Concepts of Data-Sparse Tensor-Product Approximation in Many-Particle Modelling

Heinz-Jürgen Flad[†], Wolfgang Hackbusch[*], Boris N. Khoromskij[*], and Reinhold Schneider[†]

[*] Max-Planck-Institute for Mathematics in the Sciences
Inselstr. 22-26, D-04103 Leipzig, Germany
{wh, bokh}@mis.mpg.de

[†] Institut für Mathematik, Technische Universität Berlin
Straße des 17. Juni 137, D-10623 Berlin, Germany
{flad, schneidr}@math.tu-berlin.de

July 14, 2008

Abstract

We present concepts of data-sparse tensor approximations to the functions and operators arising in many-particle models of quantum chemistry. Our approach is based on the systematic use of structured tensor-product representations where the low-dimensional components are represented in hierarchical or wavelet based matrix formats. The modern methods of tensor-product approximation in higher dimensions are discussed with the focus on analytically based approaches. We give numerical illustrations which confirm the efficiency of tensor decomposition techniques in electronic structure calculations.

AMS Subject Classification: 65F30, 65F50, 65N35, 65F10

Key words: Schrödinger equation, Hartree-Fock method, density functional theory, tensor-product approximation

1 Introduction

Among the most challenging problems of scientific computing nowadays are those of high dimensions, for instance, multi-particle interactions, integral or differential equations on $[0,1]^d$ and the related numerical operator calculus for $d \geq 3$. Many standard approaches have a computational complexity that grows exponentially in the dimension d and thus fail because of the well known "curse of dimensionality". To get rid of this exponential growth in the complexity one can use the idea of tensor-product constructions (cf. [85]) on all stages of the solution process. Hereby we approximate the quantity of interest in tensor-product formats and use other approximation methods for the remaining low-dimensional components. Depending on the specific properties of the problem, these low-dimensional components are

already in a data-sparse format, like band structured matrices, or can be approximated via hierarchical (low-rank) matrix and wavelet formats, respectively. In order to obtain low-rank tensor-product approximations it is convenient to start already with a separable approximation of possibly large separation rank. This is the case e.g. for hyperbolic cross approximations in tensor-product wavelet bases or for Gaussian-type and plane wave basis sets which are frequently used in quantum chemistry and solid state physics. With such a representation at hand it is possible to apply algebraic recompression methods to generate the desired low-rank approximations. We want to stress, however, that these recompression methods in multilinear algebra lead to severe computational problems since they are, in fact, equivalent to some kind of nonlinear approximation in $d \geq 3$. Despite these computational difficulties, such kind of procedure is especially favourable for smooth functions with few singularities which are actually typical for our envisaged applications to be discussed below.

A large class of translation invariant kernels of integral operators can be represented via integral transformations of a separable function, e.g. Gaussian function. Using exponentially convergent quadrature rules for the parametric integrals it is possible to derive low-rank tensor-product approximations for these integral operators. In a similar manner it is possible to derive such representations for matrix-valued functions in the tensor-product format.

It is the purpose of the present paper to discuss possible applications of the afore outlined approach to electronic structure calculations with applications in quantum chemistry and solid state physics. It will be shown in the following how to combine the different techniques, which complement each other nicely, to provide a feasible numerical operator calculus for some standard many-particle models in quantum chemistry. Within the present work, we focus on the *Hartree-Fock* method and the *Kohn-Sham* equations of *density functional theory* (DFT). We present a brief survey on existing approximation methods, and give some numerical results confirming their efficiency. Our approach aims towards a numerical solution of the Hartree-Fock and Kohn-Sham equations with computational complexity that scales almost linearly in the number of particles (atoms). In particular, large molecular systems such as biomolecules, and nanostructures, reveal severe limitations of the standard numerical algorithms and tensor-product approximations might help to overcome at least some of them.

The rest of the paper is organised as follows. Section 2 gives a brief outline of electronic structure calculations and of the Hartree-Fock method in particular. This is followed by a discussion of best N-term approximation and its generalization to tensor product wavelet bases. We present an application of this approach to the Hartree-Fock method. In Section 4, we first introduce various tensor product formats for the approximation of functions and matrices in higher dimensions. Thereafter we consider a variety of methods to obtain separable approximations of multivariate functions. These methods center around the *Sinc* interpolation and convenient integral representations for these functions. Section 5 provides an overview on different data sparse formats for the univariate components of tensor products. Finally, we discuss in Section 6 possible applications of these tensor-product techniques in order to obtain linear scaling methods for Hartree-Fock and Kohn-Sham equations.

2 Basic principles of electronic structure calculations

The physics of *stationary states*, i.e. time harmonic, *quantum mechanical systems of N particles*, is completely described by a single *wave function*

$$(\mathbf{r}_1, s_1, ..., \mathbf{r}_N, s_N) \mapsto \Psi(\mathbf{r}_1, s_1, ..., \mathbf{r}_N, s_N) \in \mathbb{C} \ , \ \mathbf{r}_i \in \mathbb{R}^3 \ , \ s_i \in S \ ,$$

which is a function depending on the spatial coordinates $\mathbf{r}_i \in \mathbb{R}^3$ of the particles $i = 1, \dots, N$ together with their spin degrees of freedom s_i. Since identical quantum mechanical particles, e.g. electrons, cannot be distinguished, the wave function must admit a certain symmetry with respect to the interchange of particles. The *Pauli exclusion principle* states that for electrons, the spin variables can take only two values $s_i \in S = \{\pm\frac{1}{2}\}$, and the wave function has to be antisymmetric with respect to the permutation of particles

$$\Psi(\mathbf{r}_1, s_1, \dots, \mathbf{r}_i, s_i, \dots \mathbf{r}_j, s_j, \dots, \mathbf{r}_N, s_N) = -\Psi(\mathbf{r}_1, s_1, \dots, \mathbf{r}_j, s_j, \dots \mathbf{r}_i, s_i, \dots, \mathbf{r}_N, s_N) \ .$$

The *Born Oppenheimer approximation* considers a quantum mechanical ensemble of N electrons moving in an exterior electrical field generated by the nuclei of K atoms. Therein the wave function is supposed to be a solution of the stationary electronic Schrödinger equation

$$\mathbf{H}\Psi = E\Psi \ ,$$

with the many-particle Schrödinger operator (non-relativistic Hamiltonian) $\mathbf{H}$ given by

$$\mathbf{H} := -\frac{1}{2}\sum_{i=1}^{N}\Delta_i - \sum_{a=1}^{K}\sum_{i=1}^{N}\frac{Z_a}{|\mathbf{r}_i - \mathbf{R}_a|} + \sum_{i<j\leq N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{a<b\leq K}\frac{Z_a Z_b}{|\mathbf{R}_a - \mathbf{R}_b|}, \tag{2.1}$$

where $Z_a, \mathbf{R}_a$ are charges and positions of the nuclei, respectively.

The basic problem in *wave function methods* is to calculate (approximately) the wave functions $\Psi(\mathbf{r}_1, s_1, ..., \mathbf{r}_N, s_N)$, $\mathbf{r}_i \in \mathbb{R}^3$, $s_i = \pm\frac{1}{2}$ as an eigenfunction of the non-relativistic Hamiltonian $\mathbf{H}$. The wave function can be assumed to be real valued, which we will pursue in the sequel, for sake of simplicity. Perhaps in case of translation symmetry, e.g. for bulk crystals, a complex setting might be helpful. One is mostly interested in the ground energy. Due to the Ritz or Courant-Fischer min-max principle [75] this problem can be casted in a variational formulation

$$E = \min\{\langle \mathbf{H}\Phi, \Phi\rangle \ : \langle \Phi, \Phi\rangle = 1\} \quad , \quad \Psi = \operatorname{argmin}\{\langle \mathbf{H}\Phi, \Phi\rangle \ : \langle \Phi, \Phi\rangle = 1\} \ . \tag{2.2}$$

The wave functions can be approximated by antisymmetric tensor products $\Psi = \sum c_k \Psi_k$ where each Ψ_k denotes a *Slater determinant*

$$\Psi_k(\mathbf{r}_1, s_1, \dots, \mathbf{r}_N, s_N) = \frac{1}{\sqrt{N!}}\det(\varphi_{k_i}(\mathbf{r}_{k_j}, s_{k_j}))_{i,j=1}^{N} \ . \tag{2.3}$$

Here the functions $\varphi_\nu : \mathbb{R}^3 \times \{\pm\frac{1}{2}\} \to \mathbb{R}$ are supposed to be pairwise orthogonal

$$\langle \varphi_\nu, \varphi_\mu\rangle = \sum_s \int_{\mathbb{R}^3} \varphi_\nu(\mathbf{r}, s)\varphi_\mu(\mathbf{r}, s)d^3r = \delta_{\mu,\nu} \ .$$

An approximation by a single Slater determinant (2.3)

$$\Psi \approx \Psi_{SL} := \frac{1}{\sqrt{N!}}\det(\varphi_i(\mathbf{r}_j, s_j))_{i,j=1}^{N} \ ,$$

which is a kind of rank one approximation by an antisymmetric tensor-product gives the Hartree-Fock energy functional

$$\mathcal{E}_{HF}(\varphi_1, \dots, \varphi_N) := \langle \mathbf{H}\Psi_{SL}, \Psi_{SL}\rangle \ .$$

Which inserted into (2.2) yields the following constraint minimization problem

$$E_{HF} = \min\{\mathcal{E}_{HF}(\varphi_1, \ldots, \varphi_N) = \langle \mathbf{H}\Phi_{SL}, \Phi_{SL}\rangle : \ \langle \varphi_i, \varphi_j\rangle = \delta_{i,j} \ , \ i,j = 1, \ldots, N\}$$

with N^2 constraint conditions $\langle \varphi_i, \varphi_j\rangle = \delta_{i,j}$, $i,j = 1, \ldots, N$.

An additional simplification can be made for even number of electrons, restricting pairs of orbitals with opposite spin to the same spatial behaviour $\varphi_i(\mathbf{r}, \frac{1}{2}) = \varphi_{N/2+i}(\mathbf{r}, -\frac{1}{2}) =: \phi_i(\mathbf{r})$, $i = 1, \ldots N/2$. This gives the so called *restricted Hartree-Fock* model for close shell systems. The corresponding Hartree-Fock energy functional can be calculated explicitly

$$\begin{aligned} \mathcal{E}_{HF}(\phi_1, \ldots, \phi_{N/2}) &= 2\sum_{i=1}^{N/2} \int\limits_{\mathbb{R}^3} \left(\frac{1}{2}|\nabla\phi_i(\mathbf{r})|^2 + V_c(\mathbf{r})|\phi_i(\mathbf{r})|^2 \right) d^3r \\ &+ 2\sum_{i=1}^{N/2}\sum_{j=1}^{N/2} \int\limits_{\mathbb{R}^3}\int\limits_{\mathbb{R}^3} \left(\frac{\phi_i(\mathbf{r})\phi_i(\mathbf{r})\phi_j(\mathbf{r}')\phi_j(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \frac{\phi_i(\mathbf{r})\phi_i(\mathbf{r}')\phi_j(\mathbf{r})\phi_j(\mathbf{r}')}{2|\mathbf{r}-\mathbf{r}'|} \right) d^3r' d^3r, \end{aligned} \tag{2.4}$$

where V_c denotes the Coulomb potential due to the nuclei. Variational calculus applied to the restricted Hartree-Fock energy functional

$$E_{HF} = \min_{\langle \phi_i, \phi_j\rangle = \delta_{i,j}} \mathcal{E}_{HF}(\phi_1, \ldots, \phi_{N/2})$$

yields the Hartree-Fock equations

$$\mathcal{F}\phi_i(\mathbf{r}) = \varepsilon_i\, \phi_i(\mathbf{r}), \tag{2.5}$$

with the Fock operator $\mathcal{F}$, as a necessary condition, that for a minimizer there exist pairwise orthogonal orbitals ϕ_i, $i = 1, \ldots, N/2$. Defining the *reduced one-electron spin density matrix*, or simply "density matrix" in the following, as the kernel function

$$\rho(\mathbf{r}, \mathbf{r}') := \sum_{i=1}^{N/2} \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}') \tag{2.6}$$

of the corresponding spectral projection operator P_ρ, and introducing the total electron density

$$n(\mathbf{r}) := 2\rho(\mathbf{r}, \mathbf{r}), \tag{2.7}$$

the Fock operator is given by

$$\mathcal{F}\phi(\mathbf{r}) = -\frac{1}{2}\Delta\phi(\mathbf{r}) + V_c(\mathbf{r})\, \phi(\mathbf{r}) + V_H\phi(\mathbf{r}) + (\mathcal{K}\phi)\,(\mathbf{r}), \tag{2.8}$$

with *Hartree potential*

$$V_H(\mathbf{r}) := \int\limits_{\mathbb{R}^3} \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^3r', \tag{2.9}$$

and *exchange operator*

$$(\mathcal{K}\phi)\,(\mathbf{r}) := -\int\limits_{\mathbb{R}^3} \frac{\rho(\mathbf{r}, \mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, \phi(\mathbf{r}') d^3r'. \tag{2.10}$$

This is a coupled nonlinear eigenvalue problem, since the Fock operator depends on the density matrix (2.6), where ϕ_i, $i = 1, \ldots, N/2$ are the eigenfunctions corresponding to the $N/2$ lowest eigenvalues $\epsilon_1 \leq \ldots \epsilon_{N/2} < \epsilon_{N/2+1} \leq \ldots$.

