Tensor-structured methods in electronic structure calculations

Venera Khoromskaia

GAMM 2010, Karlsruhe, March 22-26, 2010.

Max-Planck Institute for Mathematics in the Sciences, Leipzig

Outline of the talk

1. Tensor-structured (TS) methods in $\mathbb{R}^d$
   - Basic tensor formats
   - Tensor product operations
   - Best Tucker approximation
   - Multigrid accelerated BTA
2. Accurate TS computation of integral operators in $\mathbb{R}^d$, $d \geq 3$
   - The Hartree-Fock equation
   - TS Computation of the Coulomb matrix
   - Computation of the exchange matrix
3. TS numerical solution of the Hartree-Fock equation
   - 3D nonlinear EVP solver by TS methods

Rank-structured representation of higher order tensors

A tensor of order d, $A = [a_{i_1 \ldots i_d} : i_\ell \in I_\ell] \in \mathbb{R}^{I_1 \times \ldots \times I_d}$,
$I_\ell = \{1, \ldots, n_\ell\}$, $\ell = 1, \ldots, d$,
for $n_\ell = n$, $N = n^d$.

tensor product of vectors $u^{(\ell)} = \{u_{i_\ell}^{(\ell)}\}_{i_\ell=1}^{n} \in \mathbb{R}^{I_\ell}$ forms the canonical rank-1 tensor

$$A_{(1)} \equiv [u_{\mathbf{i}}]_{\mathbf{i} \in \mathcal{I}} = u^{(1)} \otimes \ldots \otimes u^{(d)} \quad \text{with entries} \quad u_{\mathbf{i}} = u_{i_1}^{(1)} \cdots u_{i_d}^{(d)},$$

storage: $dn << n^d$.
the canonical format, $\mathcal{C}_{R,\mathbf{n}}$

$$A \approx A_{(R)} = \sum_{k=1}^{R} c_k \, u_k^{(1)} \otimes \ldots \otimes u_k^{(d)}, \quad c_k \in \mathbb{R}, \tag{1}$$

Tucker tensor format

The rank-$(r_1, \ldots, r_d)$ orthogonal Tucker approximation

[Tucker, De Lathauwer].

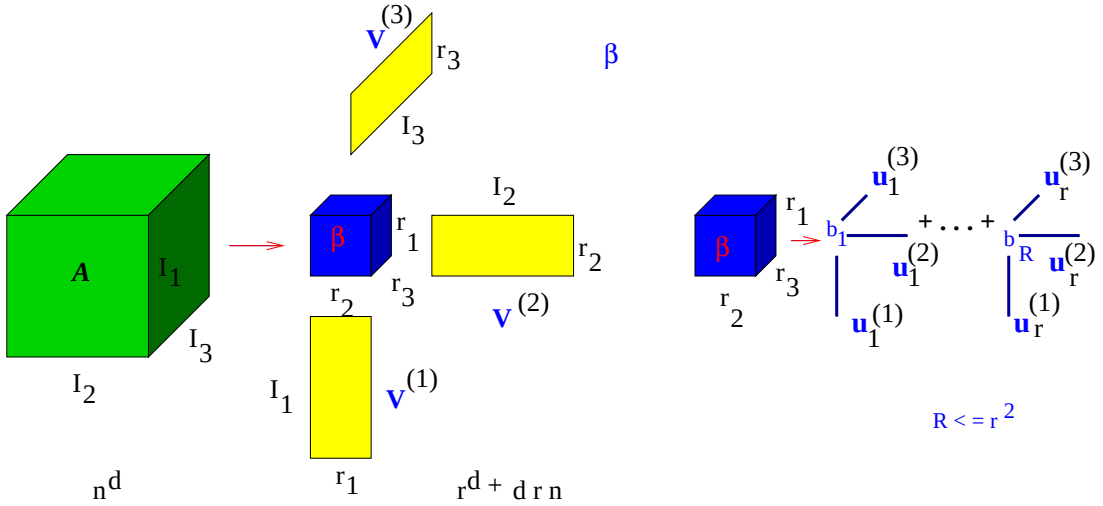
$$A \approx A_{(\mathbf{r})} = \sum_{\nu_1=1}^{r_1} \ldots \sum_{\nu_d=1}^{r_d} \beta_{\nu_1,\ldots,\nu_d} \, v_{\nu_1}^{(1)} \otimes \ldots \otimes v_{\nu_d}^{(d)}. \qquad (2)$$

