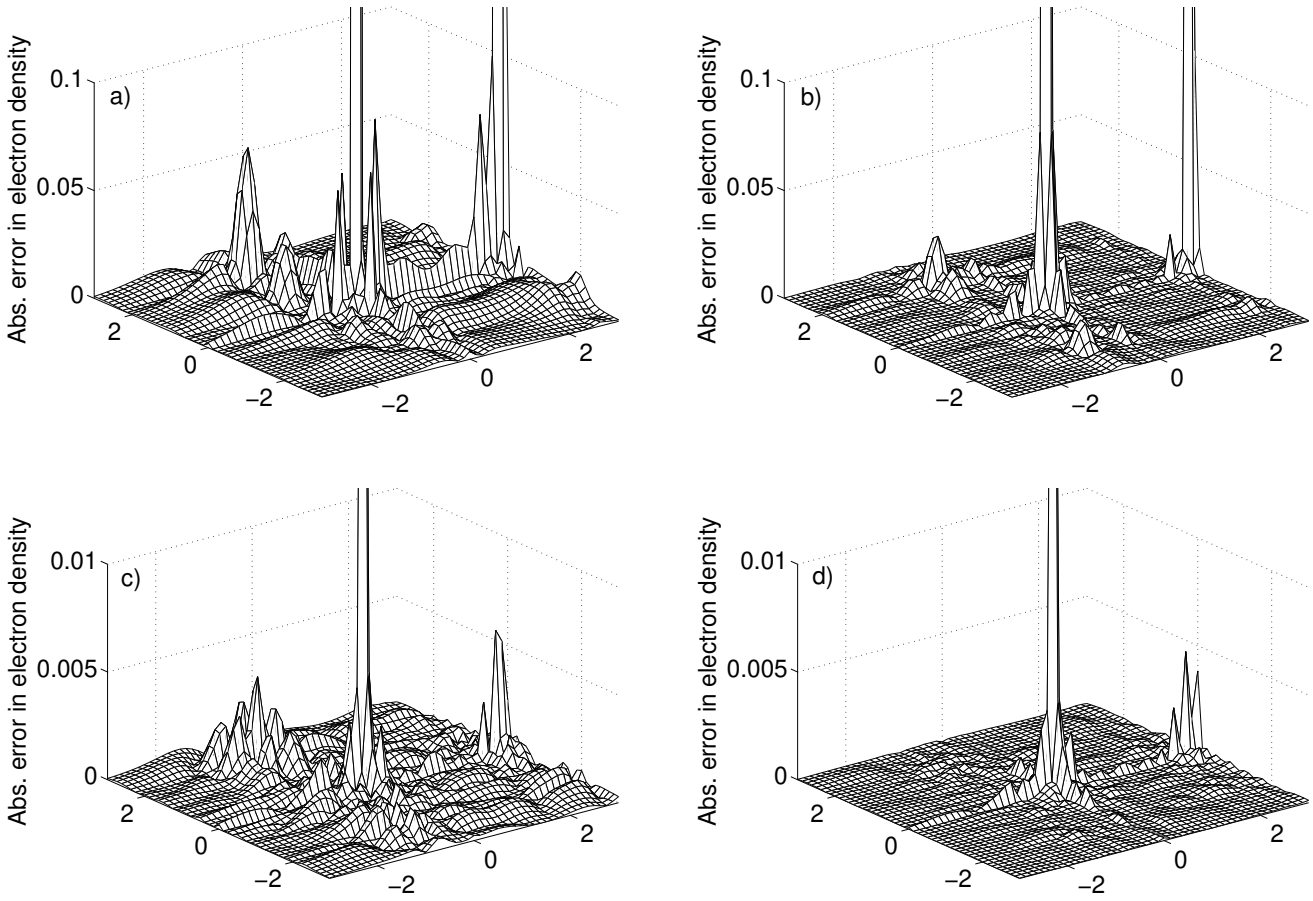

Figure 5: Absolute error of best separable approximations of the electron density for C_2H_6. The pictures show part of a plane containing both C atoms and a H atom. Different Kronecker ranks κ have been chosen for comparison. a) $\kappa = 10$, b) $\kappa = 15$, c) $\kappa = 20$, d) $\kappa = 40$.

Coulomb matrix elements is shown in Figs. 8 a) and c), where we observe a rather uniformly distributed error with slightly larger peaks along the diagonal.

We complete our discussion of the density fitting scheme by considering the accuracy of the corresponding total HF energies. To begin with, we have chosen a slightly simplified approach in order to avoid multiple compressions of the electron density within the iterative solution of the HF equation. Instead we have taken best separable approximations from already self-consistent electron densities and calculated the corresponding Coulomb matrix elements for the GTO basis set. In the following HF calculation, the Coulomb part of the Fock matrix was kept fixed and only the exchange part has been treated in a self-consistent manner. The HF calculations have been carried out with the *MOLPRO* program package [64] using the *MATROP* program for matrix operations. Before we enter into a discussion of the errors in HF energy, it remains to verify the accuracy of the exponential quadrature formula (25). For this we have calculated HF energies for CH_4 with different numbers of quadrature points $2M + 1$. The results are listed in Table 1. According to our accuracy requirements, an error of ≈ 0.1 mhartree is acceptable. This accuracy can be already achieved for $M = 64$, however in order to be on the safe side we have taken $M = 81$ in all of our calculations.

