Adaptive order polynomial algorithm in a multi-wavelet representation scheme

Antoine Durdek [a], Stig Rune Jensen[b], Jonas Juselius[c], Peter Wind[c], Tor Flå[a], Luca Frediani[c,*]

[a] *Centre for Theoretical and Computational Chemistry, Department of Mathematics, University of Tromsø, N-9037 Tromsø, Norway.*
[b] *Centre for Theoretical and Computational Chemistry, Department of Physics, University of Tromsø, N-9037 Tromsø, Norway.*
[c] *Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway.*

Abstract

We have developed a new strategy to reduce the storage requirements of a multivariate function in a multiwavelet framework. We propose that alongside the commonly used adaptivity in the grid refinement one can also vary the order of the representation k as a function of the scale n. In particular the order is decreased with increasing refinement scale. The consequences of this choice, in particular with respect to the nesting of scaling spaces, are discussed and the error of the approximation introduced is analyzed. The application of this method to some examples of mono- and multivariate functions shows that our algorithm is able to yield a storage reduction up to almost 60%. In general, values between 30 and 40% can be expected for multivariate functions. Monovariate functions are less affected but are also much less critical in view of the so called "curse of dimensionality".

Keywords:
Wavelets, Legendre polynomials, Representation, Optimization, Multiwavelets, Adaptivity, Compression

1. Introduction

Kohn–Sham DFT has proven to be a computationally cost-effective approach for both the theoretical modeling of molecules and for the modeling of extended, periodic systems [18]. Recently, linear-scaling based approaches have gradually been removing the boundaries between these two extremes[17, 11]. In current computational chemistry, the Kohn–Sham orbitals are for molecules in most cases represented in terms of basis sets consisting of Gaussian functions. The molecular orbitals $\Psi_i(\mathbf{r})$ are written as a linear combination of Gaussians:

*Corresponding author

$$\Psi_i(\mathbf{r}) = \sum_\mu C_{i\mu}\chi_\mu(\mathbf{r}_K) = \sum_\mu C_{i\mu}P_\mu(\mathbf{r}_K)\exp\left(-\alpha_\mu r_K^2\right) \quad (1)$$

where the expansion coefficients $C_{i\mu}$ are referred to as molecular orbital coefficients, and where we have indicated that the electronic coordinates are given relatively to the nuclear center K to which the Gaussian basis function is attached. $P_\mu(\mathbf{r}_K)$ denotes a Cartesian polynomial $x_K^i y_K^j z_K^k$. In principle the atomic basis set should be complete, thus infinite, but for practical reasons it is generally restricted to a few tens of functions for each atom in the molecule.

For extended periodic systems, the most convenient approach is the representation in terms of Gaussian plane waves[18, 9], which easily exploits the periodicity of the system, and allows the fast evaluation of the molecular integrals:

$$\Psi_i(\mathbf{r}) = \sum_{\mathbf{k}} C_{i\mathbf{k}}\exp(i\mathbf{k}\mathbf{r}) \quad (2)$$

where $\mathbf{k}$ is a three-dimensional wave vector.

Both approaches are somewhat inadequate when facing the challenge of modeling a large system which can be partitioned into a molecular subsystem and one or more extended or periodic structures. One would therefore like a separated representation that has approximate, algorithmic size-extensivity in the sense of a local and hierarchical scale adaptivity. More generally, finer approximations could be used in subunits of crucial importance for the molecular system at hand. For large molecules we believe a modular approach is essential to reflect the importance of the different subsystems for the quantum molecular problem under scrutiny.

A step in this direction is taken by allowing different meshes in regions of space as in multigrid [20] and multiresolution[4] techniques. Multiresolution analysis may be employed to provide a sparse and efficient representation of both operators and functions in that it allows a description of the system at different scales of resolution. Wavelet bases provide important properties for designing efficient numerical solution techniques: orthogonality, vanishing moments and compact support. The latter, which is particularly important in high dimension, enables a locally adaptive representation of functions: the grid is refined only where the current representation is not sufficient to reach the required precision in the computed results, thus yielding the coarsest grid compatible with the desired numerical precision of the result.

One important candidate multiscale method is the Multiwavelet basis which has been used by Harrison *et al.* [12, 13, 21], to represent Kohn-Sham molecular orbitals.

By making use of this approach we have in our group performed extensive tests to verify the linear scaling capabilities of the approach with respect to the system size[14] and of the ability to control the error within an arbitrary and predefined value[10, 14]. In both cases very good results have been achieved.

The main drawback of such a grid based approach compared to traditional ones based on Gaussian functions or plane waves is the large memory requirement associated with such methods: no explicit functional form is assumed, therefore the storage requirements for each function is very large, reaching several gigabytes, if high precision is requested. The problem can be partially addressed by parallelization, thereby exploiting distributed memory architectures. A complementary strategy is to reduce the memory footprint of each function. One such method has recently been proposed by Bischoff and coworkers[6, 7] who employed a rank-reduction based on Singular Value Decomposition.

In this paper, we will follow an alternative route to reducing the prefactor for the memory storage problem. We propose to make the order of the polynomial basis scale-dependent: $k = k(n)$. In particular, k will decrease with the grid refinement. The underlying assumption is that higher order polynomials are less important at finer scales to correctly represent cusp-like functions such as those needed to deal with molecular orbitals. It is instead more important to increase the grid refinement. Since the support of the basis is the same as for a fixed basis, the basis functions supported on different hypercubes will still be non-overlapping and therefore orthogonal. As will be shown Section 3, the main challenge posed by this approach is the lack of orthogonality between the scaling space V_k^n and the wavelet space $W_{k'}^n$ with $k' < k$. We have dealt with this problem by proposing an approximated representation. The algorithms necessary to construct it are given in Sec. 4 whereas a set of numerical tests is presented in Sec. 5 and discussed in Sec. 6.

2. Multiwavelet representation in 1D

Alpert was the first to describe the multiwavelet approach for the representation of functions and operators [1, 2]. His work is based on his description of Legendre scaling functions and the corresponding wavelet functions. In order to set the notations for Section 3, we briefly review here the main ideas. Let us define the scaling spaces V_k^n as:

$$V_k^n = \mathrm{Span}\{\phi_{il}^n(x) | i = 0, \cdots, k,\, l = 0, \cdots, 2^n - 1\} \tag{3}$$

where

$$\phi_{il}^n(x) = 2^{n/2}\sqrt{2i+1}\tilde{L}_i\left(2^n x - l\right), \tag{4}$$

and $\tilde{L}_i\left(x\right)$ is the $i-$th shifted Legendre polynomial on the interval $[0, 1]$:

$$\tilde{L}_i\left(x\right) = \begin{cases} L_i\left(2x-1\right) & x \in [0,1] \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

From the definition of $\tilde{L}_i\left(x\right)$ it follows that ϕ_{il}^n is zero outside the interval $[2^{-n}l, 2^{-n}(l+1)]$.

Legendre polynomials are chosen as a basis as they are obtained in a recursive manner and are orthonormal with respect to the scalar product

$$\langle f | g \rangle = \int_0^1 f(x) g(x)\, \mathrm{d}x \tag{6}$$

Moreover, the Legendre polynomial $L_i(x)$ has degree i implying that the polynomial basis spanning $V_{k'}$ $(k' < k)$ is a subset of the basis spanning V_k. We will largely exploit this in the next section: in order to change the order of the representation, one simply has to add or remove one or more basis functions keeping the other ones as they are.

By definition of the scaling spaces, one gets directly that :

$$V_k^0 \subset V_k^1 \subset \cdots \subset V_k^n \subset \cdots \tag{7}$$

and the number of basis functions at scale n is $\dim V_k^n = 2^n(k+1)$.