$$A_{(\mathbf{r})} = \boldsymbol{\beta} \times_1 V^{(1)} \times_2 V^{(2)} \times_3 \ldots \times_d V^{(d)}.$$

Vectors in $V^{(\ell)} = [v_1^{(\ell)} \ldots v_{r_\ell}^{(\ell)}]$ form the orthonormal basis.
The core tensor $\boldsymbol{\beta} = [\beta_{\nu_1,\ldots,\nu_d}] \in \mathbb{R}^{r_1 \times \ldots \times r_d}$.
$r = \max\limits_{\ell}\{r_\ell\}$ is the (maximal) Tucker rank.
Storage: $r^d + drn \ll n^d$.

Mixed TC tensor format

$$A_{(\mathbf{r})} = \left(\sum_{k=1}^{R} b_k u_k^{(1)} \otimes \ldots \otimes u_k^{(d)} \right) \times_1 V^{(1)} \times_2 V^{(2)} \times_3 \ldots \times_d V^{(d)}.$$



Advantages of the Tucker tensor decomposition:

1. Robust algorithm for decomposing full format tensors of size n^d.
2. Rank reduction of the rank-R canonical tensors with large R.

3. Very efficient for 3D tensors since r^3 is small.

Multilinear operations in $\mathcal{C}_{R,\mathbf{n}}$ tensor format

$$A_1 = \sum_{k=1}^{R_1} c_k u_k^{(1)} \otimes \ldots \otimes u_k^{(d)}, \quad A_2 = \sum_{m=1}^{R_2} b_m v_m^{(1)} \otimes \ldots \otimes v_m^{(d)}.$$

Euclidean scalar product (complexity $O(dR_1R_2n) \ll n^d$),

$$\langle A_1, A_2 \rangle := \sum_{k=1}^{R_1} \sum_{m=1}^{R_2} c_k b_m \prod_{\ell=1}^{d} \left\langle u_k^{(\ell)}, v_m^{(\ell)} \right\rangle.$$

Hadamard product of A_1, A_2

$$A_1 \odot A_2 := \sum_{k=1}^{R_1} \sum_{m=1}^{R_2} c_k b_m \left(u_k^{(1)} \odot v_m^{(1)} \right) \otimes \ldots \otimes \left(u_k^{(d)} \odot v_m^{(d)} \right).$$

Convolution, for $d = 3$ ($O(R_1R_2n \log n) \ll n^3 \log n$ (3D FFT))

$$A_1 * A_2 = \sum_{k=1}^{R_1} \sum_{m=1}^{R_2} c_k b_m (u_m^{(1)} * v_k^{(1)}) \otimes (u_m^{(2)} * v_k^{(2)}) \otimes (u_m^{(3)} * v_k^{(3)})$$

Tucker tensor approximation

$$A_0 \in \mathbb{V}_{\mathbf{n}}: \quad f(A) := \|A - A_0\|^2 \to min \quad \text{over } A \in \mathcal{T}_{\mathbf{r}}. \tag{3}$$

The minimisation problem [De Lathauwer et al., 2000] (3) is equivalent to the maximisation problem

$$g(V^{(1)}, ..., V^{(d)}) := \left\| A_0 \times_1 V^{(1)\,T} \times ... \times_d V^{(d)\,T} \right\|^2 \to max \tag{4}$$

over the set of orthogonal matrices $V^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$.