Errors in HF energy for best separable approximations of the electron density have been studied for the molecules CH_4, C_2H_2, and C_2H_6. Typically accuracies of $\approx 0.1 - 0.2$ mhartree per atom have been

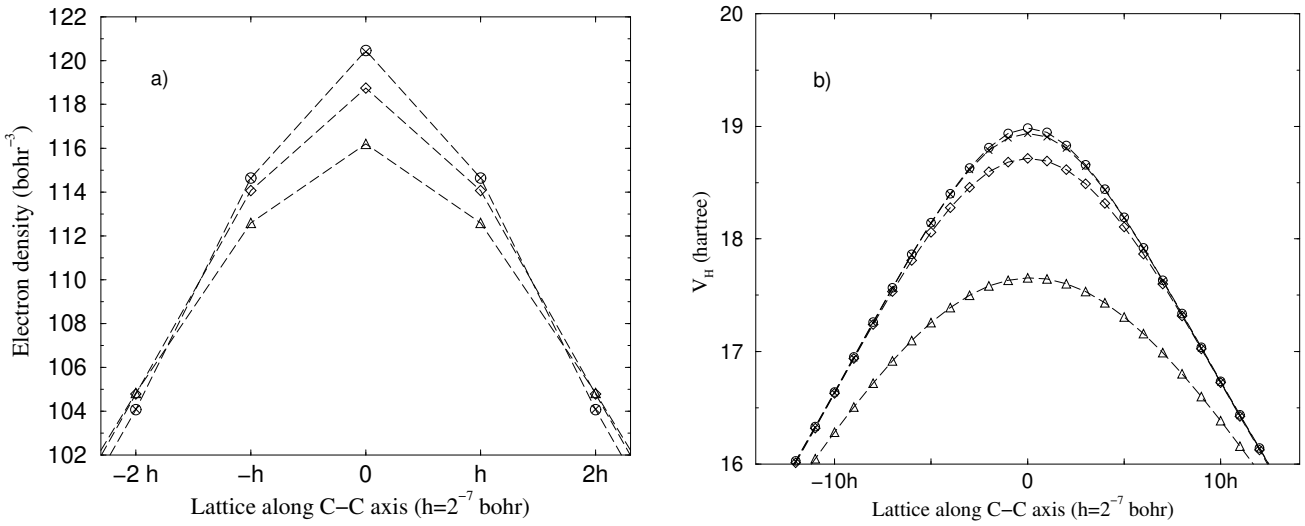

Figure 6: Electron density and Hartree potential V_H of C_2H_6 along the C-C lattice axis near a C nucleus located at the origin. Best separable approximations of a) electron density at compression ranks ($\triangle$) $\kappa = 10$, ($\diamond$) $\kappa = 15$, ($\times$) $\kappa = 20$, and b) density-fitted Hartree potential ($\kappa = 20$) at compression ranks, ($\triangle$) $\kappa' = 10$, ($\diamond$) $\kappa' = 20$, ($\times$) $\kappa' = 40$. For comparison the reference electron density and Hartree potential obtained from a VDZ basis set ($\circ$) is shown, respectively.

Table 1: Accuracy of exponential quadrature formula (25) for the Coulomb potential. Total HF energies (hartree) of CH_4 are given for different numbers $2M + 1$ of quadrature points and compression ranks of the electron density. Original HF energy for VDZ basis set -40.20266 hartree.

M	$\kappa = 20$	$\kappa = 35$	$\kappa = 45$
49	-40.20429	-40.20386	-40.20390
64	-40.20333	-40.20286	-40.20282
81	-40.20322	-40.20275	-40.20271

demanded in the literature, cf. Refs. [58, 61]. In the case of these simple molecules, Eichkorn et al., cf. Ref. [58], report total errors between 0.1 and 0.5 mhartree. We have to admit, however, that subsequent publications [62, 70, 71] discuss admissible errors which are at least one order of magnitude smaller. This has been achieved at the expense of doubling the size of the basis set compared with the initial Ref. [58]. The reason for these increased accuracy demands is the following post HF treatment where HF eigenvalues enter into the calculation. In view of the obvious shortcomming of our approach at nuclear cusps, which requires further improvements along the lines mentioned at the end of Section 3, we restrict ourselves within the present work to the modest requirements of Refs. [58, 61] adequate for pure HF or KS calculations. It can be seen from Table 2 that in order to achieve accuracies of $\approx 0.1 - 0.2$ mhartree per atom, the best separable approximations of the electron density require Kronecker ranks $\kappa \approx 20$, 45, and 40 for CH_4, C_2H_2, and C_2H_6, respectively. This has to be compared with the auxiliary basis sets of Ref. [58] which comprise 109, 122, and 186 primitive cartesian GTOs, respectively. Roughly speaking, our approach reduces the separation rank by a factor of three to five for these simple molecules. In the case of C_2H_6, the application of the trust-region method for the compression of the electron density, cf. Table 2, leads to considerable improvements at higher ranks. This indicates that the trust-region method becomes mandatory also for electron densities if high accuracies are required. Summing up, we may say that our results look encouraging. Despite various shortcommings of the L^2 error minimization (2), it is

possible to improve GTO based density fitting schemes. Problem specific modifications of the algorithm offer various possibilities for further significant improvements.