The wavelet spaces W_k^n are defined as the orthogonal complement of V_k^n with respect to V_k^{n+1}:

$$W_k^n \oplus V_k^n = V_k^{n+1}, \forall n \tag{8}$$

which implies that $\dim W_k^n = 2^n(k+1)$. If $\psi_i^0, i = 0, \cdots, k$ are the basis functions of W_k^0, then we have the following properties for the basis of W_k^n:

1. ψ_i^n is built as a piecewise polynomial function with a discontinuity in the middle of the interval since $\psi_i^n \in V_k^1$ and $\psi_i^n \notin V_k^0$.
2. $\psi_{il}^n(x) = 2^{n/2}\psi_i^n(2^n x - l)$
3. $\langle \phi_{il}^n | \psi_{jm}^{n'} \rangle = 0 \quad (n' \leq n)$
4. $\langle \psi_{il}^n | \psi_{jm}^{n'} \rangle = \delta_{nn'}\delta_{ij}\delta_{lm}$

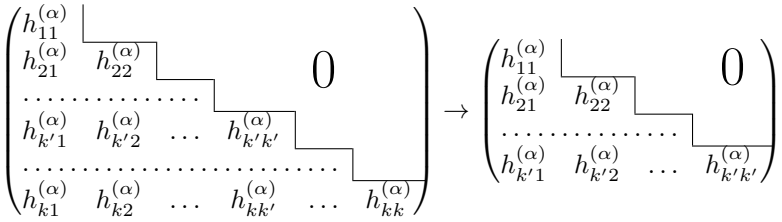
The freedom in the choice of basis functions for the wavelet space can be exploited by requiring additional properties. According to Ref. [3] it is possible to construct a basis such that:

1. ψ_i has $i + k$ vanishing moments
2. ψ_i is an odd (even) function with respect to inversion through the interval center $x = 0.5$ for even (odd) values of i.

According to Equation 8, one can describe a linear unitary transformation between the two bases via a matrix transformation, which collects the four filter matrices $G^{(0)}$, $G^{(1)}$, $H^{(0)}$ and $H^{(1)}$:

$$\begin{pmatrix} \boldsymbol{\phi}_l^n \\ \boldsymbol{\psi}_l^n \end{pmatrix} = \begin{pmatrix} H^{(0)} & H^{(1)} \\ G^{(0)} & G^{(1)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\phi}_{2l}^{n+1} \\ \boldsymbol{\phi}_{2l+1}^{n+1} \end{pmatrix} \tag{9}$$

where $\boldsymbol{\phi}_l^{n\,t} = (\phi_{1l}^n, \phi_{2l}^n \ldots \phi_{kl}^n)$ is a row-vector collecting all scaling basis functions at scale n in the l-th interval and $\boldsymbol{\psi}_l^{n\,t} = (\psi_{1l}^n, \psi_{2l}^n \ldots \psi_{kl}^n)$ similarly collects the corresponding wavelet basis functions. The transformation is unitary and scale-independent. We remark again that Legendre polynomials bases are constructed recursively adding one function to the previous basis. Consequently, the H filter matrices are lower triangular. Moreover, given two polynomial orders $k' < k$, the filters $H^{(\alpha)}(\alpha = 0, 1)$ for k' are simply submatrices of their k counterparts: they are obtained by removing $k - k'$ rows and columns at the bottom and on the right side, respectively. This structure is illustrated below and has been exploited in the design of our algoritms (see Sec. 4 for details).



As shown by Alpert *et al.* [2], the use of polynomials as scaling functions is based on the following theorem:

Theorem 1. *Let V_k^n be a scaling space described as above with polynomials as scaling functions on the interval $[0,1]$.*
Thus we have the following result:

1. $\lim_{k\to\infty} V_k^n = L^2([0,1])$
2. $\lim_{n\to\infty} V_k^n = L^2([0,1])$

The theorem shows that completeness in the L_2 norm sense can be achieved both by increasing the polynomial order and by refinement of the dyadic subdivisions along the ladder of scales.

For any function $f \in L^2$, the projected function $\mathcal{P}_k^n f = f_k^n$ of f on V_k^n can be written as:

$$f_k^n = \sum_{l=0}^{2^n-1} \sum_{i=0}^{k} f_{il}^n \phi_{il}^n \tag{10}$$

$$\text{where } f_{il}^n = \langle f | \phi_{il}^n \rangle \tag{11}$$

which is the finest-scale representation of f. Alternatively can f be decomposed the ladder of wavelet spaces:

$$f_k^n = f_k^0 + \sum_{m=0}^{n-1} df_k^m \tag{12}$$

$$= \sum_{i=0}^{k} f_i \phi_i + \sum_{m=0}^{n-1} \sum_{l=0}^{2^m-1} \sum_{i=0}^{k} df_{il}^m \psi_{il}^m \tag{13}$$

$$\text{where } f_i = \langle f | \phi_i \rangle \tag{14}$$

$$\text{and } df_{il}^m = \langle f | \psi_{il}^m \rangle \tag{15}$$

The two representations are equivalent and can be interconverted in one another by recursive application of the two-scale relation:

$$f_k^n + df_k^n = f_k^{n+1} \tag{16}$$

The two operations are generally called *reconstruction* (from the left-hand side to the right-hand side) and *decomposition*[2].

The error committed by projecting the function onto V_k^n is fully controlled and can be computed [15, 19]. The accuracy is set as a parameter and the approximation can be done arbitrarily close to the true function via scale refinement and variation on the order.

It is also useful to introduce a projector notation. If we indicate $\mathcal{P}_k^n$ and $\mathcal{Q}_k^n$ the projector onto V_k^n and W_k^n respectively. It then follows that

$$\mathcal{P}_k^n + \mathcal{Q}_k^n = \mathcal{P}_k^{n+1} \tag{17}$$

For $k' < k$ we will also define a *residual projector* $\mathcal{P}_{k,k'}^n$ as

$$\mathcal{P}_{k,k'}^n = \mathcal{P}_k^n - \mathcal{P}_{k'}^n \tag{18}$$

By definition of the wavelet projectors, and the previous relations the following relations can be easily proven:

$$\mathcal{Q}_k^n \mathcal{P}_k^n = \mathcal{P}_k^n \mathcal{Q}_k^n = \mathcal{Q}_k^n \mathcal{P}_{k'}^n = \mathcal{P}_{k'}^n \mathcal{Q}_k^n = \mathcal{Q}_k^n \mathcal{P}_{k,k'}^n = \mathcal{P}_{k,k'}^n \mathcal{Q}_k^n = 0 \tag{19}$$

$$\mathcal{Q}_{k'}^n \mathcal{P}_k^n = \mathcal{Q}_{k'}^n \mathcal{P}_{k,k'}^n \tag{20}$$

$$\mathcal{P}_k^n \mathcal{Q}_{k'}^n = \mathcal{P}_{k,k'}^n \mathcal{Q}_{k'}^n \tag{21}$$

$$\mathcal{P}_k^n \mathcal{P}_{k'}^n = \mathcal{P}_{k'}^n \mathcal{P}_k^n = \mathcal{P}_{k'}^n \tag{22}$$

As a corollary of the completeness theorem, for any normalized function $f \in L^2$ the following relations can be written for the projection operators:

$$\lim_{k\to\infty} \|\mathcal{P}_k^n f\|_{L^2} = \lim_{n\to\infty} \|\mathcal{P}_k^n f\|_{L^2} = 1 \tag{23}$$

$$\lim_{k\to\infty} \|\mathcal{Q}_k^n f\|_{L^2} = \lim_{n\to\infty} \|\mathcal{Q}_k^n f\|_{L^2} = 0 \tag{24}$$

3. Adaptive polynomial order representation

The representation of a multivariate function f at scale n in d dimensions with a tensorial multiwavelet basis of order k requires $2^{nd}(k+1)^d$ coefficients for the reconstructed representation at scale n. The accuracy of the representation can be increased either by augmenting the polynomial basis (larger k) or by further refinements (larger n), thus increasing drastically the data storage. In order to limit the memory requirement adaptivity is introduced, thereby refining the representation only where the predefined accuracy is not met.