A rank one approximation as in the Hartree-Fock model seems to be a rather poor approximation to the wave function. One way to pursue is to use more than one Slater determinant to approximate the wave function. This results in methods like *configuration interaction* (CI), *multi configuration self-consistent-field* (MCSCF), *coupled cluster* (CC) methods and other. All these methods have in common that they require a much larger computational effort. As a general reference for these methods, we refer to the monograph [62].

In contrast to these wave function methods, DFT [74] tries to replace the linear, but high-dimensional Schrödinger operator by nonlinear but low dimensional partial differential equations with the same ground state energy and which look formally like the Hartree-Fock equations. The major difference is that in the Kohn-Sham equations, the nonlocal exchange term $\mathcal{K}$ is replaced by a local *exchange-correlation potential* V_x depending only on the electron density n. However the functional dependence turns out to be very complicated and is not known explicitly. The most simple form is called the *local density approximation* (LDA) [74] where the exchange correlation potential is a nonlinear function of the density

$$V_x(\mathbf{r}) = -\frac{1}{\pi}(3\pi^2 n(\mathbf{r}))^{\frac{1}{3}} + \text{correction terms}\,.$$

Electron correlation can be incorporated into the correction term, which is usually taken from quantum Monte Carlo calculations for a homogeneous electron gas. The corresponding Euler-Lagrange equations take the form (2.5) with Fock operator

$$\mathcal{F}\phi(\mathbf{r}) := -\frac{1}{2}\Delta\phi(\mathbf{r}) + \Big(V_c(\mathbf{r}) + V_H(\mathbf{r}) + V_x[n](\mathbf{r})\Big)\phi(\mathbf{r}).$$

Since there are no nonlocal terms, apparently the computation of these density functional models becomes more simpler and efficient as Hartree-Fock computations. Nevertheless, the exact functional is still unknown and all known approximate functionals have certain deficiencies. Therefore a model error is still intrinsical in all of these methods. It should be mentioned, that actually so-called hybride models, i.e. convex combinations of Kohn-Sham and Hartree-Fock models, have shown the best performance in benchmark calculations for molecules.

Usually the orbitals ϕ_i are approximated either by atomic centered basis functions e.g. *Gaussian-type orbitals* (GTO) or even harmonic polynomials (plane waves). A new alternative approach is based on multi-scale wavelet bases, cf. [2, 34, 43, 61] and references therein. Due to the Coulomb singularity of the nuclear potential, the orbitals are only smooth away from the nuclei. To get rid of these electron nuclear cusps the core electrons are added to smooth the nuclear potential. The Coulomb potential is thereby replaced by a pseudopotential and only the valence electrons are considered explicitly in an external field generated by the pseudopotentials [30].

2.1 Density matrix formulation of Hartree-Fock and DFT models

Solving the spectral problem corresponding to the Hartree-Fock equations (2.5) leads to $\mathcal{O}(N^3)$ complexity at least. In order to achieve algorithms which scale linear in the number of electrons N it therefore turns out to be necessary to circumvent the computation of the

eigenvalue problem. Instead it is possible to reformulate the Hartree-Fock method in terms of the density matrix (2.6). It has been already mentioned before that the nonlinear part of the Fock operator (2.8) can be expressed in terms of the density matrix. Furthermore it is not hard to see that the Hartree-Fock energy functional (2.4) can be rewritten as a functional of the density matrix $\mathcal{E}_{HF}(\rho)$ where the orthogonality constraints of the orbitals have to be replaced by the idempotency condition of the density matrix considered as a spectral projector, i.e.,

$$\int_{\mathbb{R}^3} \rho(\mathbf{r},\mathbf{r}'')\rho(\mathbf{r}'',\mathbf{r}')\,d^3r'' - \rho(\mathbf{r},\mathbf{r}') = 0.$$

The density matrix plays the key role in order to achieve linear scaling in Hartree-Fock and DFT methods. For instance it is well known that the density matrix exhibits exponential decay $\rho(\mathbf{r},\mathbf{r}') \sim \exp(-\lambda|\mathbf{r}-\mathbf{r}'|)$ for nonmetallic systems. This so-called shortsightedness of the density matrix [67] enables e.g. an efficient treatment of the nonlocal exchange term (2.10) in the Fock operator. Various computational schemes, entirely based on the density matrix, exist to perform Hartree-Fock or DFT calculations, cf. [4, 7, 69, 70, 72] and references therein. In Section 6, we discuss an approach using the *sign* function of an operator which seems to be especially suitable for our purposes.

3 Hyperbolic cross approximation in wavelet bases

The idea of *sparse grids* or *hyperbolic cross approximation* is based on the following observation. Let us consider a complete orthonormal basis in $L_2(\mathbb{R})$, $\{\psi_{l,k} : l \in \mathbb{N}\cup\{-1,0\},\ k \in \mathcal{I}_l\}$ with $\sharp\mathcal{I}_l \sim 2^l$, which obeys the following approximation property

$$\|u - \sum_{l=-1}^{L-1}\sum_{k} u_{l,k}\psi_{l,k}\| \le C_R 2^{-Ls}\|u\|_s \quad \text{with} \quad u_{l,k} = \langle\psi_{l,k}, u\rangle,$$

with respect to the Sobolev spaces $H^s(\mathbb{R})$, $s > 0$. Typical bases are e.g. wavelet bases or trigonometric polynomials $\psi_{l,k}(x) = e^{2\pi i(2^l+k)x}$, $l \ge 0$. The orthogonality can be relaxed to a biorthogonal setting, which also includes hierarchical nodal basis functions as originally proposed by Zenger. We refer to [49, 80] and references therein for the detailed exposition of this approach.

Setting $W_l := \text{span}\,\{\psi_{l,k} : k \in \mathcal{I}_l\}$, the tensor-products $\Psi_{\mathbf{l},\mathbf{k}}(\mathbf{x}) = \psi_{l_1,k_1}(x_1)\cdots\psi_{l_d,k_d}(x_d)$, $l_i \ge -1$, $k_i \in \mathcal{I}_{l_i}$, $i = 1,\ldots,d$, form an orthogonal basis in $L_2(\mathbb{R}^d)$. Taking the spaces $Z_L = \sum_{l_1+\cdots+l_d\le L} W_{l_1}\otimes\cdots\otimes W_{l_d}$, we then get $\dim Z_L = L^d 2^L$ and

$$\|u - u_{Z_l}\| \lesssim 2^{-Ls}\|u\|_{s,\ldots,s}.$$

This means that one obtains the convergence rate 2^{-Ls} with $L^d 2^L$ degrees of freedom, instead of the usual complexity 2^{Ld}. The price to pay is that one has to require a slightly higher regularity in terms of the mixed or tensor-product Sobolev norms $\|.\|_{s,\ldots,s}$ in $H^{s,\ldots,s} = \bigotimes_{i=1}^d H^s(\mathbb{R})$. Therefore, for functions satisfying this regularity requirement we can get rid of the curse of dimensions. Up to a logarithmic term, we can achieve asymptotically the same complexity as in one dimension. It has been proven by Yserentant [87] that eigenfunctions of the many-particle Schrödinger operator (2.1) belong to Sobolev spaces of mixed partial

derivatives. This result enables the construction of sparse grid approximations for the entire wavefunction [47, 48, 88].

Practically the sparse grid approach is limited because the Riesz constant C_R of the basis enters by C_R^d. Usually this becomes rather large with increasing d. Recent experiences demonstrate that hyperbolic cross approximations or sparse grid approximations can be applied successfully for a moderate number of dimensions $d \leq 10, ..., 30$. In the application we have in mind, namely the numerical solution of the Hartree-Fock or Kohn-Sham equation, the orbitals are functions in $\mathbb{R}^3$ and the operators have kernels in $\mathbb{R}^6$. This makes the sparse grid approach highly attractive for the present problem.

Hyperbolic cross approximations can be used also in an adaptive setting. Results for graded meshes have been obtained by Griebel et al. (c.f. [47]). Schwab and Nitsche (cf. [71]) have considered point singularities, and demonstrated that an adaptive sparse grid approximation works well. In fact, wavelet bases are highly advantageous for local adaptive approximation. This can be explained best in the framework of a *best N-term approximation* [16, 17, 20, 28]. For $\tau \leq 2$, the space $l_\tau^w(\mathcal{J})$ is the collection of sequences, respectively infinite vectors, $\mathbf{u} \in l_2(\mathcal{J})$, satisfying

$$\sharp \{\lambda \in \mathcal{J} : |u_\lambda| > \eta\} \lesssim \eta^{-\tau}$$

for all $\eta > 0$. The quantity

$$\|\mathbf{u}\|_{l_\tau^w(\mathcal{J})} := \|\mathbf{u}\|_{l_2(\mathcal{J})} + |\mathbf{u}|_{l_\tau^w(\mathcal{J})} \text{ with } |\mathbf{u}|_{l_\tau^w(\mathcal{J})} := \sup_{\eta>0}(\eta \, [\sharp \{\lambda \in \mathcal{J} : |u_\lambda| > \eta\}]^{1/\tau})$$

defines a quasi-norm in l_τ^w. Rearranging $\mathbf{u} \in l_2(\mathcal{J})$ by a non-increasing sequence $\mathbf{u}^* = (u_k^*)_{k\in\mathbb{N}}$, i.e. $|u_k^*| \leq |u_{k-1}^*|$, we have an alternative representation of this quasi-norm via

$$|\mathbf{u}|_{l_\tau^w(\mathcal{J})} = \sup_{k>0} (k^{1/\tau} \, |u_k^*|)$$

and, if $\tau < 2$,

$$\|\mathbf{u}\| = \|\mathbf{u}\|_{l_2(\mathcal{J})} \leq \|\mathbf{u}\|_{l_\tau^w(\mathcal{J})}.$$

The quantity

$$\sigma_N(\mathbf{u}) := \inf_{\sharp \operatorname{supp} \mathbf{v} \leq N} \|\mathbf{u} - \mathbf{v}\| = (\sum_{k>N} |u_k^*|^2)^{1/2}$$

denotes the error of the *best N-term approximation of* $\mathbf{u}$. An approximation $\mathbf{v}$ satisfying $\|\mathbf{u} - \mathbf{v}\| \lesssim \sigma_N(\mathbf{u})$ is given by $\mathbf{v}^*$ where $v_k^* = u_k^*$ if $k = 1, \ldots, N$ and $v_k^* = 0$, for $k > N$.

Proposition 3.1 [16, 21]

1. *For* $\mathbf{u} \in l_2(\mathcal{J})$ *and* $s > 0$ *the estimate* $\sigma_N(\mathbf{u}) \lesssim N^{-s}$ *holds if and only if* $\mathbf{u} \in l_\tau^w(\mathcal{J})$ *with* $\frac{1}{\tau} = s + \frac{1}{2}$ *and* $\sigma_N(\mathbf{u}) \lesssim N^{-s}\|\mathbf{u}\|_{l_\tau^w(\mathcal{J})}$.

2. *For* $0 < \tau < \tau' \leq 2$ *there holds*

$$l_\tau(\mathcal{J}) \subset l_\tau^w(\mathcal{J}) \subset l_{\tau'}(\mathcal{J}).$$

3. *The wavelet expansion of a function* $u = \sum_{l,k} u_{l,k}\psi_{l,k}$ *belongs to the Besov space* $B_{\tau,\tau}^s(\mathbb{R})$ *if and only if* $(u_{l,k}) \in l_\tau(\mathcal{J})$ *where* $\frac{1}{\tau} = s + \frac{1}{2}$.

In a recent paper, Nitsche [71] has shown that the quasi Banach spaces $l_\tau(\mathbb{Z}^d)$, $0 < \tau < 1$, are also tensor-product spaces.

Theorem 3.2 [71] *For* $0 < \tau \leq 2$ *there holds* $l_\tau(\mathbb{Z}^d) = \bigotimes_{i=1}^d l_\tau(\mathbb{Z})$.

In the case of a tensor-product wavelet expansion of a function $v = \sum v_\lambda \psi_\lambda$ in the Besov space $\bigotimes_{i=1}^d B^{s^*}_{\tau,\tau}(\mathbb{R})$ we therefore obtain the following rate for the best N-term approximation

$$\inf_{\sharp \operatorname{supp} \mathbf{v}_N \leq N} \|v - v_N\| \lesssim N^{-s} \|v\|_{\bigotimes_{i=1}^d B^{s^*}_{\tau,\tau}(\mathbb{R})}$$

for all $0 \leq s < s^*$. Note that due to $s < s^*$ no logarithmic term is present in the above estimate. A similar but more complicated result is true for the energy space $H^1(\mathbb{R}^3)$. Let us note that this means that asymptotically the number of degrees of freedoms required to obtain an accuracy ϵ behaves (almost) like the corresponding number for a one-dimensional problem.