4.2 Best separable approximations of the Hartree potential

Reducing the computational complexity of the Hartree potential via density-fitting schemes is routinely applied for a long time in the case of HF and KS equations. Within the present work, we want to go one step further and consider best separable approximations of the Hartree potential itself. This can be considered as a special case of Eq. (38), where η corresponds to the electron density ρ. The initial Hartree potential is given by a convolution of the compressed density at rank κ with a separable approximation of the Coulomb potential (16) at rank $2M+1$. This yields an initial rank of $(2M+1)\kappa$, which has been compressed to final ranks κ' using the trust-region version of the Newton algorithm. It can be seen from Fig. 7 that compressed Hartree potentials and electron densities suffer from similar errors in the immediate neighbourhoods of nuclei. A more detailed picture presented in Fig. 6 b) shows that the error at the apex of the Hartree potential is smeared out compared to the cusp of the electron density in Fig. 6 a). The effect of compression on the corresponding Coulomb matrix (43) can be seen in Fig. 8 b) and d), where the error is sharply peaked at matrix elements of s-type GTOs with large exponents. Although these peaks seem to be rather large, it turns out that their effect on the error of HF energies is fairly small. This is due to small entries in the density matrix for such basis functions. In Table 2, we have listed errors in the HF energy with compressed Hartree potentials for the molecules CH_4, C_2H_2, and C_2H_6. It can be seen that the desired accuracy of $\approx 0.1 - 0.2$ mhartree per atom has been achieved for these molecules with ranks $\kappa' \approx 20$, 40, and 40, respectively, similar to the corresponding ranks for the electron density.

In order to perform the computation of the Hartree potential and a subsequent evaluation of Coulomb matrix elements in an efficient manner it is necessary to use sparse wavelet representations for univariate components of the tensor product. Such representations can be obtained through compression of the wavelet expansions (14) which can be simply achieved by neglecting coefficients smaller than a certain threshold. This is possible due to a norm equivalence between L^2 and ℓ^2 spaces discussed in Section 1.2. In Fig. 9 we have shown for the electron density and Hartree potential of C_2H_6 the percentage of wavelet coefficients on different levels of refinement with absolute value $\geq 10^{-8}$. To simplify the presentation, we have chosen best separable representations at rank $\kappa = \kappa' = 5$ which is comparatively small compared with the ranks discussed in the preceding paragraphs. Nevertheless this is sufficient, since we are mainly concerned in a qualitative discussion of the subject. It can be seen that in both cases, the fraction of significant wavelet coefficients decreases exponentially with the level of refinement, where the Hartree potential seems to allow for a higher compression rate.

5 Conclusions

Auxiliary basis sets for the approximation of the electron density and other quantities like products of orbitals turned out to be highly efficient to improve the performance of HF, KS and post HF methods. Typically GTO basis sets have been chosen for that purpose. We have proposed a new look on this subject from the point of view of "optimal" tensor product approximations. The basic idea behind the representation of certain quantities in terms of tensor products is to factorize expensive parts of the calculation in order to reduce the dimensionality and thereby the computational complexity. This has been demonstrated for the evaluation of Coulomb integrals. Starting from a given tensor product, e.g. a linear combination of GTOs, best separable approximations have been obtained using a Newton algorithm to minimize the L^2 error for a target Kronecker rank κ. In particular we have studied best separable approximations for electron densities and Hartree potentials of small molecules. It turned out that the Kronecker ranks of these tensor products compare favourably with conventional auxiliary GTO basis sets. Furthermore this approach can be merged with methods from multiresolution analysis. Especially wavelet bases seem to be an appropriate choice for the approximation of univariate components of tensor products