We propose an additional way to reduce the data storage. Namely, instead of keeping the same polynomial order k at all scales we will assume that k can be chosen as a function of n with the limitation that $k(n) \leq k(n')$ for $n > n'$. Especially in high dimension, this could determine a reduction of the data storage requirements.

The challenging point of this approach is represented by the loss of exact inclusion of the vector space $V^n_{k(n)}$ into $V^{n+1}_{k(n+1)}$:

$$V^n_{k(n)} \subsetneq V^{n+1}_{k(n+1)} \text{ unless } k(n+1) = k(n) \tag{25}$$

Let us define $V^n_{\Delta k}$ implicitly as:

$$V^n_{k(n)} \stackrel{\text{def}}{=} V^n_{k(n+1)} \oplus V^n_{\Delta k} \tag{26}$$

$V^n_{\Delta k}$ is the subspace of $V^n_{k(n)}$ which is not entirely contained in $V^{n+1}_{k(n+1)}$. However $V^{n+1}_{k(n+1)}$ can be employed to approximate a function belonging to $V^n_{\Delta k}$. More specifically, since

$$V^{n+1}_{k(n+1)} = V^n_{k(n+1)} \oplus W^n_{k(n+1)} \tag{27}$$

then $V^n_{\Delta k}$ can be approximated by a corresponding subspace in $W^n_{k(n+1)}$. As an example let us consider V_3 and $V_2 \oplus W_2$. The cubic function in V_3 is orthogonal to V_2 but can be approximated as a piecewise quadratic function which belongs to W_2.

We have the following theorem for any polynomial of order k:

Theorem 2. *Let V^n_k be the scaling space of order k, V^n_{k-1} the scaling space of order $k-1$, at scale n Let $\mathcal{P}^n_k$, $\mathcal{P}^n_{k-1}$, $\mathcal{Q}^n_{k-1}$ be the projectors onto V^n_k, V^n_{k-1} and W^n_{k-1} respectively. Let us define:*

$$d^n_{k,k-1} = \sup_f \frac{||(1-\mathcal{Q}^n_{k-1})\mathcal{P}^n_{k,k-1}f||_{L^2}}{||\mathcal{P}^n_{k,k-1}f||_{L^2}} \tag{28}$$

where $f \in C^{(k)}([0,1])$. Then $d^n_{k,k-1} = 2^{-k}$

To put it simply, the theorem states that if f is locally smooth, the norm of the component of $\mathcal{P}^n_{k,k-1}f$ which falls outside W^n_{k-1} decays faster than $\mathcal{P}^n_{k,k-1}f$ itself, such that their ratio goes exponentially to zero with increasing k.

Proof. We assume, without loss of generality that $n = 0$. The result comes from the fact that truncated Legendre series converges with an exponential decay for finite support functions [8, 16]. By writing the projection of f onto V^k_0 as

$$\mathcal{P}^0_k f = \sum_{i=0}^{k} c_k \phi^0_k, \qquad c_i = \left\langle f | \phi^0_i \right\rangle \tag{29}$$

and substituting into Eq. (28) one gets:

$$\begin{aligned} d^0_{k,k-1} = &\frac{||c_k\phi^0_k - \mathcal{Q}^0_{k-1}c_k\phi^0_k||_{L^2}}{||c_k\phi^0_k||_{L^2}} = ||\phi^0_k - \mathcal{Q}^0_{k-1}\phi^0_k||_{L^2} = \\ &||\phi^0_k - (\mathcal{P}^1_{k-1} - \mathcal{P}^0_{k-1})\phi^0_k||_{L^2} = ||(I - \mathcal{P}^1_{k-1})\phi^0_k||_{L^2}. \end{aligned} \tag{30}$$

The first step follows from the normalization condition of the basis, and the last one is due to the orthogonality of ϕ_k^0 with respect to V_{k-1}^0. In order to simplify the rightmost expression we recall the two scale difference equation [2]:

$$\phi_{il}^n(x) = \sum_{j=0}^{k} h_{ij}^{(0)} \phi_{j,2l}^{n+1}(x) + h_{ij}^{(1)} \phi_{j,2l+1}^{n+1}(x) \tag{31}$$

where $h_{ij}^{(\alpha)}, (\alpha = 0, 1)$ is the ij element of the filter matrix H^α.

By expanding ϕ_k^0 in V_k^1 one gets:

$$(I - \mathcal{P}_{k-1}^1)\phi_k^0 = \sum_{i=0}^{k} \sum_{l=0}^{1} \phi_{jl}^1 \left\langle \phi_{jl}^1 \,\middle|\, \phi_k^0 \right\rangle - \sum_{i=0}^{k-1} \sum_{l=0}^{1} \phi_{jl}^1 \left\langle \phi_{jl}^1 \,\middle|\, \phi_k^0 \right\rangle = h_{kk}^{(0)} \phi_{k0}^1 + h_{kk}^{(1)} \phi_{k1}^1 \tag{32}$$

where we have made use of the definition of the filter coefficients in terms of the inner product of basis functions of V_k^0 and V_k^1 and we have exploited the construction of the Legendre basis to eliminate all the common terms. The norm expressed in Eq. (30) is then simply

$$||(I - \mathcal{P}_{k-1}^1)\phi_k^0||_{L^2} = \sqrt{(h_{kk}^{(0)})^2 + (h_{kk}^{(1)})^2} = \sqrt{2}|h_{kk}^{(0)}| \tag{33}$$

since $h_{ij}^{(0)} = (-1)^{i+j} h_{ij}^{(1)}$(See Ref. [2] for details).

In order to prove the theorem we need to show that $h_{kk}^{(0)} = 2^{-k-1/2}$. Starting from:

$$h_{ij}^0 = \left\langle \phi_{i0}^0 \,\middle|\, \phi_{j0}^1 \right\rangle = \sqrt{2} \int_0^{1/2} \phi_i(x) \phi_j(2x) \, \mathrm{d}x, \tag{34}$$

we recall that $\phi_j(x)$ are the (normalized) shifted Legendre polynomials (see Eq. (4)) and we make the substitution $y = 2x$ obtaining:

$$h_{ij}^0 = \frac{\sqrt{(2i+1)(2j+1)}}{\sqrt{2}} \int_0^1 \tilde{L}_i\left(y/2\right) \tilde{L}_j\left(y\right) \, \mathrm{d}y \tag{35}$$

For the shifted legendre polynomials, the following formulation of the Rodrigues formula holds:

$$\tilde{L}_i\left(x\right) = \frac{1}{i!} \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^i \left[x(x-1) \right]^i. \tag{36}$$