In order to apply Nitsche's theorem to solutions of Hartree-Fock and Kohn-Sham equations it becomes necessary to establish an asymptotic smoothness property for the behaviour of the solutions at the atomic nuclei. This has been achieved in the Hartree-Fock case by applying the calculus of pseudodifferential operators on manifolds with conical singularities. Our results can be summarized in the following theorem:

Theorem 3.3 [35, 37] *The self-consistent-field solutions* ϕ_i *of the Hartree-Fock equations, obtained via the level-shifting algorithm, satisfy the asymptotic smoothness property*

$$|\partial^\beta_{\mathbf{x}} \phi_i(\mathbf{x})| \leq C_\beta |\mathbf{x} - \mathbf{A}|^{1-|\beta|} \qquad \textit{for} \;\; \mathbf{x} \neq \mathbf{A} \;\; \textit{and} \;\; |\beta| \geq 1$$

in a bounded neighbourhood $\Omega \subset \mathbb{R}^3$ *of a nucleus at* $\mathbf{A}$. *This is sufficient to ensure* $\phi_i \in \bigotimes_{i=1}^3 B^{s^*}_{\tau,\tau}(\Omega)$ *for each* $s^* > 0$ *and* $\frac{1}{\tau} = s^* + \frac{1}{2}$.

Here we have used the standard short-hand notation for mixed partial derivatives

$$\partial^\beta := \frac{\partial^{\beta_1}}{\partial x_1^{\beta_1}} \frac{\partial^{\beta_2}}{\partial x_2^{\beta_2}} \frac{\partial^{\beta_3}}{\partial x_3^{\beta_3}},$$

with the absolute value of the multi-index $|\beta| := \beta_1 + \beta_2 + \beta_3$. The iterative solutions of the Hartree-Fock equations to which the theorem applies refer to the so-called level-shifting algorithm with an appropriate initial guess. This algorithm corresponds to a fixed point iteration scheme for which convergence has been proven by Cancès and Le Bris [13]. It should be mentioned however that so far no proof exists that every solution of the Hartree-Fock equations can be obtained via the level-shifting algorithm. In view of (2.6), this theorem can be immediately extended to the density matrix.

Corollary 3.4 *The one-electron reduced density matrix* $\rho(\mathbf{x}, \mathbf{y})$ *belongs to* $\bigotimes_{i=1}^6 B^{s^*}_{\tau,\tau}(\Omega)$ *for each* $s^* > 0$ *and* $\frac{1}{\tau} = s^* + \frac{1}{2}$.

For further applications of the best N-term approximation to post Hartree-Fock methods, we refer to [36].

Since the univariate basis functions and therefore the tensor-products are fixed from the beginning, a sparse grid approximation will be by far not an optimal tensor-product approximation with respect to the separation rank. In general we expect that the separation rank

of an "optimal" tensor-product approximation [1] is much smaller than for a sparse grid approximation of comparable accuracy. For example, let $f, g \in H^s(\mathbb{R})$, and $\|f - f_L\| \lesssim 2^{-Ls}$, $\|g - g_L\| \lesssim 2^{-Ls}$. Then $F(x, y) = f(x)g(y)$ has separation rank $r_{opt} = 1$. The sparse grid approximation F_{Z_L} has the same approximation rate 2^{-Ls} and $\mathcal{O}(L^2 2^L)$ degrees of freedom, but a relatively large separation rank $L2^L >> 1$. We will see in a moment that the sparse grid approximation is not too bad. Because, to store both functions f_L and g_L with respect to the given basis requires $2 \cdot 2^L$ coefficients, whereas the sparse grid approximation requires $\mathcal{O}(L^2 2^L)$ nonzero coefficients in contrast to $\mathcal{O}(2^{dL})$ for the full product. Keeping in mind that a really optimal tensor-product approximation for $d > 2$ is still an unsolved problem, and in general it might be quite expensive, the sparse grids approximation is simple and cheap from the algorithmic point of view. It achieves also an almost optimal complexity for storage requirements. It is a trivial task to convert an "optimal" tensor-product representation into a sparse grid approximation. The opposite direction is a highly nontrivial task and requires fairly sophisticated compression algorithms.

It is worthwhile to mention that previous wavelet matrix compression approaches are based on some Calderón-Zygmund type estimates for the kernels. The sparse grid approximation is intimately related to *wavelet matrix compression* of integral operators with globally smooth kernels. The kernel functions of *Calderón-Zygmund operators* are not globally smooth. Nevertheless, it can be shown that they can be approximated within *linear* or *almost linear complexity* by means of wavelet Galerkin methods see e.g. [8, 17, 18, 19, 76], since they are smooth in the far field region. This result is proved, provided that the Schwartz kernel $K(\mathbf{x}, \mathbf{y})$ in $\mathbb{R}^d \times \mathbb{R}^d$ is approximated by tensor-product bases $\Psi \otimes \Psi$, where Ψ is an isotropic wavelet basis in $\mathbb{R}^d$. Recently developed fast methods like wavelet matrix compression and hierarchical matrices are working well for isotropic basis functions or isotropic clusters. Corresponding results for sparse grid approximations with $\bigotimes_{i=1}^{2d} \Psi^i$ have not been derived so far. Tensor-product bases in the framework of sparse grids do not have this geometric isotropy, which might spoil the efficiency of these methods. This is not the case for more general tensor-product approximations of these operators discussed in Sections 4.2.2 and 4.2.3 below. Therefore tensor-product approximations will provide an appropriate and efficient tool handling nonlocal operators acting on functions which are represented by means of tensor-product (sparse grid) bases. The development of such a tool will play a fundamental role for dealing with operators in high dimensions.

4 Toolkit for tensor-product approximations

The numerical treatment of operators in higher dimensions arising in traditional *finite element methods* (FEM) and *boundary element methods* (BEM) as well as in quantum chemistry, material sciences and financial mathematics all have in common the fundamental difficulty that the computational cost of traditional methods usually has an exponential growth in d even for algorithms with linear complexity $\mathcal{O}(N)$ in the problem size N (indeed, N scales exponentially in d as $N = n^d$, where n is the "one dimensional" problem size).

There are several approaches to remove the dimension parameter d from the exponent (cf. [5, 41, 49, 53, 57]). For the approximation of functions, such methods are usually based on

[1] It should be mentioned that in our applications at best almost optimal tensor-product approximations can be achieved. This is not of particular significance since we are aiming at a certain accuracy and small variations of the separation rank, required in order to achieve this accuracy, do not cause much harm.

different forms of the separation of variables. Specifically, a multivariate function $F : \mathbb{R}^d \to \mathbb{R}$ can be approximated in the form

$$F_r(x_1, ..., x_d) = \sum_{k=1}^{r} s_k \Phi_k^{(1)}(x_1) \cdots \Phi_{(k)}^{d}(x_d) \approx F,$$

where the set of functions $\{\Phi_k^{(\ell)}(x_\ell)\}$ can be fixed, like the best N-term approximation discussed in Section 3, or chosen adaptively. The latter approach tries to optimize the functions $\{\Phi_k^{(\ell)}(x_\ell)\}$ in order to achieve for a certain *separation rank* r at least the almost optimal approximation property. By increasing r, the approximation can be made as accurate as desired. In the case of globally analytic functions there holds $r = \mathcal{O}(|\log \varepsilon|^{d-1})$, while for analytic functions with point singularities one can prove $r = \mathcal{O}(|\log \varepsilon|^{2(d-1)})$ (cf. [53]).

In the following we want to give a short overview of various approaches to generate separable approximations with low separation rank. We first introduce in Section 4.1 two different tensor-product formats which have been used in the following. Section 4.2 provides a succint discussion of low rank tensor-product approximations of special functions, including the Coulomb and Yukawa potential, for which a certain type of "seperable" integral representation exists. This integral representation can be used to obtain separable approximations either by applying the *Sinc* approximation (Section 4.2.1) or directly through a best N-term approximation of exponential sums (Section 4.2.2).

4.1 Tensor-product representations in higher dimension

Let a d-th order tensor $\mathcal{A} = [a_{i_1 \dots i_d}] \in \mathbb{C}^{\mathcal{I}}$ be given, defined on the product index set $\mathcal{I} = I_1 \times ... \times I_d$. It can be approximated via the *canonical decomposition* (CANDECOMP) or *parallel factors* (PARAFAC) model (shortly, canonical model) in the following manner

$$\mathcal{A} \approx \mathcal{A}_{(r)} = \sum_{k=1}^{r} b_k V_k^{(1)} \otimes ..... \otimes V_k^{(d)}, \quad b_k \in \mathbb{C}, \tag{4.1}$$

where the Kronecker factors $V_k^{(\ell)} \in \mathbb{C}^{I_\ell}$ are unit-norm vectors which are chosen such that for a certain approximation only a minimal number r of components in the representation (4.1) are required. The minimal number r is called the *Kronecker rank* of a given tensor $\mathcal{A}_{(r)}$. Here and in the following we use the notation $\otimes$ to represent the canonical tensor

$$\mathcal{U} \equiv [u_{\mathbf{i}}]_{\mathbf{i} \in \mathcal{I}} = b\, U^{(1)} \otimes ... \otimes U^{(d)} \in \mathbb{C}^{\mathcal{I}},$$

defined by $u_{i_1 \dots i_d} = b \cdot u_{i_1}^{(1)} \cdots u_{i_d}^{(d)}$ with $U^{(\ell)} \equiv [u_{i_\ell}^{(\ell)}]_{i_\ell \in I_\ell} \in \mathbb{C}^{I_\ell}$. We make use of the multi-index notation $\mathbf{i} := (i_1, ..., i_d) \in \mathcal{I}$.

The Tucker model deals with the approximation

$$\mathcal{A} \approx \mathcal{A}_{(\mathbf{r})} = \sum_{k_1=1}^{r_1} ... \sum_{k_d=1}^{r_d} b_{k_1 \dots k_d} V_{k_1}^{(1)} \otimes ... \otimes V_{k_d}^{(d)}, \tag{4.2}$$

where the Kronecker factors $V_{k_\ell}^{(\ell)} \in \mathbb{C}^{I_\ell}$ ($k_\ell = 1, ..., r_\ell$, $\ell = 1, ..., d$) are complex vectors of the respective size $n_\ell = |I_\ell|$, $\mathbf{r} = (r_1, ..., r_d)$ (the Tucker rank) and $b_{k_1 \dots k_d} \in \mathbb{C}$. Without loss of generality, we assume that the vectors $\{V_{k_\ell}^{(\ell)}\}$ are orthonormal, i.e.,

$$\left\langle V_{k_\ell}^{(\ell)}, V_{m_\ell}^{(\ell)} \right\rangle = \delta_{k_\ell, m_\ell}, \quad k_\ell, m_\ell = 1, ..., r_\ell;\ \ell = 1, ..., d,$$

where δ_{k_ℓ,m_ℓ} is Kronecker's delta.

On the level of operators (matrices) we distinguish the following tensor-product structures. Given a matrix $A \in \mathbb{C}^{N\times N}$ with $N = n^d$, we approximate it with the canonical model by a matrix $A_{(r)}$ of the form

$$A \approx A_{(r)} = \sum_{k=1}^{r} V_k^{(1)} \otimes \cdots \otimes V_k^{(d)}, \tag{4.3}$$

where the $V_k^{(\ell)}$ are hierarchically structured matrices of order $n \times n$. Again the important parameter r is denoted as the Kronecker rank.

We also introduce the following rank-$(r_1, ..., r_d)$ Tucker-type tensor-product matrix format

$$\mathcal{A} = \sum_{k_1=1}^{r_1} ... \sum_{k_d=1}^{r_d} b_{k_1...k_d} V_{k_1}^{(1)} \otimes ... \otimes V_{k_d}^{(d)} \in \mathbb{R}^{I_1^2\times...\times I_d^2}, \tag{4.4}$$

where the Kronecker factors $V_{k_\ell}^{(\ell)} \in \mathbb{R}^{I_\ell\times I_\ell}$, $k_\ell = 1, ..., r_\ell$, $\ell = 1, ..., d$, are matrices of a certain structure (say, $\mathcal{H}$-matrix, wavelet based format, Toeplitz/circulant, low-rank, banded, etc.). The matrix representation in the form (4.4) is a model reduction which is a generalisation of the low-rank approximation of matrices, corresponding to the case $d = 2$. For a class of matrix-valued functions (cf. [53, 57] and Section 6.1 below) it is possible to show that $r = \mathcal{O}(|\log\varepsilon|^{2(d-1)})$. Further results on the tensor-product approximation to certain matrix-valued functions can be found in [41, 54].