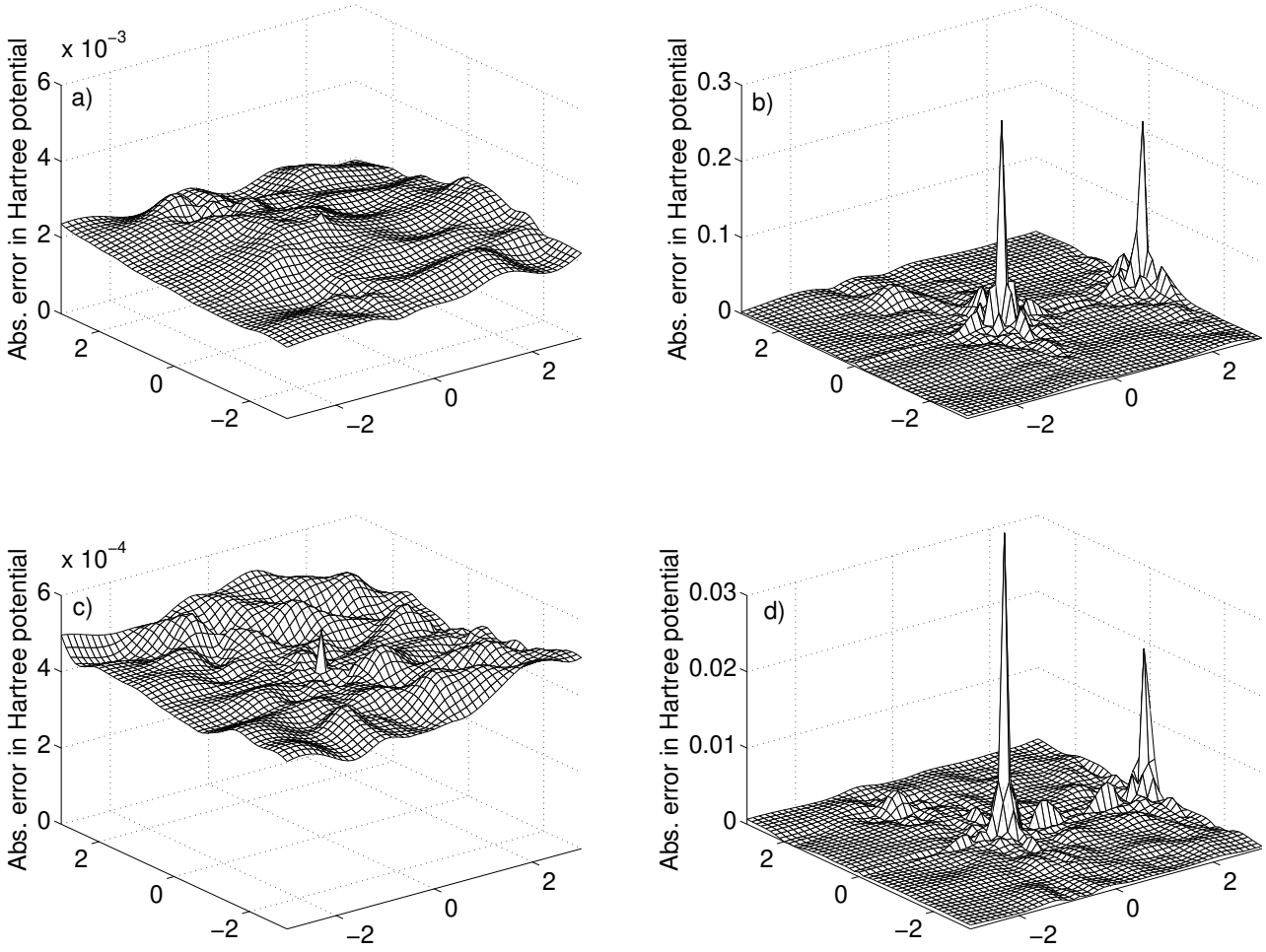

Figure 7: Local error of density fitted Hartree potentials V_H for C_2H_6. Absolute errors (hartree), before and after compression of the Hartree potential, are shown for best separable approximations of the electron density at different Kronecker rank κ. Figs. a) $\kappa = 20$, V_H before compression, b) $\kappa = 20$, V_H after compression to rank $\kappa' = 20$. c) $\kappa = 40$, V_H before compression. d) $\kappa = 40$, V_H after compression to rank $\kappa' = 40$. The pictures show a part of a plane containing both C atoms and a H atom.

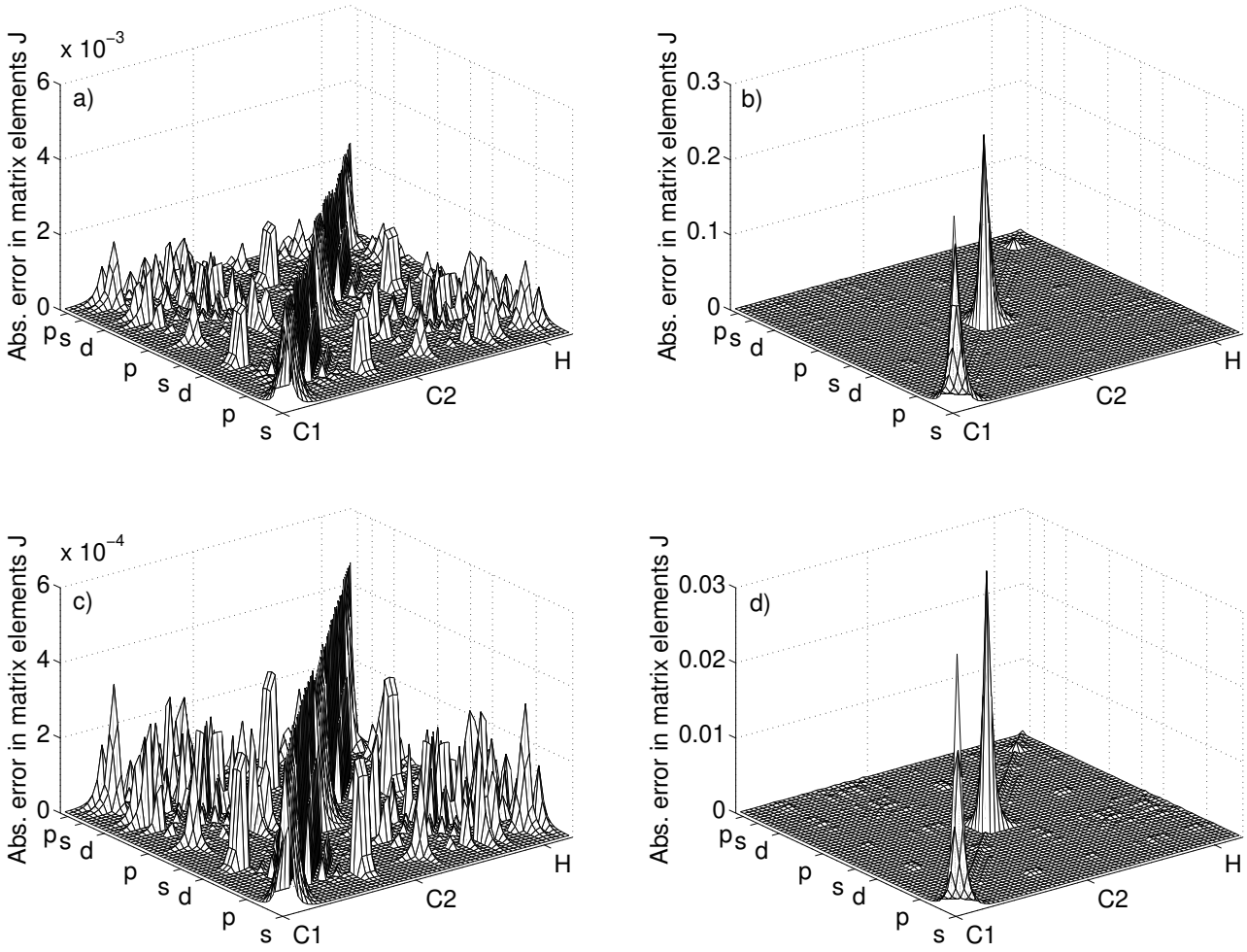

Figure 8: Error of Coulomb matrix elements J_{ab} with density fitted Hartree potential V_H for C_2H_6. The section of the Coulomb matrix shown here contains elements with respect to GTOs on both C and a H atom. GTOs are ordered with decreasing exponents for given l value. Absolute errors (hartree), before and after compression of the Hartree potential, are shown for best separable approximations of the electron density at different Kronecker rank κ. Figs. a) $\kappa = 20$, V_H before compression, b) $\kappa = 20$, V_H after compression to rank $\kappa' = 20$. c) $\kappa = 40$, V_H before compression. d) $\kappa = 40$, V_H after compression to rank $\kappa' = 40$.