By applying the Rodrigues formula to $\tilde{L}_i\left(y/2\right)$ we get:

$$\begin{aligned}
\tilde{L}_i\left(y/2\right) =& \frac{1}{2^i i!}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^i \left[y(y-1)-y\right]^i \\
=& \frac{1}{2^i i!}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^i \sum_{p=0}^{i}\binom{i}{p}(-y)^p\left[y(y-1)\right]^{i-p} \\
=& \frac{1}{2^i i!}\sum_{p=0}^{i}\sum_{q=0}^{i}\binom{i}{p}\binom{i}{q}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^q (-y)^p \left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^{i-q}\left[y(y-1)\right]^{i-p} \\
=& \frac{1}{2^i}\sum_{p=0}^{i}\sum_{q=0}^{p}\binom{i}{q}\frac{(-1)^p y^{p-q}}{(p-q)!(i-p)!}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^{i-q}\left[y(y-1)\right]^{i-p} \\
=& \frac{1}{2^i}\sum_{p=0}^{i}\sum_{q=0}^{p}\binom{i}{q}\frac{(-1)^p y^{p-q}}{(p-q)!}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^{p-q}\frac{1}{(i-p)!}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^{i-p}\left[y(y-1)\right]^{i-p} \\
=& \frac{1}{2^i}\sum_{p=0}^{i}\sum_{q=0}^{p}\binom{i}{q}\frac{(-1)^p y^{p-q}}{(p-q)!}\left(\frac{\mathrm{d}}{\mathrm{d}y}\right)^{p-q}\tilde{L}_{i-p}\left(y\right)
\end{aligned} \tag{37}$$

thus we have expanded $\tilde{L}_i\left(y/2\right)$ in a combination of shifted Legendre polynomials. This expression can now be inserted in Eq. (35). For $i=j=k$, due to orthogonality of the Legendre polynomials, only the term where $p=q=0$ gives contribution to the integral because all other terms contain lower order polynomials which are orthogonal to $\tilde{L}_k\left(x\right)$ by construction. Recalling that

$$\int_0^1 \tilde{L}_i\left(x\right)\tilde{L}_j\left(x\right)\,\mathrm{d}x = \frac{\delta_{ij}}{2i+1}, \tag{38}$$

we finally obtain:

$$h^0_{kk} = \frac{2k+1}{2^{k+1/2}}\int_0^1 \tilde{L}_k\left(y\right)\tilde{L}_k\left(y\right)\,\mathrm{d}y = \frac{1}{2^{k+1/2}} \tag{39}$$

proving that $d^0_{k,k-1} = \sqrt{2}h^{(0)}_{kk} = 2^{-k}$.

□

For any fixed $\Delta k > 1$, one could begin by defining $d^n_{k,k'}$ in analogy to Eq. (28), then proceed by exchanging $c^0_k\phi^0_k$ for $\tilde{\phi} = \sum_{j=k'+1}^{k} c_j\phi^0_j$ in Eq. (30) and assume $\|\tilde{\phi}\|_{L^2} = 1$. As shown by Alpert [1], $\|f - \mathcal{P}^n_{k'}f\|_{L^2}$ converges exponentially to zero:

$$\|(I-\mathcal{P}^n_{k'})f\|_{L^2} \leq 2^{-nk}\frac{2}{4^k\cdot k!}\sup|f^{(k)}(x)|. \tag{40}$$

In order to show exponential convergence in the limit of $k\to\infty$, one would additionally need to assume that there exists a $C>0$ such that $\sup|\tilde{\phi}^{(k)}| \leq$

C, $\forall\ k > 0$. With this assumption the exponential decay is a consequence of Alpert's bound.

A way to interpret the result is by realizing that $d^0_{k,k'}$ represents the residual norm of a unit vector v in $V^0_{k,k'}$ after its component in $W^0_{k'}$ has been projected out. In other words, for large enough k, and smooth functions ($f \in C^{(k)}([0,1])$), the space spanned by $V^n_{k,k'}$ becomes almost collinear with a corresponding subspace of $W^n_{k'}$. This near-collinearity can be expressed in terms of the projectors as:

$$\mathcal{P}^n_{k,k'} \simeq \mathcal{P}^n_{k,k'}\mathcal{Q}^n_{k'} \simeq \mathcal{Q}^n_{k'}\mathcal{P}^n_{k,k'}, \tag{41}$$

where true equivalence would hold if the space $V^0_{k,k'}$ were a subspace of $W^0_{k'}$.

3.1. Projection onto $V^n_{k(n)}$ and $W^n_{k(n+1)}$

The projection step consists in the computation of the function representation in the ladder of scaling and wavelet spaces. More in detail for each scale n, the projection $f^n_k = \mathcal{P}^n_k f$ can for instance be obtained via a quadrature scheme.

The wavelet component $df^n_{k'}$ is obtained by noticing that:

$$f^{n+1}_{k'} = f^n_{k'} + df^n_{k'} \tag{42}$$

For the sake of brevity we have assumed that $k = k(n)$ and $k' = k(n+1)$.

In this way we obtain at each scale a scaling part f^n_k and a wavelet part $df^n_{k'}$. We underline here that the two components are not orthogonal as $W^n_{k'}$ is only orthogonal to the first k' polynomials of V^n_k.

The projection down to the finest scale requires only the knowledge of $k(n)$ for each scale n starting from a predefined maximum value $k_{max} = k(0)$ until a minimum value $k_{min} = k(n_{min})$. Thereafter the polynomial order is kept constant at $k = k_{min}$

3.2. Reconstruction: $V^n_{k(n)} + W^n_{k(n+1)} \to V^{n+1}_{k(n+1)}$

The reconstruction step consists in obtaining the scaling representation at the finest scale by making use of the scaling component at the coarsest scale $f^0_{k(0)}$ and the ladder of wavelet components $df^n_{k(n)}$. Assuming again $k = k(n)$ and $k' = k(n+1)$, the reconstruction step at each scale can be achieved by the following procedure.

First the polynomial part of f^n_k from $k'+1$ to k is projected out:

$$f^n_{k'} = (1 - \mathcal{P}^n_{k,k'})f^n_k = \mathcal{P}^n_{k'}f^n_k \tag{43}$$

then the scaling representation $f^{n+1}_{k'}$ is obtained by assembling:

$$f^{n+1}_{k'} = f^n_{k'} + df^n_{k'} \tag{44}$$

The procedure is repeated iteratively, scale by scale along the tree structure. As there is no overlap between neighboring nodes the iteration is carried on until a local finest scale, which is determined by the precision requirements.

3.3. Analysis: $V^{n+1}_{k(n+1)} \to V^n_{k(n)} + W^n_{k(n+1)}$

The analysis or compression step is the inverse transformation of the reconstruction, in the sense that it consists in obtaining the scaling component at scale $n = 0$ and the wavelet components at all scales from the reconstructed representation f^n_k at the finest scale. This is achieved iteratively, starting at the finest scale. The difference with respect to the standard algorithm is represented by the fact that, given a representation of f in $V^{n+1}_{k'}$ we want to obtain a representation in V^n_k where $k > k'$.

The first step consists in transforming $f^{n+1}_{k'}$ into the corresponding wavelet and scaling components at scale n:

$$f^{n+1}_{k'} = f^n_{k'} + df^n_{k'} \tag{45}$$

The second step consists in "transferring" the component of $df^n_{k'}$ which is collinear to V^n_k to the scaling part in an approximate way by making use of Eq. (41):

$$\begin{aligned} f^n_{k'} + df^n_{k'} &= \mathcal{P}^n_{k'} f + \mathcal{Q}^n_{k'} f \\ &= \mathcal{P}^n_{k'} f + (1 - \mathcal{P}^n_{k,k'} + \mathcal{P}^n_{k,k'}) \mathcal{Q}^n_{k'} f \\ &\simeq \mathcal{P}^n_{k'} f + \mathcal{P}^n_{k,k'} f + (1 - \mathcal{P}^n_{k,k'}) \mathcal{Q}^n_{k'} f \\ &= \mathcal{P}^n_k f + (1 - \mathcal{P}^n_{k,k'}) \mathcal{Q}^n_{k'} f = f^n_k + d\tilde{f}^n_{k'} \end{aligned} \tag{46}$$

In the last step we have implicitly defined $d\tilde{f}^n_k = (1 - \mathcal{P}^n_{k,k'}) \mathcal{Q}^n_{k'} f$.

In this way the scheme to achieve an approximate representation of f on V^n_k based on the representation in $V^{n+1}_{k'}$ is complete. Repeating this procedure iteratively from $n = n_{max}$ to $n = 0$ leads to a representation of f onto $V^0_{k(0)} \oplus W^0_{k(1)} \oplus \cdots \oplus W^{n_{max}-1}_{k(n_{max})}$.