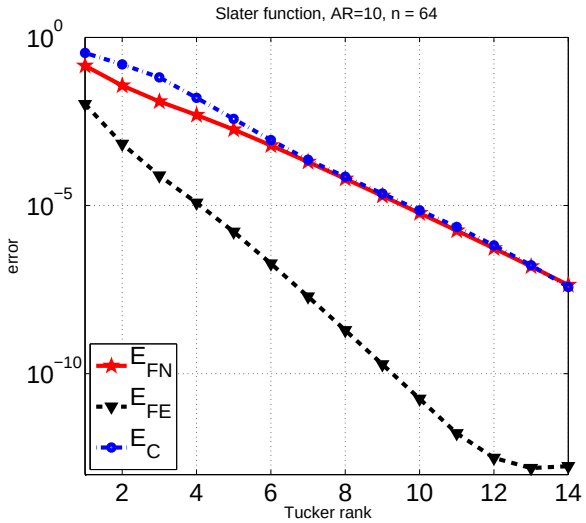
For given $V^{(\ell)}$, the core β that minimises (3) is

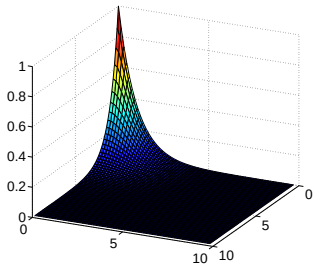
$$\beta = A_0 \times_1 V^{(1)\,T} \times ... \times_d V^{(d)\,T} \in \mathbb{R}^{r_1 \times ... \times r_d}. \tag{5}$$

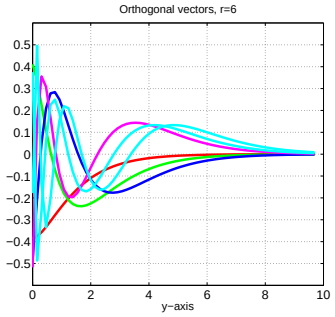
[De Lathauwer et al, 2000]: Algorithm BTA ($\mathbb{V}_{\mathbf{n}} \to \mathcal{T}_{\mathbf{r},\mathbf{n}}$).

Complexity for $d = 3$: $W_{F\to T} = O(n^4 + n^3 r + n^2 r^2 + n r^3)$.