Table 2: HF energies (mhartree) for best separable rank κ approximations of the electron density (E_κ) and with subsequent compression of the Hartree potential to rank κ' ($E_{\kappa'}$). Errors are given with respect to the HF energy for a VDZ GTO basis set (E_{GTO}). For comparison relative L^2 errors of the electron density (L^2_ρ) and Hartree potential ($L^2_{V_H}$) are given at each rank. Values given in parentheses correspond to best separable approximations of the electron density which have been obtained using the trust-region method.

	κ	$E_\kappa - E_{GTO}$	L^2_ρ	κ'	$E_{\kappa'} - E_{GTO}$	$L^2_{V_H}$
CH_4	20	-0.57	$8.1 \cdot 10^{-5}$	10	-39.50	$6.2 \cdot 10^{-4}$
				15	-1.40	$1.3 \cdot 10^{-4}$
				20	-0.63	$5.3 \cdot 10^{-5}$
	35	-0.09	$1.9 \cdot 10^{-5}$	10	-39.07	$6.2 \cdot 10^{-4}$
				15	-0.93	$1.3 \cdot 10^{-4}$
				20	-0.16	$3.9 \cdot 10^{-5}$
	45	-0.06	$9.0 \cdot 10^{-6}$	10	-63.16	$7.0 \cdot 10^{-4}$
				15	-4.86	$1.3 \cdot 10^{-4}$
				20	-0.12	$6.0 \cdot 10^{-5}$
C_2H_2	25	-3.56	$5.5 \cdot 10^{-4}$			
	30	-2.68	$2.7 \cdot 10^{-4}$			
	35	-2.41	$2.1 \cdot 10^{-4}$	30	-10.00	$1.8 \cdot 10^{-4}$
				40	-0.84	$6.4 \cdot 10^{-5}$
	40	-0.95	$1.8 \cdot 10^{-4}$	30	-4.35	$1.7 \cdot 10^{-4}$
				40	0.97	$5.3 \cdot 10^{-5}$
				50	-0.45	$2.3 \cdot 10^{-5}$
	45	-0.60	$1.5 \cdot 10^{-4}$			
	50	-0.50	$1.3 \cdot 10^{-4}$			
C_2H_6	30	-6.35	$2.6 \cdot 10^{-4}$	30	-7.14	$5.3 \cdot 10^{-5}$
		(-11.14)	$(3.0 \cdot 10^{-4})$		(-9.57)	$(4.5 \cdot 10^{-5})$
				40	-6.50	$1.8 \cdot 10^{-5}$
					(-11.25)	$(2.0 \cdot 10^{-5})$
				50	-6.45	$6.1 \cdot 10^{-6}$
					(-11.42)	$(6.9 \cdot 10^{-6})$
	40	-3.64	$1.6 \cdot 10^{-4}$	40	-3.87	$1.7 \cdot 10^{-5}$
		(-0.68)	$(1.2 \cdot 10^{-4})$		(-0.98)	$(1.6 \cdot 10^{-5})$
				50	-3.64	$6.8 \cdot 10^{-6}$
					(-0.95)	$(7.5 \cdot 10^{-6})$
				60	-3.73	$3.3 \cdot 10^{-6}$
					(-0.58)	$(3.3 \cdot 10^{-6})$
	45	(-0.17)	$(7.3 \cdot 10^{-5})$			
	50	-2.50	$1.1 \cdot 10^{-4}$			
	55	-0.40	$8.8 \cdot 10^{-5}$			