3.4. Multivariate functions

For multivariate functions a tensor product representation is employed. The projector onto the scaling space at each scale is:

$$\mathcal{P}^n_k = \bigotimes_{i=1}^{d} \mathcal{P}^{n,i}_k \tag{47}$$

whereas the projector onto the wavelet space is obtained as the difference between two successive scales:

$$\mathcal{Q}^n_k \stackrel{\text{def}}{=} \mathcal{P}^{n+1}_k - \mathcal{P}^n_k = \bigotimes_{i=1}^{d} \mathcal{P}^{n+1,i}_k - \bigotimes_{i=1}^{d} \mathcal{P}^{n,i}_k \tag{48}$$

Similarly, we can define the residual projector as:

$$\mathcal{P}^n_{k,k'} \stackrel{\text{def}}{=} \mathcal{P}^n_k - \mathcal{P}^n_{k'} = \bigotimes_{i=1}^{d} \mathcal{P}^{n,i}_k - \bigotimes_{i=1}^{d} \mathcal{P}^{n,i}_{k'} \tag{49}$$

As for the monovariate case we can write the approximate relationship (41) which can be derived from the monovariate case by exploiting the tensor product structure:

$$\begin{aligned}
\mathcal{Q}^n_{k'}\mathcal{P}^n_{k,k'} &= \left(\bigotimes_{i=1}^{d}\mathcal{P}^{n+1,i}_{k'} - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'}\right)\left(\bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k} - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'}\right) \\
&= \bigotimes_{i=1}^{d}\mathcal{P}^{n+1,i}_{k'}\mathcal{P}^{n,i}_{k} - \bigotimes_{i=1}^{d}\mathcal{P}^{n+1,i}_{k'}\mathcal{P}^{n,i}_{k'} - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'}\mathcal{P}^{n,i}_{k} + \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'}\mathcal{P}^{n,i}_{k'} \\
&= \bigotimes_{i=1}^{d}\mathcal{P}^{n+1,i}_{k'}\mathcal{P}^{n,i}_{k} - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'} = \bigotimes_{i=1}^{d}\left(\mathcal{P}^{n,i}_{k'} + \mathcal{Q}^{n,i}_{k'}\right)\mathcal{P}^{n,i}_{k} - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'} \\
&= \bigotimes_{i=1}^{d}\left(\mathcal{P}^{n,i}_{k'} + \mathcal{Q}^{n,i}_{k'}\mathcal{P}^{n,i}_{k}\right) - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'} \\
&\simeq \bigotimes_{i=1}^{d}\left(\mathcal{P}^{n,i}_{k'} + \mathcal{P}^{n,i}_{k,k'}\mathcal{P}^{n,i}_{k}\right) - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'} = \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k} - \bigotimes_{i=1}^{d}\mathcal{P}^{n,i}_{k'} = \mathcal{P}^n_{k,k'}
\end{aligned} \tag{50}$$

We further underline that in the multivariate case, when the polynomial order is reduced from k to k' the number of components which need to be discarded as described in Sec. 3.2 or transferred from $W^n_{k'}$ to V^n_k as described in Sec. 3.3 is now $(k+1)^d - (k'+1)^d$: in other words it is the difference between the d-dimensional hypercube of length $k+1$ and the one of length $k'+1$ (e.g. for $d=3$ and $k'=k-1$ the number of discarded/transferred components is $3k^2+3k+1$).

4. Algorithms

In this section, we present the details of our algorithm. Legendre basis functions are used for scaling functions: thanks to the construction of Legendre polynomials, only one scaling function is involved in the process. The construction of the wavelet basis [3] with additional vanishing moments is directly linked to the non-orthogonality between high order polynomial and the wavelet basis. In the simplest case where the polynomial order $k(n)$ is lowered by one at each successive scale, only the first wavelet function ψ_0 is not orthogonal to ϕ_k. All other inner products are zero by construction. E.g. for $k=3$, this is equivalent to approximating the cubic function ϕ_k by the piecewise polynomial ψ_0 which is made of two adjacent parabolas, supported respectively on $[0,1/2]$ and $[1/2,1]$. Increasing the order will, as proved in Theorem 2, lead to better approximations in the L_2-norm sense. In practice one only needs to "move" one projection coefficient for each node: the coefficient representing the projection onto ψ^n_{0l} will instead be used for the projection onto ϕ^n_{kl} or vice versa. This means that there is no additional loss of information or deterioration of the representation by performing successive reconstructions/decompositions.

Algorithm 1 illustrates the projection of a function employing our adaptive scheme. At each scale, starting from the coarsest one the scaling part of the function $f^n_{k(n),l}$ is computed. Then the wavelet part $df^n_{k(n+1),l}$ Is computed by switching to the polynomial basis $k(n+1)$. The norm of the wavelet part is then checked locally for each node against the required precision to determine whether refinement is necessary. We remark that $df^n_{k(n+1),l}$ is the projection of f in the wavelet space of the selected node, therefore $\|df^n_{k(n+1),l}\|_{L^2} = \sum_{j=0}^{k(n+1)} |df^n_{jl}|^2$, where $df^n_{k(n+1),l}$ is defined in Eq. (15).

Algorithm 1 Adaptive projection algorithm for a function f with a given accuracy ε

```
01 For each scale n
02     For each available node l at the current scale
03         Compute f^n_{k(n),l}
04         Compute df^n_{k(n+1),l}
05         If(‖df^n_{k(n+1),l}‖_{L^2} > 2^{-n}ε)
06         allocate child nodes and mw-transform coefficients
07     next node
08 next scale
```

Algorithm 2 describes the compression of a function: it is here assumed that the function is represented at the local finest scale as $f^n_{k(n)}$ and all child nodes are present to reconstruct the parent. Starting at the next finest scale $n = n_{max} - 1$, the scaling part $f^n_{k(n+1)}$ and the wavelet part $df^n_{k(n+1)}$ of each node are obtained from its children through a standard Multiwavelet (MW) transform. If $k(n) > k(n+1)$, the scaling part is augmented to $f^n_{k(n)}$ by making use of Eq. (46) and the wavelet part is correspondingly purged. In practice thanks to the Alpert construction of the basis set, this implies that one or more coefficients are simply transferred from the wavelet to the scaling part. The sequence is repeated for all nodes at the current scale n before moving to scale $n-1$.

Algorithm 2 Compression algorithm

```
01 For each scale from n = n_max − 1 to n = 0
02     For each node l at the current scale
03         Obtain f^n_{k(n+1)} and df^n_{k(n+1)} from f^{n+1}_{k(n+1)}
04         If (k(n) > k(n+1))
05             Transform f^n_{k(n+1)} + df^n_{k(n+1)} into f^n_{k(n)} + df̃^n_{k(n+1)}
06      next node
07 previous scale
```

Algorithm 3 shows the reconstruction of the finest-scale representation of a function. Such a function is represented through $f^0_{k(0)}$ plus the modified wavelet

part at each scale $d\tilde{f}^n_{k(n+1)}$. Starting at the coarsest scale $n = 0$, the correct scaling and wavelet components $f^n_{k(n+1)}$ and $df^n_{k(n+1)}$ are obtained by making use of Eq. (46) if $k(n) > k(n+1)$. As for the compression algorithm, this implies that one or more coefficients are simply transferred, this time from the scaling to the wavelet part. The scaling representation of the child nodes $f^n_{k(n+1)}$ is then obtained by a MW-transform. The sequence is repeated for all nodes at the current scale n before moving to scale $n + 1$.