Note that *algebraic recompression methods* based on the *singular value decomposition* (SVD) cannot be directly generalised to $d \geq 3$. We refer to [5, 6, 25, 26, 27, 33, 57, 58, 63, 65, 66, 73, 89] and references therein for detailed description of the methods of numerical multi-linear algebra. In the following, we stress the significance of analytical methods for the separable approximation of multivariate functions and related function-generated matrices/tensors.

4.2 Separable approximation of functions

Separable approximation of functions plays an important role in the design of effective tensor-product decomposition methods. For a large class of functions (cf. [83, 84]) it is possible to show that tensor-product approximations with low separation rank exist. In this section, we overview the most commonly used methods to construct separable approximations of multivariate functions.

4.2.1 *Sinc* interpolation methods

Sinc-approximation methods provide the efficient tools for interpolating $\mathbb{C}^\infty$ functions on $\mathbb{R}$ having exponential decay as $|x| \to \infty$ (cf. [79]). Let

$$S_{k,\mathfrak{h}}(x) = \frac{\sin[\pi(x - k\mathfrak{h})/\mathfrak{h}]}{\pi(x - k\mathfrak{h})/\mathfrak{h}} \qquad (k \in \mathbb{Z},\ \mathfrak{h} > 0,\ x \in \mathbb{R})$$

be the k-th *Sinc* function with step size $\mathfrak{h}$, evaluated at x. Given f in the Hardy space $H^1(D_\delta)$ with respect to the strip $D_\delta := \{z \in \mathbb{C} : |\Im z| \leq \delta\}$ for a $\delta < \frac{\pi}{2}$. Let $\mathfrak{h} > 0$ and $M \in \mathbb{N}_0$, the

corresponding *Sinc*-interpolant (cardinal series representation) and quadrature read as

$$C_M(f,\mathfrak{h}) = \sum_{k=-M}^{M} f(k\mathfrak{h}) S_{k,\mathfrak{h}}, \quad T_M(f,\mathfrak{h}) = \mathfrak{h} \sum_{k=-M}^{M} f(k\mathfrak{h}),$$

where the latter approximates the integral

$$I(f) = \int_{\mathbb{R}} f(x)dx.$$

For the interpolation error, the choice $\mathfrak{h} = \sqrt{\pi\delta/bM}$ implies the exponential convergence rate

$$\|f - C_M(f,\mathfrak{h})\|_\infty \leq CM^{1/2} e^{-\sqrt{\pi\delta bM}}.$$

Similarly, for the quadrature error, the choice $\mathfrak{h} = \sqrt{2\pi\delta/bM}$ yields

$$|I(f) - T_M(f,\mathfrak{h})| \leq Ce^{-\sqrt{2\pi\delta bM}}.$$

If f has a double-exponential decay as $|x| \to \infty$, i.e.,

$$|f(\xi)| \leq C\exp(-be^{a|\xi|}) \qquad \text{for all } \xi \in \mathbb{R} \text{ with } a,b,C > 0,$$

the convergence rate of both *Sinc*-interpolation and *Sinc*-quadrature can be improved up to $\mathcal{O}(e^{-cM/\log M})$.

For example, let $d = 2$. Given a function $F(\zeta,\eta)$ defined in the product domain $\Omega := [0,1] \times [a,b]$, $a,b \in \mathbb{R}$, we assume that for each fixed $\eta \in [a,b]$, the univariate function $F(\cdot,\eta)$ belongs to $C^\infty(0,1]$ and allows a certain holomorphic extension (with respect to ζ) to the complex plane $\mathbb{C}$ (cf. [53] for more details). Moreover, the function $F(\cdot,\eta)$ restricted onto $[0,1]$ is allowed to have a singularity with respect to ζ at the end-point $\zeta = 0$ of $[0,1]$. Specifically, it is assumed that there is a function $\phi : \mathbb{R} \to (0,1]$ such that for any $\eta \in [a,b]$ the composition $f(x) = F(\phi(x),\eta)$ belongs to the class $H^1(D_\delta)$. For this class of functions a separable approximation is based on the transformed *Sinc*-interpolation [41, 79] leading to

$$F_M(\zeta,\eta) = \sum_{k=-M}^{M} F(\phi(k\mathfrak{h}),\eta) S_{k,\mathfrak{h}}(\phi^{-1}(\zeta)) \approx F(\zeta,\eta).$$

The following error bound

$$\sup_{\zeta\in[a,b]} |F(\zeta,\eta) - F_M(\zeta,\eta)| \leq Ce^{-sM/\log M} \tag{4.5}$$

holds with $\phi^{-1}(\zeta) = \operatorname{arsinh}(\operatorname{arcosh}(\zeta^{-1}))$. In the case of a multivariate function in $[0,1]^{d-1} \times [a,b]$, one can adapt the corresponding tensor-product approximation by successive application of the one-dimensional interpolation (cf. [53]). In the numerical example shown in Fig. 1), we approximate the Euclidean distance $|x-y|$ in $\mathbb{R}^3$ on the domain $|x_i - y_i| \leq 1$ $(i = 1,2,3)$, by the *Sinc*-interpolation. To that end, the approximation (4.5) applies to the function $F(\zeta,\eta,\theta) = \sqrt{\zeta^2+\eta^2+\theta^2}$ in $\Omega := [0,1]^3$.

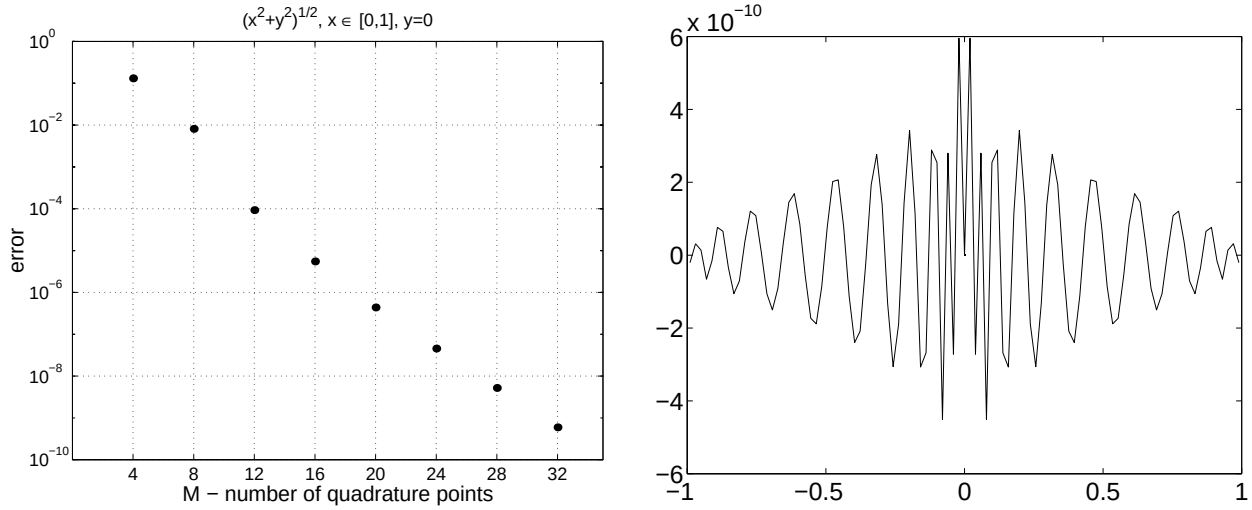

Figure 1: Convergence of the *Sinc*-approximation of $F(\zeta,\eta,\theta)=\sqrt{\zeta^2+\eta^2+\theta^2}$ on the slice $\eta=\theta=0$ (left), and the pointwise error for $M=32$ on $(\zeta,0,0)$, $\zeta\in[-1,1]$ (right).

4.2.2 Integral representation methods

Integral representation methods are based on the quadrature approximation of integral Laplace-type transforms representing spherically symmetric functions. In particular, some functions of the Euclidean distance in $\mathbb{R}^d$, say,

$$1/|x-y|,\ \ |x-y|^\beta,\ \ e^{-|x-y|},\ \ e^{-\lambda|x-y|}/|x-y|,\quad x,y\in\mathbb{R}^d,$$

can be approximated by *Sinc*-quadratures of the corresponding Gaussian integral on the semi-axis [41, 53, 54, 64].

For example, in the range $0<a\le|x-y|\le A$, one can use the integral representation

$$\frac{1}{|x-y|}=\frac{1}{\sqrt{\pi}}\int\limits_{\mathbb{R}}\exp(-|x-y|^2t^2)dt=\int\limits_{\mathbb{R}}F(\varrho;t)dt,\quad x,y\in\mathbb{R}^d \tag{4.6}$$

of the Coulomb potential with

$$F(\varrho;t)=\frac{1}{\sqrt{\pi}}e^{-\varrho^2t^2},\quad \varrho=|x-y|,\quad d=3.$$

After the substitution $t=\log(1+e^u)$ and $u=\sinh(w)$ in the integral (4.6), we apply the quadrature to obtain

$$T_M(F,\mathfrak{h}):=\mathfrak{h}\sum_{k=-M}^{M}\cosh(k\mathfrak{h})G(\varrho,\sinh(k\mathfrak{h}))\approx\int_{\mathbb{R}}F(\varrho,t)dt=\frac{1}{\varrho} \tag{4.7}$$

with $G(\varrho,u)=\frac{2}{\sqrt{\pi}}\frac{e^{-\varrho^2\log^2(1+e^u)}}{1+e^u}$ and with $\mathfrak{h}=C_0\log M/M$. The quadrature (4.7) is proven to converge exponentially in M,

$$E_M:=\left|\frac{1}{\varrho}-T_M(F,\mathfrak{h})\right|\le Ce^{-sM/\log M},$$

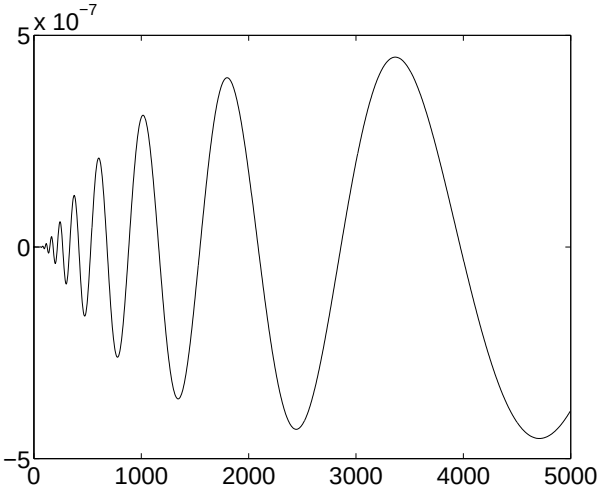

Figure 2: Error of the quadrature formula (4.7) for the Coulomb potential on the interval $[1, 5000]$ for $M = 64$.

where C, s do not depend on M (but depend on ϱ), see [53]. With the proper scaling of the Coulomb potential, one can apply this quadrature in the reference interval $\varrho \in [1, R]$. A numerical example for this quadrature with values $\varrho \in [1, R]$, $R \leq 5000$, is presented in Fig. 2. We observe almost linear error growth in ϱ.

In electronic structure calculations, the Galerkin discretisation of the Coulomb potential in tensor-product wavelet bases is of specific interest. For simplicity, we consider an isotropic $3d$-wavelet basis

$$\gamma^{(\mathbf{s})}_{j,\mathbf{a}}(\mathbf{x}) := \psi^{(s_1)}_{j,a_1}(x_1)\, \psi^{(s_2)}_{j,a_2}(x_2)\, \psi^{(s_3)}_{j,a_3}(x_3),$$

where the functions $\psi^{(0)}_{j,a}(x) := 2^{j/2}\psi^{(0)}(2^j x - a)$, $\psi^{(1)}_{j,a}(x) := 2^{j/2}\psi^{(1)}(2^j x - a)$, with $j, a \in \mathbb{Z}$, correspond to univariate scaling functions and wavelets, respectively. The *nonstandard* representation of the Coulomb potential (cf. [8, 34]) requires integrals of the form

$$\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \gamma^{(\mathbf{p})}_{j,\mathbf{a}}(\mathbf{x})\, \frac{1}{|\,\mathbf{x}-\mathbf{y}\,|}\, \gamma^{(\mathbf{q})}_{j,\mathbf{b}}(\mathbf{y})\, d^3x d^3y = \frac{2^{-2j+1}}{\sqrt{\pi}} \int_0^\infty \mathcal{I}^{(\mathbf{p},\mathbf{q})}(t, \mathbf{a}-\mathbf{b}) dt,$$

with

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

and

$$G^{(p,q)}(a, t) = \int_{\mathbb{R}}\int_{\mathbb{R}} \psi^{(p)}(x-a)\, e^{-(x-y)^2 t^2}\, \psi^{(q)}(y)\, dxdy.$$

In order to benefit from the tensor-product structure, it is important to have a uniform error bound with respect to the spatial separation $|\mathbf{a}-\mathbf{b}|$ of the wavelets. Recently, the following theorem was proven by Schwinger [78]

Theorem 4.1 *Given a univariate wavelet basis $\psi^{(p)}_{j,a}$ which satisfies*

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

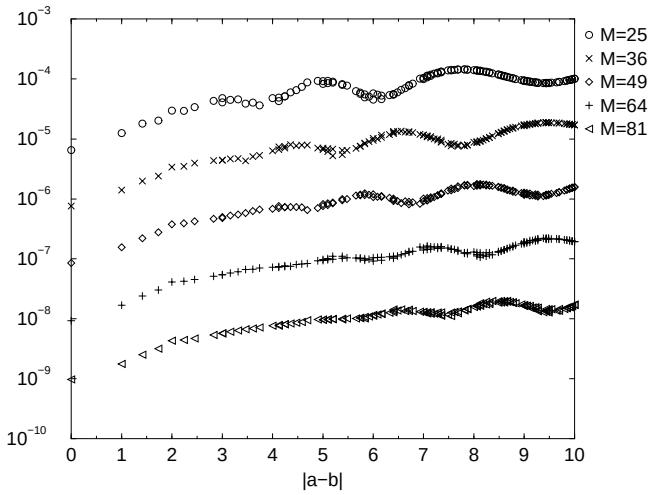

Figure 3: Relative error of Coulomb integrals (4.8) for the case of pure scaling functions, i.e., $\mathbf{p} = \mathbf{q} = (0, 0, 0)$, for different number of quadrature points M. The isotropic $3d$-scaling functions were generated from the univariate Deslauriers Dubuc interpolating scaling function [29] which is exact up to 5'th order.