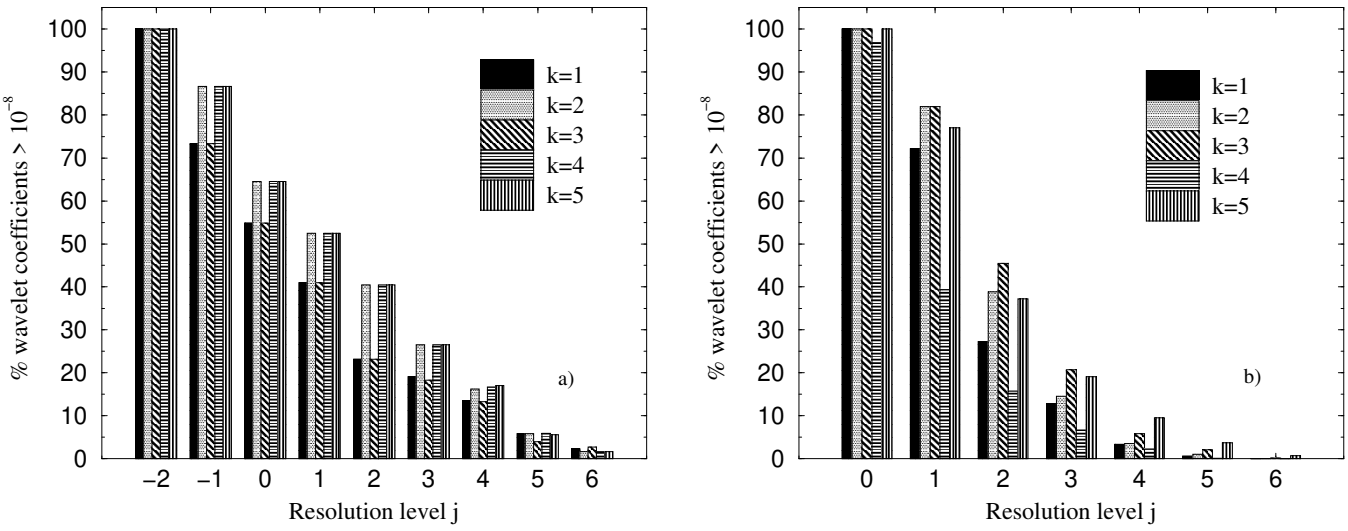

Figure 9: Wavelet compression of univariate tensor product components of the electron density and Hartree potential of C_2H_6 using Daubechies wavelets with five vanishing moments. a) Electron density at rank $\kappa = 5$. b) Hartree potential at rank $\kappa' = 5$.

which combines beneficial with a multiscale representation of the Coulomb interaction. Nevertheless, there seems to be still considerable room for further improvements. A detailed numerical analysis revealed that the largest approximation errors are typically located near the nuclei due to cusps in the electron density. These regions require special attention either by means of weighted L^2 norms or by predetermined terms in the tensor product which have been optimizied e.g. for individual atoms.

6 Acknowledgments

The authors gratefully acknowledge Dr. L. Grasedyck, and S. Schwinger (Leipzig) as well as Prof. R. Schneider (Kiel) for useful discussions. This work was supported by the Deutsche Forschungsgemeinschaft (SPP 1145).

Appendix

A Basic notions of multiresolution analysis

The purpose of this appendix is to provide some basic definitions and properties of wavelets which are required for this paper. For a complete exposition of this subject, we refer to the excellent monographs of Daubechies [72] and Mallat [73]. 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 $\varphi(x)$ provides a basis

$$\varphi_{j,a}(x) := 2^{j/2}\varphi(2^j x - a) \quad \text{with } a \in \mathbb{Z}, \tag{44}$$

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. An explicit embedding of

V_j into the larger space V_{j+1} is given by the refinement relation

$$\varphi(x) = 2\sum_a h_a\, \varphi(2x-a), \tag{45}$$

where the number of nonzero filter coefficients h_a is finite for the scaling functions used in our applications. 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. (44)

$$\psi_{j,a}(x) := 2^{j/2}\psi(2^j x - a) \quad \text{with } a \in \mathbb{Z}. \tag{46}$$

This construction leads to a hierarchical decomposition of $L^2 = \bigoplus_{j\in\mathbb{Z}} W_j$ into wavelet subspaces W_j [72]. Wavelets can be represented by scaling functions on the next finer level and vice versa a scaling function can be expressed in terms of wavelets and scaling functions on the next coarser level. In a biorthogonal wavelet basis there exists a sequence of dual spaces $\tilde{V}_j$, $\tilde{W}_j$, which satisfy the orthogonality relations $\tilde{W}_j \perp V_j$ and $\tilde{V}_j \perp W_j$. The corresponding dual wavelets $\tilde{\psi}_{j,a} := 2^{j/2}\tilde{\psi}(2^j x - a)$ and scaling functions $\tilde{\varphi}_{j,a} := 2^{j/2}\tilde{\varphi}(2^j x - a)$ provide a biorthogonal basis in L^2 i.e.

$$\langle \varphi_{j,a} | \tilde{\varphi}_{j,b}\rangle = \delta_{a,b}, \qquad \langle \psi_{j,a} | \tilde{\psi}_{l,b}\rangle = \delta_{j,l}\,\delta_{a,b}. \tag{47}$$

In the case of orthogonal wavelet basis, like Daubechies wavelets, the spaces V_j, W_j and their duals $\tilde{V}_j$, $\tilde{W}_j$ can be identified, i.e. $\varphi = \tilde{\varphi}$ and $\psi = \tilde{\psi}$.

An arbitrary function $f \in L^2(\mathbb{R})$ can be expanded in a biorthogonal wavelet basis

$$f(x) = \underbrace{\sum_a \langle \tilde{\varphi}_{j_0,a} | f\rangle\, \varphi_{j_0,a}(x)}_{V_{j_0}} + \underbrace{\sum_{j=j_0}^{\infty}\sum_a \langle \tilde{\psi}_{j,a} | f\rangle\, \psi_{j,a}(x)}_{\bigoplus_{j_0\le j} W_j}, \tag{48}$$

where the scaling function and wavelet coefficients are given by scalar products with respect to the dual basis. Depending on the specific choice of the wavelet basis, a certain number of moments for $\psi, \tilde{\psi}$ vanish, i.e.

$$\int dx\, x^k\, \tilde{\psi}(x) = 0, \text{ for } \; k = 0, \ldots, n-1. \tag{49}$$