Algorithm 3 Reconstruction algorithm

```
01 For each scale from n = 0 to n = n_max − 1
02     For each node l at the current scale
03         If (k(n) > k(n + 1))
04             Transform f^n_{k(n)} + df̃^n_{k(n+1)} into f^n_{k(n+1)} + df^n_{k(n+1)}
05         Compute f^{n+1}_{k(n+1)} from f^n_{k(n+1)} and df^n_{k(n+1)}
06      next node
07 previous scale
```

5. Numerical results

In order to test the effectiveness of our approach we have selected some test functions and we have compared the amount of memory required to represent them on the one hand by making use of a regular MW-representation for a given polynomial order k and a given accuracy ϵ, and on the other hand with our decreasing order approach.

The chosen functions are Gaussian functions and so-called Slater type orbitals ($f(x) = Ae^{(-\alpha|x-x_0|)}$) which display a cusp-like singularity for $x = x_0$. Both examples are mutated from quantum chemistry as the former is the most widespread choice to build a basis set, whereas the latter is nowadays less common but has the appropriate behavior: a cusp at the atomic center and exponential asymptotic decay for large distances.

The parameterization employed for $k(n)$ is shown in Fig. 1. The polynomial order is kept fixed at k_{max} from $n = 0$ to a given n_0. It is then decreased by one at each successive scale up to n_1 and finally kept constant for all successive scales at $k_{min} = k_{max} - (n_1 - n_0)$. This strategy has been chosen to be able to adjust the range of scales where the order reduction takes place, keeping at the same time the structure as simple as possible.

Table 1 and Table 2 collect the results for two one-dimensional Gaussians with exponents $\alpha = 50$ and $\alpha = 10000$ respectively. For each of them we have reported the number of coefficients required to represent the function with the standard MW-representation and polynomial order k_{max} and with decreasing order scheme. The parameterization of $k(n)$ is also reported through the values of k_{min} (minimum allowed order) and n_0 (starting scale for order reduction). Our results show that a reduction of the size of the representation can

be achieved in most cases by the appropriate choice of $k(n)$. In a few cases no reduction is possible indicating that the parameterization provided by the standard MW-representation is already optimal.

The results collected for the two three-dimensional Gaussians are reported in Table 3 and Table 4, respectively. By comparison with the results obtained in the one-dimensional case, an enhancement of the compression achieved with a decreasing-order scheme can be observed. In particular the following remarks can be made: (1) the reduction of the number of coefficients needed for the representation can be achieved in all cases tested, (2) the compression achieved is consistently larger than for the monovariate case; (3) the decreasing order scheme has a stronger impact on the narrow Gaussian (large exponent α), which is also the one requiring a larger representation.

The achieved compression expressed as percent reduction of the size of the representation for the Gaussian functions of Tables 1, 2, 3 and 4 is also reported in Fig. 2.

Table 5 summarizes the same kind of information for a non-centered one-dimensional Slater-type orbital, with exponent parameter $\alpha = 100$. The function is off-centered ($x_0 = 0.27$) in order to avoid the singularity to be on a discretization point and hence take artificially advantage of it. The table contains the number of coefficients required both for the standard representation with a fixed order $k = k_{max}$, and for the corresponding adaptive order representation. Our results highlight a reduction of the total number of coefficients in all cases. We have observed that in most cases the best parameterization is achieved when $k(n)$ is chosen such that k_{min} is reached at the finest scale N.

The results for the off-centered three-dimensional Slater orbital are presented in Table 6. The parameters are $\alpha = 100$ and $x_0 = (0,27; 0,27; 0,27)$. Also in this case, compared to the monodimensional one, a more consistent behavior is observed. Compression is achieved for all choices of initial order k_{max} and a more pronounced compression rate is observed compared to the monovariate case.

The achieved compression expressed as percent reduction of the size of the representation for the Slater-type functions of Tables 5 and 6 is also reported in Fig. 3.

6. Discussion

The numerical results of the previous section, (see for a summary Fig 2 and 3) show that in most cases, a compression of the memory needed to represent a single function can be achieved. Two clear distinctions can be drawn: on the one hand the compression achieved for functions presenting short-scale variations (a Gaussian with a large exponent or a cusp) is more significant; at the same time the effect of compression is clearly more pronounced for a multivariate function than for a monovariate one. The latter consideration is motivated by the the fact that in a standard MW-representation the number of coefficients at scale n is proportional to $(k+1)^{nd}$, therefore the effect of order reduction is amplified. For

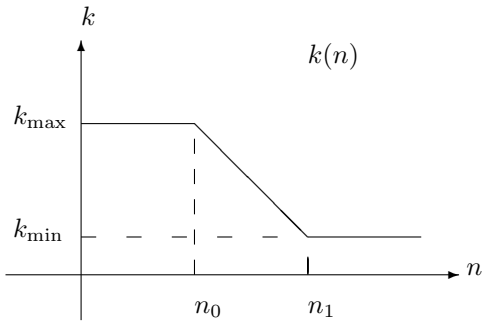


Figure 1: Generic shape of the order $k(n)$

k_{max}	SR	DOR	k_{min}	n_0	%
5	180	180	5	0	0
6	210	182	5	0	13
7	176	168	5	2	5
8	126	126	8	0	0
9	140	128	8	0	9
10	154	120	8	0	13
11	168	148	8	0	12
12	182	162	8	0	11
13	84	84	13	0	0
14	90	86	13	0	4
15	96	92	13	0	4

Table 1: Comparison of standard MW-representation (SR) with the decreasing-order representation (DOR) for a centered one-dimensional Gaussian function with $\alpha = 50$. The number of coefficients for the two representations (second and third column) is expressed as a function of the initial polynomial order k_{max}. For SR the initial order k_{max} is used throughout whereas for the DOR the function $k(n)$ is equal to k_{max} until $n = n_0$ and then decreased by one at each successive refinement until k_{min} is reached. The last column (%) is expressing the compression achieved as the percent reduction in the representation size in terms of number of coefficients.

k_{max}	SR	DOR	k_{min}	n_0	%
5	564	564	5	0	0
6	434	434	6	0	0
7	496	436	6	0	12
8	414	414	8	0	0
9	460	416	8	0	10
10	506	422	8	0	17
11	552	436	8	0	21
12	598	458	8	0	23
13	532	488	8	0	8
14	570	526	8	0	8
15	608	572	8	0	9

Table 2: Comparison of standard MW-representation (SR) with the decreasing-order representation (DOR) for a centered one-dimensional Gaussian function with $\alpha = 10000$. The number of coefficients for the two representations (second and third column) is expressed as a function of the initial polynomial order k_{max}. For SR the initial order k_{max} is used throughout whereas for the DOR the function $k(n)$ is equal to k_{max} until $n = n_0$ and then decreased by one at each successive refinement until k_{min} is reached. The last column (%) is expressing the compression achieved as the percent reduction in the representation size in terms of number of coefficients.

k_{max}	SR	DOR	k_{min}	n_0	%
5	568512	568512	5	0	0
6	375928	310904	5	2	17
7	561152	323072	5	1	42
8	425736	324808	5	0	24
9	584000	427904	8	0	27
10	777304	447896	8	0	42
11	1009152	611008	9	0	39
12	1283048	809640	9	0	37
13	197568	197568	13	0	0
14	243000	202616	13	0	17
15	294912	248768	14	0	16

Table 3: Comparison of standard MW-representation (SR) with the decreasing-order representation (DOR) for a centered three-dimensional Gaussian function with $\alpha = 50$. The number of coefficients for the two representations (second and third column) is expressed as a function of the initial polynomial order k_{max}. For SR the initial order k_{max} is used throughout whereas for the DOR the function $k(n)$ is equal to k_{max} until $n = n_0$ and then decreased by one at each successive refinement until k_{min} is reached. The last column (%) is expressing the compression achieved as the percent reduction in the representation size in terms of number of coefficients.

k_{max}	SR	DOR	k_{min}	n_0	%
5	1453248	1266880	4	7	13
6	1605240	1601144	4	6	0
7	1609728	1523200	4	6	5
8	1918728	1611464	7	0	16
9	2632000	1627520	7	0	38
10	3503192	1758616	7	0	50
11	4548096	2032832	7	0	55
12	5782504	2441384	8	0	58
13	5817280	2987264	8	0	49
14	7155000	3778936	8	0	47
15	8683520	4856768	9	0	44

Table 4: Comparison of standard MW-representation (SR) with the decreasing-order representation (DOR) for a centered three-dimensional Gaussian function with $\alpha = 100$. The number of coefficients for the two representations (second and third column) is expressed as a function of the initial polynomial order k_{max}. For SR the initial order k_{max} is used throughout whereas for the DOR the function $k(n)$ is equal to k_{max} until $n = n_0$ and then decreased by one at each successive refinement until k_{min} is reached. The last column (%) is expressing the compression achieved as the percent reduction in the representation size in terms of number of coefficients.