BTA of function related tensors





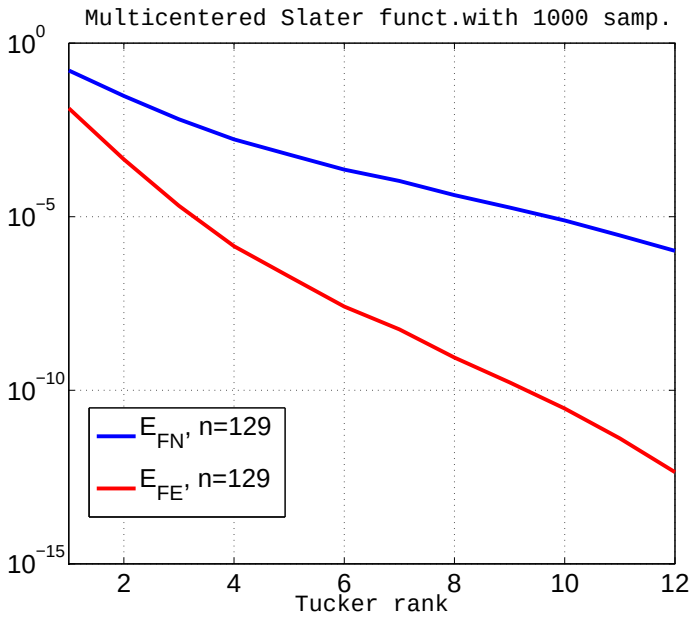


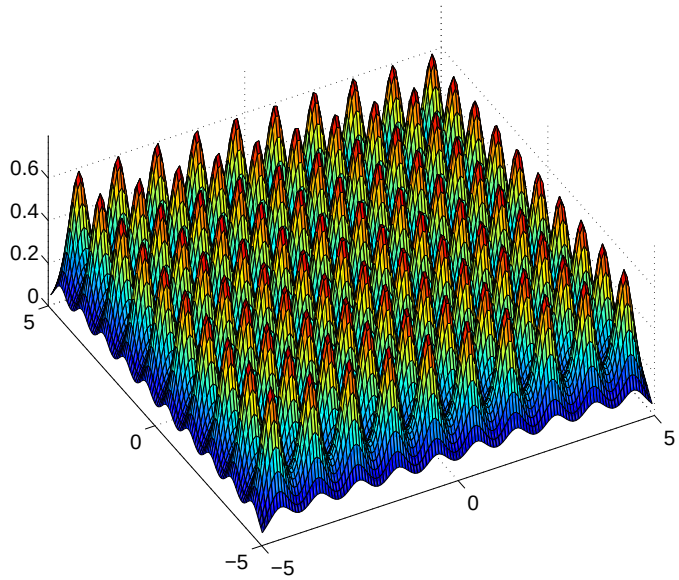
Slater function $g(x) = \exp(-\alpha\|x\|)$ with $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$

$$E_{FN} = \frac{\|A_0 - A_{(r)}\|}{||A_0||} \quad E_{FE} = \frac{\|A_0\| - \|A_{(r)}\|}{||A_0||}, \quad E_C := \frac{\max_{i\in\mathcal{I}} |a_{0,i} - a_{r,i}|}{\max_{i\in\mathcal{I}} |a_{0,i}|}.$$

[Khoromskij,VKH. '07]

MGA BTA for 3D periodic structures





The "multi-centered Slater potential" obtained by displacing a single Slater potential with respect to the $m \times m \times m$ spatial grid of size $H > 0$, (here $m = 10$)

$$g(x) = \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} e^{-\alpha \sqrt{(x_1 - iH)^2 + (x_2 - jH)^2 + (x_3 - kH)^2}}.$$

Canonical to Tucker Approximation

[B.N. Khoromskij, VKH '09]

Theorem

Let $A \in \mathcal{C}_{R,\mathbf{n}}$. Then the minimisation problem

$$A \in \mathcal{C}_{R,\mathbf{n}} \subset \mathbb{V}_{\mathbf{n}} : \quad A_{(\mathbf{r})} = \operatorname{argmin}_{V \in \mathcal{T}_{\mathbf{r},\mathbf{n}}} \|A - V\|_{\mathbb{V}_{\mathbf{r}}}, \tag{6}$$

$$[Z^{(1)}, \ldots, Z^{(d)}] = \operatorname{argmax}_{V^{(\ell)} \in \mathcal{M}_\ell} \left\| \sum_{\nu=1}^{R} c_\nu \left({V^{(1)}}^T u_\nu^{(1)} \right) \otimes \ldots \otimes \left({V^{(d)}}^T u_\nu^{(d)} \right) \right\|_{\mathbb{V}_{\mathbf{r}}}^2,$$

Init. guess: RHOSVD $Z_0^{(\ell)} \Rightarrow$ the truncated SVD of $U^{(\ell)} = [u_1^{(\ell)} \ldots u_R^{(\ell)}]$.
Error bounds for RHOSVD

$$\|A - A^0_{(\mathbf{r})}\| \le \|\mathbf{c}\| \sum_{\ell=1}^{d} \Big(\sum_{k=r_\ell+1}^{\min(n,R)} \sigma_{\ell,k}^2 \Big)^{1/2}, \qquad \text{where } \|\mathbf{c}\|^2 = \sum_{\nu=1}^{R} c_\nu^2.$$

Canonical to Tucker Approximation

The maximizer $Z^{(\ell)} = [z_1^{(\ell)} ... z_{r_\ell}^{(\ell)}] \in \mathbb{R}^{n \times r_\ell}$ is obtained by ALS of the matrix unfolding

$$B_{(q)} \in \mathbb{R}^{n \times \bar{r}_q}, \quad \bar{r}_q = r_1 ... r_{q-1} r_{q+1} ... r_d. \tag{7}$$