This property has a significant effect on the magnitude of wavelet coefficients of sufficiently regular functions, i.e. $f \in C^n(\mathbb{R})$, as can be seen from their local Taylor series expansion

$$f(x) = c_0 + \cdots + c_{n-1}(x - 2^{-j}a)^{n-1} + R_{n-1}(x)(x-2^{-j}a)^n, \tag{50}$$

around the center of a wavelet $\tilde{\psi}_{j,a}(x)$. Inserting the Taylor series expansion (50) into the scalar product yields the following estimate for the wavelet coefficient

$$|v_{j,a}| = \left| \int dx\, f(x)\, \tilde{\psi}_{j,a}(x) \right| \le \sup_{\mathrm{supp}\{\tilde{\psi}_{j,a}\}} |R_{n-1}(x)|\; 2^{-j(n+1/2)} \int dx\, \left| x^n \tilde{\psi}(x) \right|, \tag{51}$$

where the supremum of the remainder $|R_{n-1}(x)|$ has to be taken with respect to the support of a wavelet $\tilde{\psi}_{j,a}(x)$. For functions with rapidly converging local Taylor series, the corresponding wavelet expansion (48), therefore, converges very fast with respect to the level j. This enables an efficient compression of the wavelet expansion by introducing an appropriate threshold for the wavelet coefficients, which leads to sparse wavelet representations for such functions. Because of the following norm equivalence, cf. Ref. [72], between $L^2(\mathbb{R})$ and ℓ^2 (sequence space of wavelet coefficients)

$$C_1 \|f\|^2_{L^2(\mathbb{R})} \le \sum_{j,a} \left| \langle \tilde{\psi}_{j,a} | f\rangle \right|^2 \le C_2 \|f\|^2_{L^2(\mathbb{R})}, \tag{52}$$

the compression step can be performed in a controlled manner.

For our numerical studies, we have used the lifted biorthogonal wavelets with six vanishing moments of Sweldens [74] and the corresponding scaling function of Deslauriers and Dubuc [25]. In the text, we refer to it as the SDD6 wavelet basis. Furthermore we have used orthogonal Daubechies wavelets with five vanishing moments [72], denoted as D10 wavelet basis according to the number of filter coefficients.

Any uniformly continuous function f can be approximated in a Deslauriers and Dubuc [25] interpolating scaling function basis $\varphi_{j,a}^{DD}$ through

$$f(x) \approx 2^{-j/2} \sum_a f(2^{-j}a)\, \varphi_{j,a}^{DD}(x), \tag{53}$$

which requires only pointwise evaluations of the function on a grid. These scaling functions have been constructed in such a way that the approximation (53) becomes exact for polynomials with degree $< n$. Provided that the n'th derivate of f is uniformly continuous, the following error estimate

$$\left\| f - 2^{-j/2} \sum_a f(2^{-j}a)\, \varphi_{j,a}^{DD} \right\|_{L^\infty(\mathbb{R})} \leq C\, 2^{-jn}, \tag{54}$$

can be obtained.

Sometimes it is convenient to switch between different wavelet basis. For example it is appropriate to approximate smooth functions in an interpolating scaling function basis, whereas the corresponding biorthogonal wavelet basis turns out be less convenient for certain computations within the tensor product format than orthogonal Daubechies wavelets. A transformation from Deslauriers and Dubuc φ^{DD} to Daubechies φ^{Daub} scaling functions can be easily accomplished through

$$\varphi_{j,a}^{DD} \approx \sum_b \langle \varphi_{j,a}^{DD} | \varphi_{l,b}^{Daub} \rangle \varphi_{l,b}^{Daub}, \tag{55}$$

where the level l has to be chosen large enough in order to make the approximation error sufficiently small. The coefficients $\langle \varphi_{j,a}^{DD} | \varphi_{l,b}^{Daub} \rangle$ can be easily evaluated via Beylkin's iterative approach [45].

References

[1] H. Yserentant, Numer. Math. **98**, 731 (2004); ibid. **101**, 381 (2005).

[2] J. Almlöf, in *Modern Electronic Structure Theory, Vol. II*, D.R. Yarkony, Ed., (World Scientific, Singapore, 1995) p. 110.

[3] M. Griebel and S. Knapek, Constr. Approx. **16**, 525 (2000).

[4] T. Gerstner and M. Griebel, Computing **71**, 65 (2003).

[5] H.-J. Bungartz and M. Griebel, Acta Numerica **13**, 147 (2004).

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

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

[8] H.-J. Flad, W. Hackbusch, and R. Schneider, ESAIM: M2AN **40**, 49 (2006).

[9] H.-J. Flad, W. Hackbusch, and R. Schneider, ESAIM: M2AN, to appear.

[10] C.F. Van Loan, J. Comp. Appl. Math. **123**, 85 (2000).

[11] L. De Lathauwer, B. De Moor, J. Vandewalle, SIAM J. Matrix Anal. Appl. **26**, 295 (2004).

[12] E. Tyrtyshnikov, Sb. Math. **194**, 941 (2003).

[13] E. Tyrtyshnikov, Linear Algebra Appl. **379**, 423 (2004).

[14] J.M. Ford and E. Tyrtyshnikov, SIAM J. Sci. Comput. **25**, 961 (2003).

[15] J.M. Ford and E. Tyrtyshnikov, Numer. Algorithms **40**, 125 (2005)

[16] W. Hackbusch, B.N. Khoromskij and E. Tyrtyshnikov, J. Numer. Math. **13**, 119 (2005).

[17] W. Hackbusch and B.N. Khoromskij, SIAM J. Matr. Anal. Appl., to appear.

[18] W. Hackbusch and B.N. Khoromskij, Preprint 139, MPI MIS, Leipzig 2006.

[19] A. Smilde, R. Bro and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences* (John Wiley & Sons, Chichester, 2004).

[20] G. Beylkin and M.J. Mohlenkamp, Proc. Natl. Acad. Sci. USA **99**, 10246 (2002).

[21] G. Beylkin and M.J. Mohlenkamp, SIAM J. Sci. Comput. **26**, 2133 (2005).

[22] I. Ibraghimov, Numer. Linear Algebra Appl. **9**, 551 (2002).

[23] B.N. Khoromskij, Comput. Methods Appl. Math. **6**, 194 (2006).

[24] P.D. Dacre and M. Elder, Mol. Phys. **22**, 593 (1971).

[25] G. Deslauriers and S. Dubuc, Constr. Approx. **5**, 49 (1989).