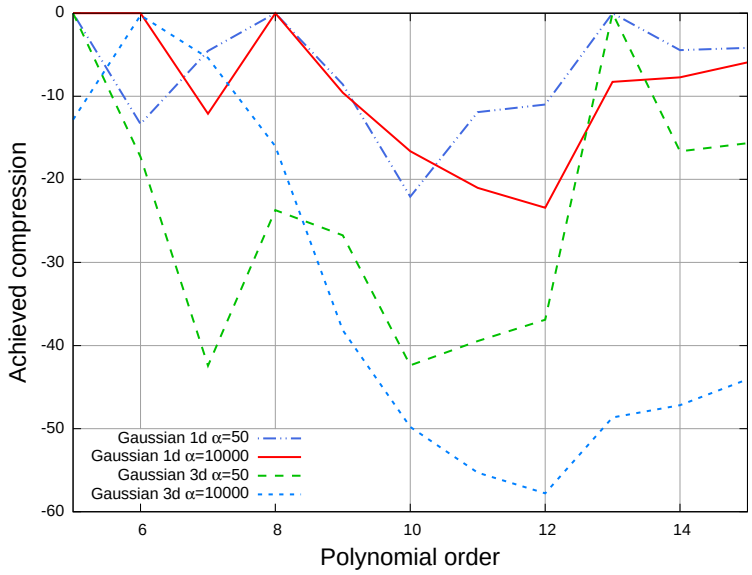


Figure 2: Percentage of coefficients gain in function of the order k_{max} for the Gaussian-type function in the one- and three-dimensional case and $\alpha = 50, 10000$. The data corresponds to the last column of the corresponding Tables.

k_{max}	SR	DOR	k_{min}	n_0	%
5	792	792	5	0	0
6	840	796	5	0	5
7	832	740	5	7	11
8	936	776	5	6	17
9	960	752	5	6	22
10	1056	780	5	5	26
11	1152	800	5	4	31
12	1196	816	5	3	32
13	1176	824	5	1	30
14	1320	828	5	0	37
15	1344	840	5	0	39

Table 5: Comparison of standard MW-representation (SR) with the decreasing-order representation (DOR) for a off-centered one-dimensional Slater function with $\alpha = 100$. The number of coefficients for the two representations (second and third column) is expressed as a function of the initial polynomial order k_{max}. For SR the initial order k_{max} is used throughout whereas for the DOR the function $k(n)$ is equal to k_{max} until $n = n_0$ and then decreased by one at each successive refinement until k_{min} is reached. The last column (%) is expressing the compression achieved as the percent reduction in the representation size in terms of number of coefficients.

k_{max}	SR	DOR	k_{min}	n_0	%
5	2004481	2004481	5	0	0
6	2129344	2012608	5	0	5
7	2195456	2054268	5	0	6
8	2472768	2091456	5	0	14
9	3008000	2174464	5	0	28
10	4174016	2216000	6	0	47
11	4091904	2367872	6	0	42
12	4921280	2640064	6	0	46
13	5795328	2679168	5	0	54
14	7128000	3049152	5	0	57
15	8650752	3706496	5	0	57

Table 6: Comparison of standard MW-representation (SR) with the decreasing-order representation (DOR) for a off-centered three-dimensional Slater function with $\alpha = 100$. The number of coefficients for the two representations (second and third column) is expressed as a function of the initial polynomial order k_{max}. For SR the initial order k_{max} is used throughout whereas for the DOR the function $k(n)$ is equal to k_{max} until $n = n_0$ and then decreased by one at each successive refinement until k_{min} is reached. The last column (%) is expressing the compression achieved as the percent reduction in the representation size in terms of number of coefficients.

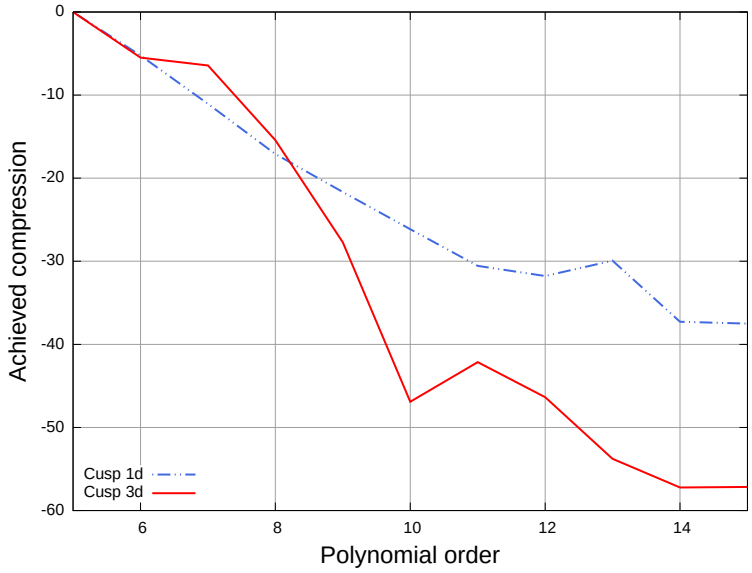


Figure 3: Percentage of coefficients gain in function of the order k_{max} for the the Slater-type function in the one- and three- dimensional case with $\alpha = 100$. The data corresponds to the last column of the corresponding Tables.

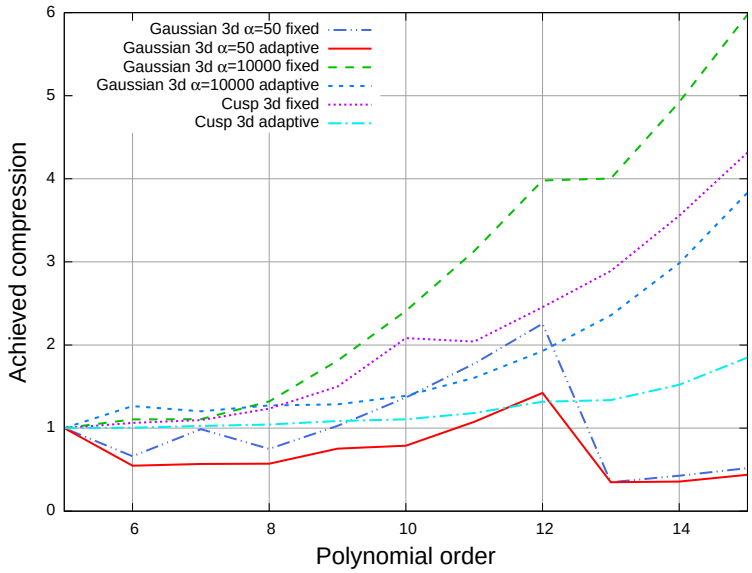


Figure 4: Relative variation on the number of coefficients for the Gaussian type function with $\alpha = 10000$ and Slater type with $\alpha = 100$ in the three-dimensional case. For the two functions, the SR and the DOR are presented. The relative variation $r(k)$ is obtained with respect to the order $k_{ref} = 5$. Writing $N(k)$, the number of coefficients needed at order k, we compute $r(k)$ as $r(k) = N(k)/N(k_{ref})$ (so that $r(5) = 1$ for any of the representation)

the least-effective case (a monovariate Gaussian with small exponent, $\alpha = 50$) the representation is however small to start with and the lack of a significant compression is to be expected.