Then for any $\delta < \frac{\pi}{4}$, the integration error of the exponential quadrature rule (cf. [79]*) with $h = \sqrt{\frac{\pi\delta}{M}}$ ($h = \sqrt{\frac{2\pi\delta}{M}}$ pure scaling functions, i.e., $\mathbf{p} = \mathbf{q} = (0, 0, 0)$) satisfies*

$$\left| \int_0^\infty \mathcal{I}^{(\mathbf{p},\mathbf{q})}(t, \mathbf{a}) dt - h \sum_{m=-M}^{M} e^{mh} \mathcal{I}^{(\mathbf{p},\mathbf{q})}(e^{mh}, \mathbf{a}) \right| \leq C e^{-\alpha\sqrt{M}} \tag{4.8}$$

for $\alpha = 2\sqrt{\pi\delta}$ ($\alpha = \sqrt{2\pi\delta}$ pure scaling functions) with constant C independent of the translation parameter $\mathbf{a}$.

We illustrate the theorem for the case of pure scaling functions in Fig. 4.2.2. Similar results for wavelets are presented in [14].

4.2.3 On the best approximation by exponential sums

Using integral representation methods, the *Sinc*-quadrature can be applied, for example, to the integrals

$$\frac{1}{\varrho} = \int_0^\infty \mathrm{e}^{-\varrho\xi} \mathrm{d}\xi, \quad \text{and} \quad \frac{1}{\varrho} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty \mathrm{e}^{-\varrho^2 t^2} \mathrm{d}t$$

to obtain an exponentially convergent sum of exponentials approximating the inverse function $\frac{1}{\varrho}$. Instead, one can directly determine the best approximation of a function with respect to a certain norm by exponential sums $\sum_{\nu=1}^{n} \omega_\nu \mathrm{e}^{-t_\nu x}$ or $\sum_{\nu=1}^{n} \omega_\nu \mathrm{e}^{-t_\nu x^2}$, where $\omega_\nu, t_\nu \in \mathbb{R}$ are to be chosen optimally. For some applications in quantum chemistry of approximation by exponential sums we refer e.g. to [1, 59, 61].

We recall some facts from the approximation theory by exponential sums (cf. [10] and the discussion in [53]). The existence result is based on the fundamental *Big Bernstein Theorem*:

If f is completely monotone for $x \geq 0$, i.e.,

$$(-1)^n f^{(n)}(x) \geq 0 \qquad \text{for all } n \geq 0,\ x \geq 0,$$

then it is the restriction of the Laplace transform of a measure to the half-axis:

$$f(z) = \int_{\mathbb{R}_+} \mathrm{e}^{-tz} \mathrm{d}\mu(t).$$

For $n \geq 1$, consider the set E_n^0 of exponential sums and the extended set E_n:

$$E_n^0 := \left\{ u = \sum_{\nu=1}^{n} \omega_\nu \mathrm{e}^{-t_\nu x} :\ \omega_\nu,\ t_\nu \in \mathbb{R} \right\},$$
$$E_n := \left\{ u = \sum_{\nu=1}^{\ell} p_\nu(x) \mathrm{e}^{-t_\nu x} :\ t_\nu \in \mathbb{R},\ p_\nu \text{ polynomials with } \sum_{\nu=1}^{\ell} (1 + \mathrm{degree}(p_\nu)) \leq n \right\}.$$

Now one can address the problem of finding the best approximation to f over the set E_n characterised by the best N-term approximation error

$$d_\infty(f, E_n) := \inf_{v \in E_n} \|f - v\|_\infty.$$

We recall the complete elliptic integral of the first kind with modulus κ,

$$\mathbf{K}(\kappa) = \int_0^1 \frac{\mathrm{d}t}{\sqrt{(1-t^2)(1-\kappa^2 t^2)}} \qquad (0 < \kappa < 1)$$

(cf. [12]), and define $\mathbf{K}'(\kappa) := \mathbf{K}(\kappa')$ by $\kappa^2 + (\kappa')^2 = 1$.

Theorem 4.2 [2] ([10]) *Assume that f is completely monotone and analytic for $\Re\mathrm{e}\, z > 0$, and let $0 < a < b$. Then for the uniform approximation on the interval $[a, b]$,*

$$\lim_{n\to\infty} d_\infty(f, E_n)^{1/n} \leq \frac{1}{\omega^2}, \quad \textit{where } \omega = \exp \frac{\pi \mathbf{K}(\kappa)}{\mathbf{K}'(\kappa)} \quad \textit{with } \kappa = \frac{a}{b}.$$

In the case discussed below, we have $\kappa = 1/R$ for possibly large R. Applying the asymptotics

$$\mathbf{K}(\kappa') = \ln \tfrac{4}{\kappa} + C_1 \kappa + ... \qquad \text{for } \kappa' \to 1,$$
$$\mathbf{K}(\kappa) = \tfrac{\pi}{2}\{1 + \tfrac{1}{4}\kappa^2 + C_1 \kappa^4 + ...\} \qquad \text{for } \kappa \to 0,$$

of the complete elliptic integrals (cf. [44]), we obtain

$$\frac{1}{\omega^2} = \exp\left(-\frac{2\pi \mathbf{K}(\kappa)}{\mathbf{K}(\kappa')}\right) \approx \exp\left(-\frac{\pi^2}{\ln(4R)}\right) \approx 1 - \frac{\pi^2}{\ln(4R)}.$$

The latter expression indicates that the number n of different terms to achieve a tolerance ε is asymptotically

$$n \approx \frac{|\log \varepsilon|}{|\log \omega^{-2}|} \approx \frac{|\log \varepsilon| \ln(4R)}{\pi^2}.$$

[2] The same result holds for E_n^0, but the best approximation may belong to the closure E_n of E_n^0.

This result shows the same asymptotical convergence in n as the corresponding bound in the *Sinc*-approximation theory.

Optimisation with respect to the maximum norm leads to the nonlinear minimisation problem $\inf_{v\in E_n^0}\|f-v\|_{L^\infty[1,R]}$ involving $2n$ parameters $\{\omega_\nu, t_\nu\}_{\nu=1}^n$. The numerical implementation is based on the Remez algorithm (cf. [12]). For the particular application with $f(x)=x^{-1}$, we have the same asymptotical dependence $n=n(\varepsilon,R)$ as in the *Sinc*-approximation above, however, the numerical results [3] indicate a noticeable improvement compared with the quadrature method, at least for $n\le 15$.

The best approximation to $1/\varrho^\mu$ in the interval $[1,R]$ with respect to a W-weighted L^2-norm can be reduced to the minimisation of an explicitly given differentiable functional

$$d_2(f,E_n) := \inf_{v\in E_n}\|f-v\|_{L^2_W}.$$

Given $R>1$, $\mu>0$, $n\ge 1$, find the $2n$ real parameters $t_1,\omega_1,...,t_n,\omega_n\in\mathbb{R}$, such that

$$F_\mu(R;t_1,\omega_1,...,t_n,\omega_n) := \int_1^R W(x)\Big(\frac{1}{x^\mu}-\sum_{i=1}^n \omega_i \mathrm{e}^{-t_i x}\Big)^2 \mathrm{d}x = \min. \tag{4.9}$$

In the particular case of $\mu=1$ and $W(x)=1$, the integral (4.9) can be calculated in a closed form[4]:

$$F_1(R;t_1,\omega_1,...,t_n,\omega_n) = 1-\frac{1}{R}-2\sum_{i=1}^n \omega_i\left[\mathrm{Ei}(-t_i)-\mathrm{Ei}(-t_iR)\right]$$
$$+\frac{1}{2}\sum_{i=1}^n \frac{\omega_i^2}{t_i}\left[\mathrm{e}^{-2t_i}-\mathrm{e}^{-2t_iR}\right]+2\sum_{1\le i<j\le n}\frac{\omega_i\omega_j}{t_i+t_j}\left[\mathrm{e}^{-(t_i+t_j)}-\mathrm{e}^{-(t_i+t_j)R}\right]$$

with the integral exponential function $\mathrm{Ei}(x)=\int_{-\infty}^x \frac{\mathrm{e}^t}{t}\mathrm{d}t$ (cf. [12]). In the special case $R=\infty$, the expression for $F_1(\infty;\ldots)$ even simplifies. Gradient or Newton type methods with a proper choice of the initial guess can be used to obtain the minimiser of F_1 (cf. [56]).

5 Data sparse formats for univariate components

5.1 Hierarchical matrix techniques

The hierarchical matrix ($\mathcal{H}$-matrix) technique [46, 50, 51, 55] (see also the mosaic-skeleton method [82]) allows an efficient treatment of dense matrices arising, e.g., from BEM, evaluation of volume integrals and multi-particle interactions, certain matrix-valued functions, etc. In particular, it provides matrix formats which enable the computation and storage of inverse FEM stiffness matrices corresponding to elliptic problems as well as of BEM matrices.

The hierarchical matrices are represented by means of a certain block partitioning. Fig. 4 shows typical admissible block structures. Each block is filled by a submatrix of a rank not exceeding k. Then, for the mentioned class of matrices, it can be shown that the exact dense matrix A and the approximating hierarchical matrix $A_\mathcal{H}$ differ by $\|A-A_\mathcal{H}\|\le\mathcal{O}(\eta^k)$ for a

[3] Numerical results for the best approximation of x^{-1} by sums of exponentials can be found in [10] and [11]; a full list of numerical data is presented in `www.mis.mpg.de/scicomp/EXP_SUM/1_x/tabelle`.

[4] In the general case, the integral (4.9) may be approximated by certain quadratures.

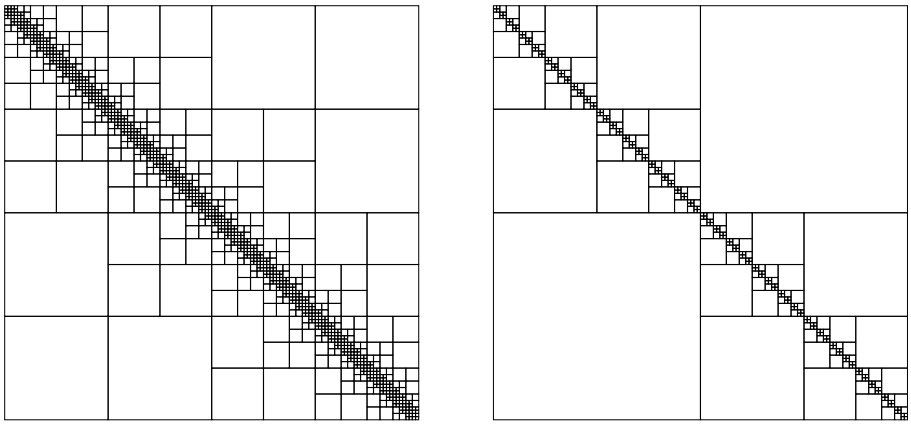

Figure 4: Hierarchical partitioning by the standard and weak admissibility conditions

certain number $\eta < 1$. This exponential decrease allows to obtain an error ε by the choice $k = \mathcal{O}(\log(1/\varepsilon))$. It is shown (cf. [50, 51, 52]) that the $\mathcal{H}$-matrix arithmetic exhibits almost linear complexity in N:

- *Data compression.* The storage of $N \times N$ $\mathcal{H}$-matrices as well as the matrix-by-vector multiplication and matrix-matrix addition have a cost $\mathcal{O}(kN \log N)$, where the local rank k is the parameter determining the approximation error.
- *Matrix-by-matrix and matrix-inverse complexity.* The approximate matrix-matrix multiplication and the approximate inversion both take $\mathcal{O}\left(k^2 N \log^2 N\right)$ operations.
- *The Hadamard (entry-wise) matrix product.* The exact Hadamard product of two rank-k $\mathcal{H}$-matrices leads to an $\mathcal{H}$-matrix of the block-rank k^2 (see Section 5.2 below).