(c) The minimiser in (6) is then

$$A_{(\mathbf{r})} = \beta \times_1 Z^{(1)} \times_2 Z^{(2)} \times_3 \ldots \times_d Z^{(d)}.$$

$$\beta = \sum_{\nu=1}^{R} c_\nu ({Z^{(1)}}^T u_\nu^{(1)}) \otimes ... \otimes ({Z^{(d)}}^T u_\nu^{(d)}) \in \mathcal{C}_{R,\mathbf{r}}.$$

Complexity for $d = 3$, $n \geq R$: $W_{C \to T} = O(nR^2 + nr^4)$.

C2T $\Rightarrow$ T2C: $R \to R' \leq r^2 \ll R$

MGA Tucker approximation for $\mathcal{C}_{R,\mathbf{n}}$ tensors

1. Sequence of nonlinear appr. problems for $A = A_n$,$n = n_m := n_0 2^m$, $m = 0, 1, \ldots, M$, on a sequence of refined grids $\omega_{\mathbf{3},n_m}$.
2. $Z_0^{(q)}$ on grid $\omega_{\mathbf{3},n_m} \rightarrow$ by linear interp. of $Z^{(q)} \in \mathbb{R}^{n_{m-1} \times r_q}$ from $\omega_{\mathbf{3},n_{m-1}}$.
3. The *restricted ALS iteration*, is based on "most important fibers" (MIFs) of unfolding matrices.
   Positions of MIFs are extracted at the coarsest grid

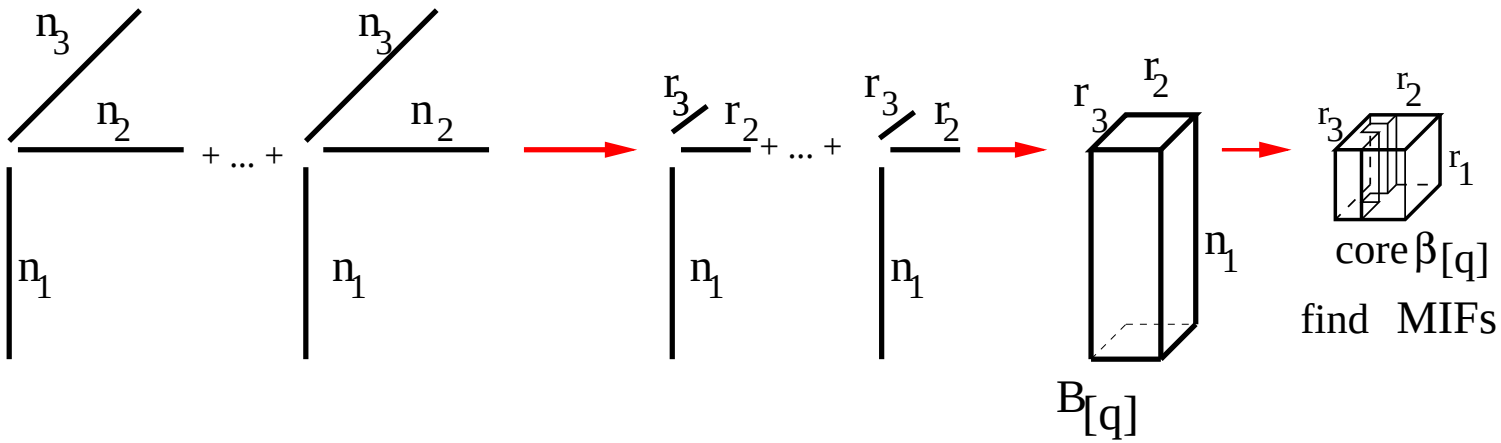
$$\beta_{(q)} = {Z^{(q)}}^T B_{(q)} \in \mathbb{R}^{r_q \times \overline{r}_q}. \tag{8}$$