Concerning the parameterization of $k(n)$ (the order k employed at each scale n) we observed that within a certain range, for all the examples shown a certain degree of compression can be achieved. In practice, the parameterization $k_{max} \in [8, 12], k_{min} = 5, n_0 = 0$ leads to a moderate compression for the monovariate functions and 30% or better in the multivariate case.

It is also interesting to observe what happens to the total number of coefficients needed while increasing the order k_{max}. Such data are summarized in Fig. 4 for the multivariate functions. In the standard case, the representation size soon becomes larger with increasing k (the representation of the chosen narrow multivariate Gaussian with $k = 15$ becomes six times larger than the one with $k = 5$) both for the narrow Gaussian and the cusp. The wide Gaussian is however less sensitive to the choice of k until $k = 13$, when a significant reduction is observed. By decreasing the order one sees that the overall size of the representation stays almost constant in the beginning and becomes larger only for $k_{max} = 12$ or larger. In other words, decreasing the order helps in maintaining an optimal degree of compression: smooth and slowly varying functions (Gaussian with $\alpha = 50$) are best represented with large degree polynomials which are able to yield an accurate representation with very few refinements. For high frequency variations (Gaussian with $\alpha = 10000$) and cusps, deep refinement levels are anyway necessary; the order reduction scheme employed here is able to keep the complexity close to optimal values by gradually removing unnecessary degrees of freedom.

We also notice that for the cusp and the narrow Gaussian, when $k_{max} = 12$ or larger, also the decreasing order scheme leads to slightly larger representations, albeit not as large as the standard scheme. We argue that a more pronounced order decrease (e.g. $k(n + 1) = k(n) - 2$) could help reduce the complexity in such cases but we have not pursued this route yet.

Another consideration regards the choice of n_0, namely the last scale with order $k = k_{max}$. We have often seen (*cf.* Table 5 on the Cusp-like example) that an optimal representation with the decreased-order approach is obtained when the order k_{min} is reached at the finest scale N. This requirement is however function-dependent and therefore difficult to exploit fully in practical applications, where the same $k(n)$ shall be employed for all functions. This consideration could nevertheless guide the final choice of the order function $k(n)$.

In the future we plan to apply the decreasing order scheme $k(n)$ to the application of operators in the Non-Standard form[5]. The main challenge in this case will be the construction of the components of the operator at each scale. However, as the Non-Standard form virtually decouples scales when the operator is applied (the coupling is afterwards restored by applying the filters to the resulting functions) we believe this to be a feasible prosecution of the present work.

Acknowledgments

This work has been supported by the Research Council of Norway through a Centre of Excellence Grant (Grant No. 179568/V30). This work has received support from the Norwegian Supercomputing Program (NOTUR) through a grant of computer time (Grant No. NN4654K).

[1] B. Alpert, Sparse representation of smooth linear operators, Ph.D. thesis, Yale University, Department of Mathematics, 10 Hillhouse Avenue, P.O. Box 208283 New Haven, CT 06520-8283, 1990. Available online at: `http://www.math.yale.edu/pub/papers/`.

[2] B. Alpert, G. Beylkin, D. Gines, L. Vozovoi, Adaptive solution of partial differential equations in multiwavelet bases, Journal of computational physics 182 (2002) 149–190.

[3] B.K. Alpert, A class of bases in L^2 for the sparse representation of integral operators, SIAM J. On Math. Analysis 24 (1993) 246–262.

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

[5] G. Beylkin, J. Keiser, On the adaptive numerical solution of nonlinear partial differential equations in wavelet bases, Journal of Computational Physics 132 (1997) 233–259.

[6] F.A. Bischoff, R.J. Harrison, E.F. Valeev, Computing many-body wave functions with guaranteed precision: The first-order Møller-Plesset wave function for the ground state of helium atom., The Journal of Chemical Physics 137 (2012) 104103.

[7] F.A. Bischoff, E.F. Valeev, Low-order tensor approximations for electronic wave functions: Hartree-Fock method with guaranteed precision., The Journal of Chemical Physics 134 (2011) 104104–104104–10.

[8] C. Canuto, M.Y. Hussaini, T.A. Zang, Spectral methods in fluid dynamics, Springer-Verlag, 1988.

[9] R. Car, M. Parrinello, Unified approach for molecular dynamics and density-functional theory, Phys. Rev. Lett. 55 (1985) 2471–2474. URL: `http://link.aps.org/doi/10.1103/PhysRevLett.55.2471`. doi:10.1103/PhysRevLett.55.2471.

[10] L. Frediani, E. Fossgaard, T. Flå, K. Ruud, Fully adaptive algorithms for multivariate integral equations using the non-standard form and multiwavelets with applications to the Poisson and bound-state Helmholtz kernels in three dimensions, Mol Phys 111 (2013) 1143–1160. URL: `http://www.tandfonline.com/doi/abs/10.1080/00268976.2013.810793`. doi:10.1080/00268976.2013.810793.

[11] S. Goedecker, Linear scaling electronic structure methods, Rev. Mod. Phys. 71 (1999) 1085–1123. URL: http://link.aps.org/doi/10.1103/RevModPhys.71.1085. doi:10.1103/RevModPhys.71.1085.

[12] R. Harrison, G. Fann, T. Yanai, G. Beylkin, Multiresolution quantum chemistry in multiwavelet bases, Multiresolution Quantum Chemistry in Multiwavelet Bases, Springer, Heidelberg, 2003, pp. 103–110.

[13] R. Harrison, G. Fann, T. Yanai, Z. Ghan, G. Beylkin, Multiresolution quantum chemistry: Basic theory and initial applications, Journal of Chemical Physics 121 (2004) 11587–11598.

[14] S.R. Jensen, J. Juselius, A. Durdek, P. Wind, T. Flå, L. Frediani, Linear scaling coulomb interaction in the multiwavelet basis, a parallel implementation, 2013. Submitted.

[15] F. Keinert, Wavelets and multiwavelets, Studies in Advanced Mathematics, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[16] N. Kovvali, Theory and Applications of Gaussian Quadrature Methods, Synthesis Lectures on Algorithms and Software in Engineering, Morgan Claypool Publishers, 2011. URL: http://dx.doi.org/10.2200/S00366ED1V01Y201105ASE008.

[17] K.N. Kudin, G.E. Scuseria, Linear-scaling density-functional theory with gaussian orbitals and periodic boundary conditions: Efficient evaluation of energy and forces via the fast multipole method, Phys. Rev. B 61 (2000) 16440–16453. URL: http://link.aps.org/doi/10.1103/PhysRevB.61.16440. doi:10.1103/PhysRevB.61.16440.

[18] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, J.D. Joannopoulos, Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients, Rev. Mod. Phys. 64 (1992) 1045–1097. URL: http://link.aps.org/doi/10.1103/RevModPhys.64.1045. doi:10.1103/RevModPhys.64.1045.

[19] G. Strang, T. Nguyen, Wavelets and filter banks, Wellesley-Cambridge Press, 1997. URL: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0961408871.

[20] U. Trottenberg, C. Oosterlee, A. Schueller, Multigrid, Academic Press, 2001.

[21] T. Yanai, G. Fann, Z. Ghan, R. Harrison, G. Beylkin, Multiresolution quantum chemistry in multiwavelet bases: Hartree-fock exchange, Journal of chemical physics 121 (2004) 6680–6688.