5.2 Hierarchical Kronecker tensor-product approximations

Since n is much smaller than N, one can apply the hierarchical (or low-rank) matrix structure to represent the Kronecker factors V_k^ℓ in (4.3) with the complexity $\mathcal{O}(n \log^q n)$ or even $\mathcal{O}(n)$ that finally leads to $\mathcal{O}(rn) = \mathcal{O}(rN^{1/d})$ data to represent the compressed matrix A_r. We call by $\mathrm{HKT}(r, s)$ the class of Kronecker rank-r matrices, where the Kronecker factors V_k^ℓ are represented by the block-rank s $\mathcal{H}$-matrices (shortly, HKT-matrices). It was shown in [57] that the advantages of replacing A with A_r (cf. (4.3)), where all the Kronecker factors possess the structure of general $\mathcal{H}$-matrices, are the following:

- *Data compression.* The storage for the V_k^ℓ matrices of (4.3) is only $\mathcal{O}(rn) = \mathcal{O}(rN^{1/d})$ while that for the original (dense) matrix A is $\mathcal{O}(N^2)$, where $r = \mathcal{O}(\log^\alpha N)$ for some $\alpha > 0$. Consequently, we enjoy a *linear-logarithmic complexity* of $\mathcal{O}(n \log^\alpha n)$ in the univariate problem size n.
- *Matrix-by-vector complexity.* Instead of $\mathcal{O}(N^2)$ operations to compute Ax, $x \in \mathbb{C}^N$, we now need only $\mathcal{O}(rkn^d \log n) = \mathcal{O}(rkN \log n)$ operations. If the vector can be represented in a tensor-product form (say, $x = x^1 \otimes \ldots \otimes x^d$, $x^i \in \mathbb{C}^n$) the corresponding cost is reduced to $\mathcal{O}(rkn \log n) = \mathcal{O}(rkN^{1/d} \log n)$ operations.

- *Matrix-by-matrix complexity.* Instead of $\mathcal{O}(N^3)$ operations to compute AB, we now need only $\mathcal{O}(r^2 n^3) = \mathcal{O}(r^2 N^{3/d})$ operations for rather general structure of the Kronecker factors. Remarkably, this result is much better than the corresponding matrix-by-vector complexity for a general vector x.

- *Hadamard product.* The Hadamard (entry-wise) product of two HKT-matrices $A * B$ is presented in the same format: $(U_1 \times V_1) * (U_2 \times V_2) = (U_1 * U_2) \times (V_1 * V_2)$. In turn, the exact Hadamard product $U_1 * U_2$ (same for $V_1 * V_2$) of two rank-k $\mathcal{H}$-matrices results in an $\mathcal{H}$-matrix of the block-rank k^2 and with the corresponding "skeleton" vectors defined by the Hadamard products of those in the initial factors (since there holds $(a \otimes b) * (a_1 \otimes b_1) = (a * a_1) \otimes (b * b_1)$).

Therefore, basic linear algebra operations can be performed in the tensor-product representation using one-dimensional operations, thus avoiding an exponential scaling in the dimension d.

The exact product of two HKT-matrices can be represented in the same format, but with squared Kronecker rank and properly modified block-rank [57]. If $A, B \in \mathrm{HKT}(r, s)$, where s corresponds to the block-rank of the $\mathcal{H}$-matrices involved, then in general $AB \notin \mathrm{HKT}(r, s)$. However,

$$A = \sum_{k=1}^{r} U_k^A \otimes V_k^A, \quad B = \sum_{l=1}^{r} U_l^B \otimes V_l^B, \quad U_k^A, V_k^A, U_l^B, V_l^B \in \mathbb{C}^{n \times n}, \tag{5.1}$$

leads to

$$AB = \sum_{k=1}^{r} \sum_{l=1}^{r} (U_k^A U_l^B) \otimes (V_k^A V_l^B).$$

It can be proven that the $U_k^A U_l^A$ and $V_k^A V_l^B$ matrices possess the same hierarchical partitioning as the initial factors in (5.1) with blocks of possibly larger (than s) rank bounded, nevertheless, by $s_{AB} = \mathcal{O}(s \log N)$. Thus, $AB \in \mathrm{HKT}(r^2, s_{AB})$ with $s_{AB} = \mathcal{O}(s \log N)$.

5.3 Wavelet Kronecker tensor-product approximations

Wavelet matrix compression was introduced in [8]. This techniques has been considered by one of the authors during the past decade in a series of publications (cf. [76]). The compression of the Kronecker factors $V_i \in \mathbb{R}^{n \times n}$ is not so obvious, since it is not clear to what extend they satisfy a Calderón-Zygmund condition. It is more likely that they obey more or less a hyperbolic cross structure. An underlying truncation criterion based on the size of the coefficients will provide an automatic way to find the optimal structure independent of an a priori assumption. A basic thresholding or a posteriori criterion has been formulated by Harbrecht [60] and in [22]. With this criterion at hand, we expect linear scaling with respect to the size of the matrices.

- *Data compression.* The matrices in (4.3) V_k^ℓ can be compressed requiring total storage size about $\mathcal{O}(rn) = \mathcal{O}(rN^{1/d})$, where $r = \mathcal{O}(\log^\alpha N)$ is as above. The data vector requires at most $\mathcal{O}(n \log^d n)$ nonzero coefficients.

- *Matrix-by-vector complexity.* Instead of $\mathcal{O}(N^2)$ operations to compute Ax, $x \in \mathbb{C}^N$, we now need only $\mathcal{O}(rn^d) = \mathcal{O}(rN)$ operations. If the vector is represented in a tensor-product form (say, $x = x^1 \otimes ... \otimes x^d$, $x^i \in \mathbb{C}^n$) or in sparse grid representation, then the corresponding cost is reduced to $\mathcal{O}(rn)$, resp. $\mathcal{O}(rn \log^d n)$ operations .

- *Matrix-by-matrix complexity.* Using the compression of the Lemarié algebra [81], instead of $\mathcal{O}(N^3)$ operations to compute AB, we need only $\mathcal{O}(r^2 n \log^q n) = \mathcal{O}(r^2 N^{1/d} \log^q N)$, or even $\mathcal{O}(r^2 n)$ operations.

Adaptive wavelet schemes for nonlinear operators have been developed in [3, 24] and for nonlocal operators in [23]. Corresponding schemes for hyperbolic cross approximations have not been worked out up to now. Perhaps basic ideas can be transfered immediately to the tensor-product case.

6 Linear scaling methods for Hartree-Fock and Kohn-Sham equations

Operator-valued functions $\mathcal{G}(\mathcal{L})$ of elliptic operators $\mathcal{L}$ play a prominent role in quantum many-particle theory. A possible representation of the operator $\mathcal{G}(\mathcal{L})$ is given by the Dunford-Cauchy integral (cf. [38, 39, 40, 41])

$$\mathcal{G}(\mathcal{L}) = \frac{1}{2\pi i} \int_\Gamma \mathcal{G}(z)(zI - \mathcal{L})^{-1} dz,$$

where Γ envelopes the spectrum $\mathrm{spec}(\mathcal{L})$ of the operator $\mathcal{L}$ in the complex plane. This kind of representation is especially suitable for tensor-product approximation using *Sinc* or Gauss-Lobatto quadratures for the contour integral to get an approximate operator of the form

$$\mathcal{G}(\mathcal{L}) \approx \sum c_k \mathcal{G}(z_k)(z_k I - \mathcal{L})^{-1}. \tag{6.1}$$

An important example for an operator valued function is the *sign* function of the shifted Fock operator which can be directly related to the spectral projector P_ρ associated with the density matrix ρ. This relation

$$P_\rho = \frac{1}{2}\left[I - \mathrm{sign}(\mathcal{F} - \mu I)\right] = -\frac{1}{2\pi i} \int_\Gamma (\mathcal{F} - zI)^{-1} dz,$$

where $\Gamma \cap \mathrm{spec}(\mathcal{F}) = \emptyset$ encloses the $N/2$ lowest eigenvalues of the Fock operator, has been first noticed by Beylkin, Coult and Mohlenkamp [7]. In order to be applicable, the method requires a finite gap between the highest occupied $\varepsilon_{N/2}$ and lowest unoccupied $\varepsilon_{N/2+1}$ eigenvalue to adjust the parameter $\varepsilon_{N/2} < \mu < \varepsilon_{N/2+1}$. This constraint, in particular, excludes metallic systems.

In general, the approximability of inverse matrices, required in (6.1), within the HKT format is still an open problem. First results on fast approximate algorithms to compute inverse matrices in the HKT format for the case $d \geq 2$ can be found in [41]. In Fig. 6, we consider the HKT representation to the discrete Laplacian inverse $(-\Delta_h)^{-1}$ (homogeneous Dirichlet boundary conditions) in $\mathbb{R}^d$, which can be obtained with $\mathcal{O}(dn \log^q n)$ cost. Numerical examples for still higher dimensions $d \leq 1024$ are presented in [45]. For comparison, the

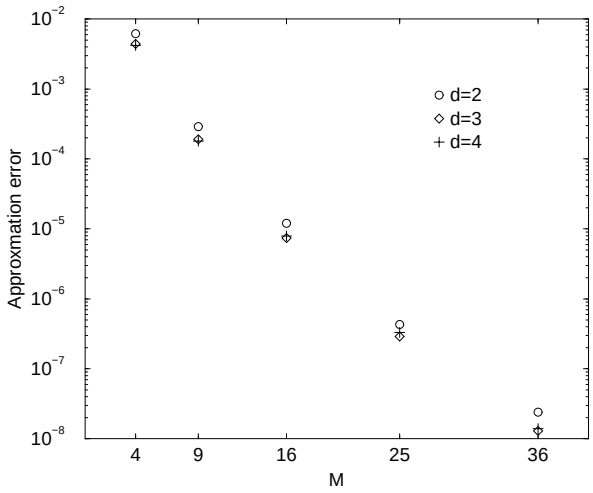

Figure 5: Error for canonical tensor product approximations of $(-\Delta_h)^{-1}$ in $[0,1]^d$ with $d = 2, 3$ and 4. Here M denotes the number of *Sinc*-quadrature points so that the separation rank is given by $r = 2M + 1$, cf. [41] for further details.

following numerical example manifests the optimal Kronecker rank of the discrete elliptic inverse in $d = 2$. Let $-\Delta_h$ now correspond to a five-point stencil discretization of the Laplacian on a uniform mesh in the unit rectangle in $\mathbb{R}^2$ (Dirichlet boundary conditions). It is easy to see that the Kronecker rank of $-\Delta_h$ is 2. The Kronecker ranks of $(-\Delta_h)^{-1}$ for different relative approximation accuracies (in the Frobenius norm) are given in Table 6. Our results indicate a logarithmic bound $\mathcal{O}(\log \varepsilon^{-1})$ for the approximate Kronecker rank r.

6.1 Matrix-valued functions approach for density matrices

Let $\mathbb{F} \in \mathbb{R}^{M\times M}$ be the Fock matrix that represents the Fock operator $\mathcal{F}$ (cf. (2.8)) in an orthogonal basis $\{\varphi_i\}_{i=1}^M$, $M \geq N/2$. There exist two different approaches to compute the Galerkin discretization $\mathbb{D} \in \mathbb{R}^{M\times M}$ of the density matrix (2.6) via the matrix *sign* of the shifted Fock matrix

$$\mathbb{D} = \frac{1}{2}[\mathbb{I} - \operatorname{sign}(\mathbb{F} - \mu\mathbb{I})], \quad \text{with } \mu \in (\varepsilon_{N/2}, \varepsilon_{N/2+1}).$$

The first approach uses an exponentially convergent quadrature for the integral to obtain an expansion into resolvents (6.1) whereas the second approach is based on a Newton-Schultz iteration scheme. Concerning the tensor-product approximation of resolvents in the HKT

Table 1: Canonical tensor product approximation with optimal separation rank of the inverse Laplacian $(-\Delta_h)^{-1}$ on a uniform rectangular grid in $[0,1]^2$. Here $n = 64$ and, hence, $-\Delta_h \in \mathbb{R}^{N\times N}$ with $N = 4096$.