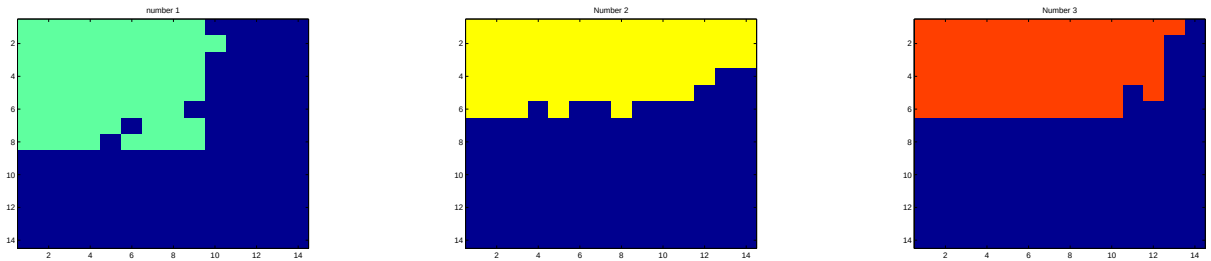
   Location of MIFs corresponds to pr columns in $\beta_{(q)}$ with maximal Euclidean norms (maximal energy principle).
4. Complexity for $d = 3$: $O(Rrn_M + p^2 r^2 n_M)$.
   (Linear w.r.t. n and R !)

MGA C2T ($\mathcal{C}_{R,\mathbf{n}}$)

[B.N. Khoromskij, VKH '08]

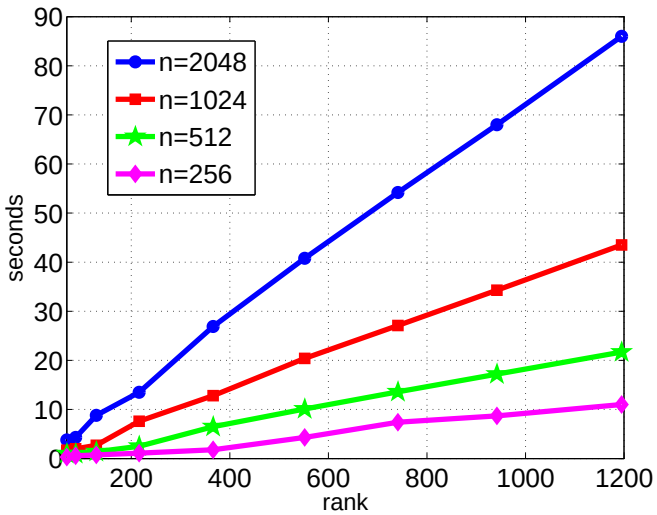
$B_{(q)} \to$ unfolding of $(B_{[q]})$, $\quad \beta_{(q)} = Z^{(q)\,T} B_{(q)}$; $n_0 = 64$, $n_M = 16384$

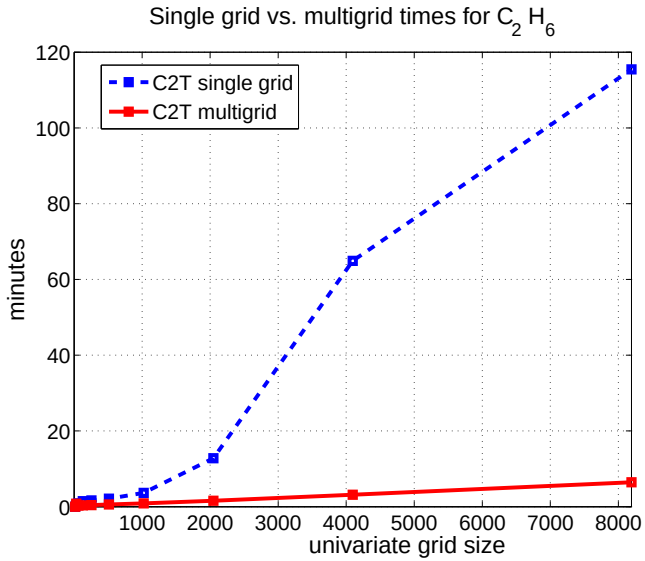




MIFs for H_2O

MGA Tucker vs. single grid





Linear scaling in R and in n (left) for C2T algorithm.