[26] L. Grasedyck, private communication.

[27] M. Espig, PhD thesis, MPI-MIS Leipzig (2007).

[28] T. Zhang and G.H. Golub, SIAM J. Matrix Anal. Appl. **23**, 534 (2001).

[29] E. Polak, *Optimization: Algorithms and Consistent Approximations* (Springer, New York, 1997).

[30] B.N. Khoromskij and V. Khoromskaia, Preprint 105, MPI MIS, Leipzig 2006.

[31] D. Braess, J. Approx. Theory **83**, 93 (1995).

[32] W. Kutzelnigg, Int. J. Quantum Chem. **51**, 447 (1994).

[33] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

[34] W. Hackbusch, Technical report MPI-MIS Leipzig **4** (2005).

[35] T.H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[36] G. Beylkin and L. Monzón, Appl. Comput. Harmon. Anal. **19**, 17 (2005).

[37] W. Hackbusch, B.N. Khoromskij, Computing **76**, 177 (2006).

[38] W. Hackbusch, Computing **76**, 359 (2006).

[39] R.J. Harrison, G.I. Fann, T. Yanai, Z. Gan and G. Beylkin, J. Chem. Phys. **121**, 11587 (2004).

[40] T. Yanai, G.I. Fann, Z. Gan, R.J. Harrison and G. Beylkin, J. Chem. Phys. **121**, 2866 (2004); ibid. **121**, 6680 (2004).

[41] BIGDFT is a wavelet based DFT program by T. Deutsch, S. Goedecker, X. Gonze and R. Schneider, with contributions from P.-M. Anglade, A. Bergman, D. Caliste, L. Genovese, A. Ghasemi, F. Krüger, F. Lancon, A. Neelov, J. Pivonski, and M. Rayson.

[42] L. Genovese, T. Deutsch, A. Neelov, S. Goedecker and G. Beylkin, arXiv:cond-mat/0605371v1.

[43] M. Dupuis, J. Rys and H.F. King, J. Chem. Phys. **65**, 111 (1976).

[44] G. Beylkin, R.R. Coifman and V. Rokhlin, Commun. Pure Appl. Math. **44**, 141 (1991).

[45] G. Beylkin, SIAM J. Numer. Anal. **6**, 1716 (1992).

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

[47] W. Dahmen, S. Prößdorf, R. Schneider, Math. Z. **215**, 583 (1994).

[48] W. Dahmen, S. Prößdorf, R. Schneider, Adv. Comp. Maths. **1**, 259 (1993).

[49] F. Stenger *Numerical Methods Based on Sinc and Analytic Functions* (Springer, New York, 1993).

[50] S. Schwinger, Preprint, MPI MIS, Leipzig 2007.

[51] D. Sundholm, J. Chem. Phys. **122**, 194107 (2005).

[52] E.J. Baerends, D.E. Ellis and P. Ros, Chem. Phys. **2**, 41 (1973).

[53] H. Sambe and R.H. Felton, J. Chem. Phys. **62**, 1122 (1975).

[54] B.I. Dunlap, J.W.D. Connolly and J.R. Sabin, J. Chem. Phys. **71**, 3396 (1979).

[55] J. Andzelm and E. Wimmer, J. Chem. Phys. **96**, 1280 (1992).

[56] N. Godbout, D.R. Salahub, J. Andzelm and E. Wimmer, Can. J. Chem. **70**, 560 (1992).

[57] V. Termath and N.C. Handy, Chem. Phys. Lett. **230**, 17 (1994).

[58] K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).

[59] C.-K. Skylaris, L. Gagliardi, N.C. Handy, A.G. Ioannou, S. Spencer, A. Willetts, J. Mol. Struct. (Theochem) **501-502** 229 (2000).

[60] F.R. Manby and P.J. Knowles, Phys. Rev. Lett. **87**, 163001 (2001).

[61] F.R. Manby, P.J. Knowles and A.W. Lloyd, J. Chem. Phys. **115**, 9144 (2001).

[62] R. Polly, H.-J. Werner, F.R. Manby and P.J. Knowles, Mol. Phys. *102*, 2311 (2004).

[63] S.R. Chinnamsetty, PhD thesis, MPI-MIS Leipzig 2007.

[64] MOLPRO is a package of ab initio programs written by H.-J. Werner and P.J. Knowles with contributions from R.D. Amos, A. Bernhardsson, A. Berning, P. Celani, D.L. Cooper, M.J.O. Deegan, A.J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, T. Korona, R. Lindh, A.W. Lloyd, S.J. McNicholas, F.R. Manby, W. Meyer, M.E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schütz, U. Schumann, H. Stoll, A.J. Stone, R. Tarroni and T. Thorsteinsson.

[65] C. Fonseca Guerra, J.G. Snijders, G. te Velde and E.J. Baerends, Theor. Chem. Acc. **99**, 391 (1998).

[66] B.I. Dunlap, Phys. Chem. Chem. Phys. **2**, 2113 (2000).

[67] P.M.W. Gill, B.G. Johnson, J.A. Pople and S.W. Taylor, J. Chem. Phys. **96**, 7178 (1992).

[68] Y. Jung, A. Sodt, P.M.W. Gill and M. Head-Gordon, Proc. Natl. Acad. Sci. USA **102**, 6692 (2005).

[69] P.M.W. Gill, A.T.B. Gilbert, S.W. Taylor, G. Friesecke, M. Head-Gordon, J. Chem. Phys. **123**, 061101 (2005).

[70] F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).

[71] F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).

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

[73] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, San Diego, 1998).

[74] W. Sweldens, Appl. Comput. Harmon. Anal. **3**, 186 (1996).