Kronecker rank for $(-\Delta_h)^{-1}$	6	8	9	11	12	14
Relative Frobenius error	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}

format we refer to our discussion in Section 5.2. For the Newton-Schultz iteration scheme proposed in [7]

$$S^{(n+1)} = S^{(n)} + \frac{1}{2}\left[I - (S^{(n)})^2\right] S^{(n)} \quad , \quad S^{(0)} = (\mathbb{F} - \mu\mathbb{I})\,/||\mathbb{F} - \mu\mathbb{I}||_2, \tag{6.2}$$

the sequence $S^{(n)}$ converges to $sign(\mathbb{F} - \mu\mathbb{I})$. First applications in quantum chemistry by Németh and Scuseria [70] demonstrate the practicability of this approach. Iterations schemes of the form (6.2) seem to be especially favourable for tensor-product formats. Starting from an initial approximation of the Fock matrix $\mathbb{F}$, with low separation rank one has to perform matrix-matrix multiplications which can be handled in an efficient manner in the tensor-product format, cf. our discussion in Section 5.2. After each iteration step a recompression of the tensor-product decomposition of $S^{(n+1)}$ becomes necessary. For the recompression one can apply the simple *alternating least squares* (ALS) method [5, 86, 89] or Newton-type and related algebraic iterative methods [33]. The ALS algorithm starts with an initial decomposition of $S^{(n+1)}$ with separation rank r and obtains the best approximation with separation rank $\tilde{r} \leq r$ by iteratively solving an optimisation problem for each coordinate separately. Assume that r is actually much larger than necessary, i.e., $\tilde{r} << r$, then the ALS algorithm costs $\mathcal{O}(d\,\tilde{r}(\tilde{r}^2 + rn^2))$. We refer to [77] for the discussion of wavelet methods for density matrix computation in electronic structure calculation.

6.2 Computation of Hartree potentials in tensor-product formats

A common bottleneck for the numerical solution of Hartree-Fock and Kohn-Sham equations is the computation of the Hartree-potential (2.9). Traditionally, highly adapted GTO basis sets are used for the approximation of electron densities which enable an analytic evaluation of the subsequent convolution with the Coulomb potential. This kind of approach became widely known as *density-fitting* or *resolution of the identity* method (cf. [31, 32, 68]) and turned out to be an essential ingredient for computational efficiency. We want to discuss two alternative approaches based on canonical (4.1) and Tucker (4.2) tensor-product decompositions which have been presented in [14] and [15], respectively. Both approaches have to be considered as a generalization of the density-fitting method with respect to the fact that they provide tensor-product approximations for both, the electron density and the Hartree potential. The latter becomes possible via separable approximations of the Coulomb potential obtained via *Sinc* interpolation or best approximation by exponential sums discussed in Sections 4.2.2 and 4.2.3, respectively.

The canonical tensor-product approach is based on approximations for the electron density (2.7) in the format

$$n(\mathbf{x}) = \sum_{k=1}^{K} n_k^{(1)}(x_1) n_k^{(2)}(x_2) n_k^{(3)}(x_3) \approx \sum_{k=1}^{\kappa} \eta_k^{(1)}(x_1)\, \eta_k^{(2)}(x_2)\, \eta_k^{(3)}(x_3),$$

which can be obtained from separable approximations with large separation rank K, e.g. expansions in terms of GTO basis sets, via algebraic compression methods [33]. Using a separable approximation of the Coulomb potential with separation rank $2M + 1$, cf. (4.7), we obtain from an intermediate tensor-product representation with separation rank $\kappa(2M + 1)$, after a further compression step, the Hartree potential in the canonical format with low separation

rank

$$V_H(\mathbf{x}) = \int\limits_{\mathbb{R}^3} \frac{1}{\mid \mathbf{x}-\mathbf{y} \mid} n(\mathbf{y}) d^3y \approx \sum_{k=1}^{\kappa'} v_k^{(1)}(x_1)\, v_k^{(2)}(x_2)\, v_k^{(3)}(x_3).$$

With such an approximation at hand, it is straightforward to assemble the Galerkin matrix with respect to an arbitrary tensor-product basis $\{G_\Lambda(\mathbf{x}) := g_{\lambda_1}(x_1) g_{\lambda_2}(x_2) g_{\lambda_3}(x_3)\}$, $\Lambda := (\lambda_1, \lambda_2, \lambda_3) \in \mathcal{I}$, in an efficient manner by utilising the tensor-product decomposition of the constituents

$$\int\limits_{\mathbb{R}^3}\int\limits_{\mathbb{R}^3} G_\Lambda V_H(\mathbf{x}) G_\Omega d^3x d^3y = \sum_{k=1}^{\kappa} \langle g_{\lambda_1} v_k^{(1)}, g_{\omega_1}\rangle \langle g_{\lambda_2} v_k^{(2)}, g_{\omega_2}\rangle \langle g_{\lambda_3} v_k^{(3)}, g_{\omega_3}\rangle.$$

This approach is therefore especially suitable in combination with conventional GTO or tensor-product wavelet bases. It turned out that significant improvements beyond standard GTO bases can be achieved, however the unconstraint optimization of univariate components required for the compression steps represents a generically ill-posed problem and rather sophisticated optimization techniques are necessary in order to achieve the required high accuracies.

The second approach uses the Tucker format (4.2) to obtain data sparse representations of electron densities. In contrast to the canonical approach, the Tucker approximation is typically a well posed problem. We have studied Tucker approximations of electron densities and Hartree potentials represented on regular cartesian grids. This kind of representation underlies the recently developed BigDFT code [9] for DFT based large scale electronic structure calculations. The presently most efficient approach to compute the Hartree potential on a cartesian grid uses the *Fast Fourier Transform* (FFT) to perform the convolution with the Coulomb potential [42]. Concerning the computational complexity FFT scales $\mathcal{O}(n^3 \log n)$ on a cubic grid, where n is the number of grid points in each direction. Within tensor-product formats it is possible to perform this step with sublinear, i.e., $\mathcal{O}(n^\alpha)$, $1 \leq \alpha < 3$ complexity. The Tucker format is not as convenient for the convolution as the canonical format. Therefore it is favourable to simply rearange the Tucker tensor with Tucker rank r into a canonical tensor with separation rank r^2 and perform the convolution in the canonical format as described above. In a subsequent step it is possible to compress the resulting Hartree potential of rank $r^2(2M+1)$ again within the Tucker format.

We have studied canonical and Tucker type tensor-product approximations of electron densities and Hartree potentials for a series of small molecules. We refer to [14, 15] for a detailed discussion of the canonical and Tucker approach including benchmark calculations for some small molecules. As an illustrative example we present in Fig. 6.2 results for the C_2H_6 molecule. The relative errors of the tensor-product approximations refer to the $L^2(\mathbb{R}^3)$ and discrete Frobenius norm for the canonical and Tucker format, respectively. It can be seen from Fig. 6.2 that the error in the canonical format decreases approximately like $e^{-c\sqrt{\kappa}}$ whereas the Tucker format shows an exponential convergence with respect to the Tucker rank. As already noticed in the previous paragraph it is always possible to rearange the Tucker into the canonical format where the Tucker rank r corresponds to a canonical separation rank $\kappa = r^2$. Here we observe a similar behaviour for electron densities and Hartree potentials.

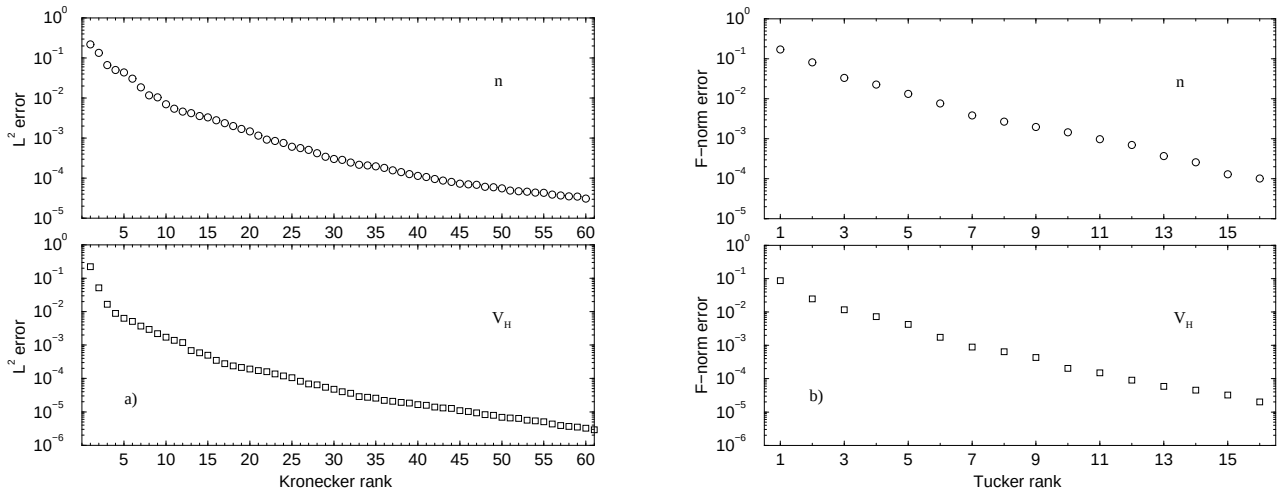

Figure 6: Errors for tensor product approximations of the electron density n and Hartree potential V_H of C_2H_6. a) Kronecker rank of the canonical format versus error in the L^2 norm. b) Tucker rank of the Tucker format versus error in the Frobenius norm.

7 Conclusions

The importance of tensor-product approximations for electronic structure calculations in quantum chemistry and solid state physics can be hardly overestimated. Their scope of applications ranges from basic issues related to the problem to find a convenient ansatz for the many-particle wavefunction in terms of Slater determinants up to more technical issues concerning the efficient computation of integrals involving the Coulomb potential. Within the present work we reviewed some recent developments in numerical analysis comprising best N-term approximation in tensor-product wavelet bases as well as more general canonical and Tucker type tensor-product formats, which can be combined with data sparse representations for the low dimensional components using $\mathcal{H}$ matrices or wavelets. Furthermore, separable approximations of certain kernel functions, like the Coulomb or Yukawa potential, enable fast convolutions in tensor-product formats. For large scale Hartree-Fock and DFT electronic structure calculations, based on GTO bases or cartesian grids, the computation of the nonlinear Hartree potential becomes a dominant step. Tensor-product formats provide possible improvements with respect to conventional approaches based on density-fitting schemes and FFT for GTO and grid based methods, respectively. An essential prerequisite, however, is the availability of accurate and fast compression algorithms, which have to be successively applied in order to avoid a disproportionate increase of the separation rank e.g. within tensor-product convolutions. Such kind of algorithms are presently under development [33, 73].

8 Acknowledgments

The authors gratefully acknowledge S. R. Chinnamsetty, M. Espig, Dr. L. Grasedyck, V. Khoromskaia and S. Schwinger (Leipzig) for useful discussions. This work was supported by the Deutsche Forschungsgemeinschaft (SPP 1145).

References

[1] J. Almlöf, *Elimination of energy denominators in Møller-Plesset perturbation theory by a Laplace transform approach*, Chem. Phys. Lett., 181 (1991), pp. 319–320.

[2] T. A. Arias, *Multiresolution analysis of electronic structure: Semicardinal and wavelet bases*, Rev. Mod. Phys., 71 (1999), pp. 267–312.

[3] A. Barinka, W. Dahmen and R. Schneider, *Fast computation of adaptive wavelet expansions*, Numer. Math., 105 (2007), pp. 549–589.

[4] K. R. Bates, A. D. Daniels and G. E. Scuseria, *Comparison of conjugate gradient density matrix search and Chebyshev expansion methods for avoiding diagonalization in large-scale electronic structure calculations*, J. Chem. Phys., 109 (1998), pp. 3308–3312.

[5] G. Beylkin and M. J. Mohlenkamp, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251.

[6] G. Beylkin and M. J. Mohlenkamp, *Algorithms for numerical analysis in high dimension*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.

[7] G. Beylkin, N. Coult and M. J. Mohlenkamp, *Fast spectral projection algorithms for density-matrix computations*, J. Comp. Phys., 152 (1999), pp. 32–54.

[8] G. Beylkin, R. Coifman and V. Rokhlin, *Fast wavelet transforms and numerical algorithms I*, Comm. Pure Appl. Math., 44 (1991), pp. 141–183.

[9] 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. Lançon, A. Neelov, J. Pivonski, and M. Rayson.

[10] D. Braess, *Nonlinear Approximation Theory*, Springer-Verlag, Berlin, 1986.

[11] D. Braess and W. Hackbusch, *Approximation of* $1/x$ *by exponential sums in* $[1, \infty)$, IMA J. Numer. Anal., 25 (2005), pp. 685–697.

[12] I. N. Bronshtein, K. A. Semendyayev, G. Musiol and H. Muehlig, *Handbook of Mathematics, 4th ed.* Springer-Verlag, Berlin, 2004.

[13] E. Cancès and C. Le Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, ESAIM: M2AN, 34 (2000), pp. 749–774.

[14] S. R. Chinnamsetty, M. Espig, W. Hackbusch, B. N. Khoromskij and H.-J. Flad, *Tensor product approximation with optimal rank in quantum chemistry*, J. Chem. Phys., 127 (2007), 084110 (14 pages).

[15] S. R. Chinnamsetty, H.-J. Flad, V. Khoromskaia and B. N. Khoromskij, *Tensor decomposition in electronic structure calculations on Cartesian grids*, MPI MIS 65, Leipzig 2007 (submitted).

[16] A. Cohen, *Numerical analysis of wavelet methods*, Elsevier, Amsterdam, 2003.

[17] A. Cohen, W. Dahmen and R. A. DeVore, *Adaptive wavelet methods for elliptic operator equations – convergence rates*, Math. Comp., 70 (2001), pp. 27–75.

[18] W. Dahmen, S. Prössdorf and R. Schneider, *Wavelet approximation methods for pseudodifferantial equations II: Matrix compression and fast solution*, Adv. Comput. Math., 1 (1993), pp. 259–335.

[19] W. Dahmen, S. Prössdorf and R. Schneider, *Multiscale methods for pseudo-differential equations on smooth closed manifolds*, Wavelet Anal. Appl., 5, Academic Press, San Diego, 1994.

[20] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica, 6 (1997), pp. 55–228.

[21] W. Dahmen, *Multiscale and wavelet methods for operator equations*, in Multiscale problems and methods in numerical simulations, Lecture Notes in Math., 1825, Springer–Verlag, Berlin, 2003, pp. 31–96.

[22] W. Dahmen, H. Harbrecht and R. Schneider, *Compression techniques for boundary integral equations—asymptotically optimal complexity estimates*, SIAM J. Numer. Anal., 43 (2006), pp. 2251–2271.

[23] W. Dahmen, H. Harbrecht and R. Schneider, *Adaptive methods for boundary integral equations: complexity and convergence estimates*, Math. Comp., 76 (2007), pp. 1243–1274.

[24] W. Dahmen, R. Schneider and Y. Xu, *Nonlinear functionals of wavelet expansions – Adaptive reconstruction and fast evaluation*, Numer. Math., 86 (2000), pp. 40–101.

[25] L. De Lathauwer, B. De Moor and J. Vandewalle, *On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.

[26] L. De Lathauwer, B. De Moor and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[27] L. De Lathauwer, B. De Moor and J. Vandewalle, *Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 295–327.

[28] R. A. DeVore, *Nonlinear Approximation*, Acta Numerica, 7 (1998), pp. 51–150.

[29] G. Deslauriers and S. Dubuc, *Symmetric iterative interpolation processes*, Constr. Approx., 5 (1989), pp. 49–68.

[30] M. Dolg, *Effective core potentials* in: Modern Methods and Algorithms of Quantum Chemistry, J. Grotendorst, ed., John Neumann Institute for Computing, NIC Series Vol. 1, Jülich, 2000 , pp. 479–508.

[31] B. I. Dunlap, *Robust and variational fitting*, Phys. Chem. Chem. Phys., 2 (2000), pp. 2113–2116.

[32] K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Auxiliary basis sets to approximate Coulomb potentials*, Chem. Phys. Lett., 240 (1995), pp. 283–290.

[33] M. Espig, *Efficient Best-Approximation with Sums of Elementary-Tensors in High Dimensions*, PhD thesis (in German), MPI-MIS Leipzig 2007.

[34] H.-J. Flad, W. Hackbusch, D. Kolb and R. Schneider, *Wavelet approximation of correlated wavefunctions. I. Basics*, J. Chem. Phys., 116 (2002), pp. 9641–9657.

[35] H.-J. Flad, W. Hackbusch and R. Schneider, *Best N-term approximation in electronic structure calculation. I. One-electron reduced density matrix*, ESAIM: M2AN, 40 (2006), pp. 49–61.

[36] H.-J. Flad, W. Hackbusch and R. Schneider, *Best N-term approximation in electronic structure calculation. II. Jastrow factors*, ESAIM: M2AN, 41 (2007), pp. 261–279.

[37] H.-J. Flad, R. Schneider and B.-W. Schulze, *Asymptotic regularity of solutions of Hartree-Fock equations with Coulomb potential*, preprint 5, Universität Potsdam 2007.

[38] I. P. Gavrilyuk, W. Hackbusch and B. N. Khoromskij, *$\mathcal{H}$-Matrix approximation for the operator exponential with applications*, Numer. Math., 92 (2002), pp. 83–111.

[39] I. P. Gavrilyuk, W. Hackbusch and B. N. Khoromskij, *Data-sparse approximation to operator-valued functions of elliptic operators*, Math. Comp. 73 (2004), pp. 1297–1324.

[40] I. P. Gavrilyuk, W. Hackbusch and B.N. Khoromskij, *Data-sparse approximation to a class of operator-valued functions*, Math. Comp. 74 (2005), pp. 681–708.

[41] I. P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij, *Tensor-product approximation to elliptic and parabolic solution operators in higher dimensions*, Computing 74 (2005), pp. 131–157.

[42] L. Genovese, T. Deutsch, A. Neelov, S. Goedecker and G. Beylkin, *Efficient solution of Poisson's equation with free boundary conditions*, arXiv:cond-mat/0605371v1 (2006).

[43] S. Goedecker, *Wavelets and their Application for the Solution of Differential Equations*. Presses Polytechniques Universitaires et Romandes, Lausanne, 1998.

[44] I. S. Gradshteyn and I.M. Ryzhik, *Tables of integrals, series and products, 6th edition*, Academic Press, San Diego, 2000.

[45] L. Grasedyck, *Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure*, Computing 72 (2004), pp. 247–265.

[46] L. Grasedyck and W. Hackbusch, *Construction and arithmetics of $\mathcal{H}$-matrices*, Computing 70 (2003), pp. 295–334.

[47] J. Garcke and M. Griebel, *On the computation of the eigenproblems of hydrogen and helium in strong magnetic and electric fields with the sparse grid combination technique*, J. Comp. Phys., 165 (2000), pp. 694–716.

[48] M. Griebel and J. Hamaekers, *Sparse grids for the Schrdinger equation*, ESAIM: M2AN, 41 (2007), pp. 215–247.

[49] M. Griebel and S. Knapek, *Optimized tensor-product approximation spaces*, Constr. Approx., 16 (2000), pp. 303–332.

[50] *W. Hackbusch, A sparse matrix arithmetic based on $\mathcal{H}$-matrices. I. Introduction to $\mathcal{H}$-matrices*, Computing, 62 (1999), pp. 89–108.

[51] W. Hackbusch and B.N. Khoromskij, *A sparse $\mathcal{H}$-matrix arithmetic. II. Application to multi-dimensional problems*, Computing, 64 (2000), pp. 21–47.

[52] W. Hackbusch and B. N. Khoromskij, *A sparse $\mathcal{H}$-matrix arithmetic: General complexity estimates*, J. Comp. Appl. Math., 125 (2000), pp. 479–501.

[53] *W. Hackbusch and B. N. Khoromskij, Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part I. Separable approximation of multi-variate functions*, Computing, 76 (2006), pp. 177–202.

[54] W. Hackbusch and B. N. Khoromskij, *Low-rank Kronecker product approximation to multi-dimensional nonlocal operators. Part II. HKT representations of certain operators*, Computing **76** (2006), pp. 203–225.

[55] W. Hackbusch, B. N. Khoromskij and R. Kriemann, *Hierarchical matrices based on a weak admissibility criterion*, Computing, 73 (2004), pp. 207–243.

[56] W. Hackbusch, B. N. Khoromskij and A. Klaus, *Approximation of functions by exponential sums based on Newton-type optimisation*, Techn. Rep. 3, MIP MIS, Leipzig 2005.

[57] W. Hackbusch, B. N. Khoromskij and E. E. Tyrtyshnikov, *Hierarchical Kronecker tensor-product approximations*. J. Numer. Math., 13 (2005), pp. 119–156.

[58] W. Hackbusch, B. N. Khoromskij and E. E. Tyrtyshnikov, *Approximate iterations for structured matrices*, Numer. Math., to appear.

[59] M. Häser and J. Almlöf, *Laplace transform techniques in Møller-Plesset perturbation theory*, J. Chem. Phys., 96 (1992), pp. 489–494.

[60] sc H. Harbrecht, *Wavelet Galerkin schemes for the boundary element method in three dimensions*, PhD thesis, Technische Universität Chemnitz, 2001.

[61] R. J. Harrison, G. I. Fann, T. Yanai, Z. Gan and G. Beylkin, *Multiresolution quantum chemistry: Basic theory and initial applications*, J. Chem. Phys., 121 (2004), pp. 11587–11598.

[62] T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory*, Wiley, New York, 1999.

[63] B. N. Khoromskij, *Structured rank-$(r_1, ..., r_d)$ decomposition of function-related tensors in $\mathbb{R}^d$*, Comp. Meth. in Applied Math., 6 (2006), pp. 194–220.

[64] B. N. Khoromskij, *Structured data-sparse approximation to high-order tensors arising from the deterministic Boltzmann equation*, Math. Comp., 76 (2007), pp. 1291–1315.

[65] B. N. Khoromskij, *Multi-linear approximation of higher-dimensional convolution in linear cost*, preprint MPI MIS Leipzig, 2007.

[66] B. N. Khoromskij and V. Khoromskaia, *Low Rank Tucker Tensor Approximation to the Classical Potentials*, Cent. Eur. J. Math., 5 (2007), pp. 1–28.

[67] W. Kohn, *Density functional theory for systems of very many atoms*, Int. J. Quantum Chem., 56 (1995), pp. 229–232.

[68] F. R. Manby, P. J. Knowles and A. W. Lloyd, *The Poisson equation in density fitting for the Kohn-Sham Coulomb problem*, J. Chem. Phys., 115 (2001), pp. 9144–9148.

[69] J. M. Millam and G. E. Scuseria, *Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations*, J. Chem. Phys., 106 (1997), pp. 5569–5577.

[70] K. Németh and G. E. Scuseria, *Linear scaling density matrix search based on sign matrices*, J. Chem. Phys., 113 (2000), pp. 6035–6041.

[71] P.-A. Nitsche, *Best N term approximation spaces for tensor product wavelet bases*, Constr. Approx., 24 (2006), pp. 49–70.

[72] C. Ochsenfeld and M. Head-Gordon, *A reformulation of the coupled perturbed self-consistent field equations entirely within a local atomic orbital density matrix-based scheme*, Chem. Phys. Lett., 270 (1997), pp. 399-405.

[73] I. V. Oseledets, D. V. Savostianov and E. E. Tyrtyshnikov, *Tucker dimensionality reduction of three-dimensional arrays in linear time*, SIAM J. Matrix Anal. Appl., 2007, to appear.

[74] R. G. Parr and W. Yang, *Density-functional Theory of Atoms and Molecules*, Oxford University Press, 1989.

[75] M. Reed and B. Simon, *Methods of Modern Mathematical Physiscs. IV: Analysis of Operators*, Academic Press, San Diego, 1978.

[76] R. Schneider, *Multiskalen- und Wavelet-Matrixkompression*, Teubner, Stuttgart, 1998.

[77] R. Schneider and T. Weber, *Wavelets for density matrix computation in electronic structure calculation.* Appl. Numer. Math., 56 (2006), pp. 1383-1396.

[78] S. Schwinger, Diploma thesis, MPI Leipzig and TU Chemnitz, 2005.

[79] F. Stenger, *Numerical Methods Based on Sinc and Analytic Functions*, Springer-Verlag, Heidelberg, 1993.

[80] V. N. Temlyakov, *Approximation of functions with bounded mixed derivative*, Proc. Steklov Inst. Math., 1989 (1989), pp. 275-293.

[81] P. Tchamitchian, *Wavelet functions and operators*, in Wavelets, Theory and Applications, G. Erlebacher, Y. Hussaini, L. Jameson, eds., Oxford University Press, 1996, pp. 87–178.

[82] E. E. Tyrtyshnikov, *Incomplete cross approximation in the mosaic-skeleton method*, Computing, 64 (2000), pp. 367–380.

[83] E. E. Tyrtyshnikov, *Tensor approximations of matrices generated by asymptotically smooth functions*, Sb. Math., 194 (2003), pp. 941–954.

[84] E. Tyrtyshnikov, *Kronecker-product approximations for some function-related matrices*, Linear Algebra Appl., 379 (2004), pp. 423-437.

[85] C. F. Van Loan, *The ubiquitous Kronecker product*, J. Comput. Appl. Math., 123 (2000), 85–100.

[86] C. F. Van Loan and N. P. Pitsianis, *Approximation with Kronecker products*, NATO Adv. Sci. Ser E Appl. Sci., 232, Kluwer, Dordrecht, 1993, pp. 293–314.

[87] H. Yserentant, *On the regularity of the electronic Schrödinger equation in Hilbert spaces of mixed derivatives*, Numer. Math., 98 (2004), pp. 731-759.

[88] H. Yserentant, *Sparse grid spaces for the numerical solution of the electronic Schrödinger equation*, Numer. Math., 101 (2005), pp. 381-389.

[89] T. Zhang and G.H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534-550.