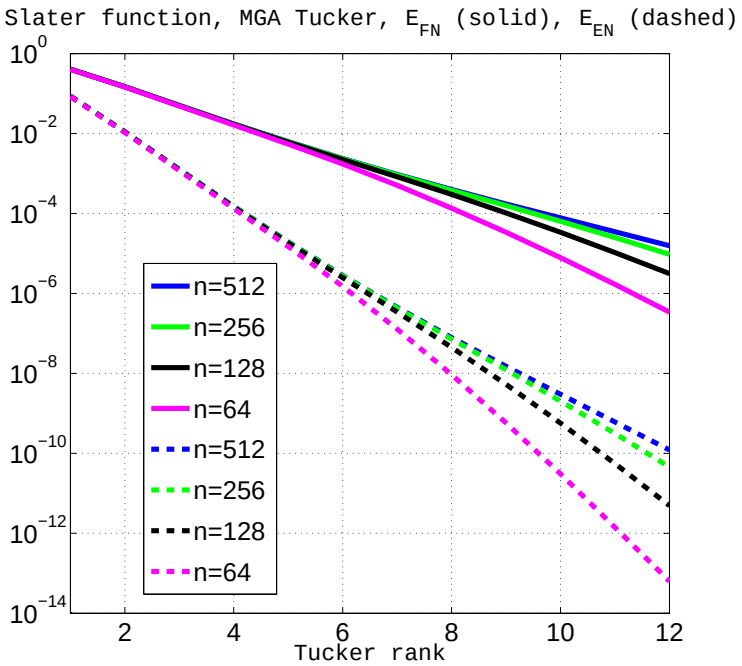
Multigrid vs. single grid times for the C2T transform in computations of V_H for C_2H_6 (right).

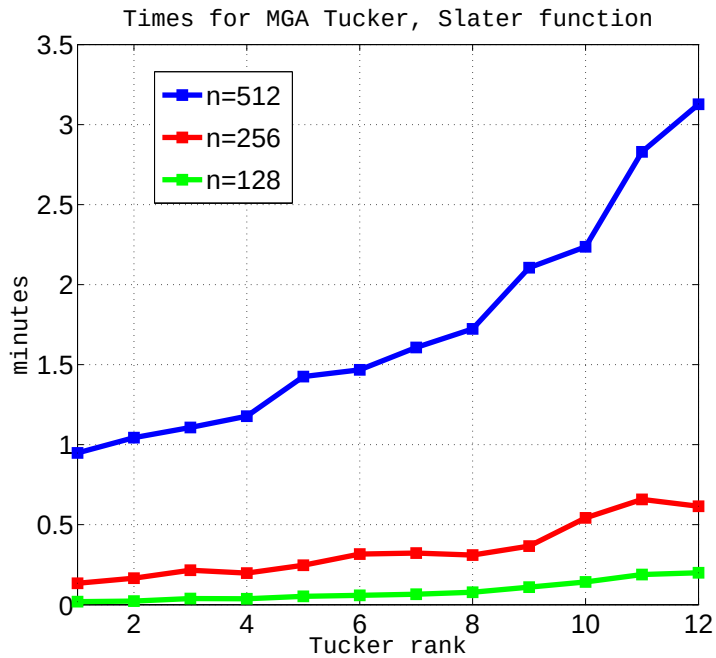
MGA BTA for full format tensors

MGA best Tucker decomposition for full format tensors,

[B. Khoromskij,VKH'09], [VKH '10, in progress]

$$W_{F\to T} = O(n^3) \text{ for MGA BTA, instead of } O(n^4) \text{ for BTA.}$$

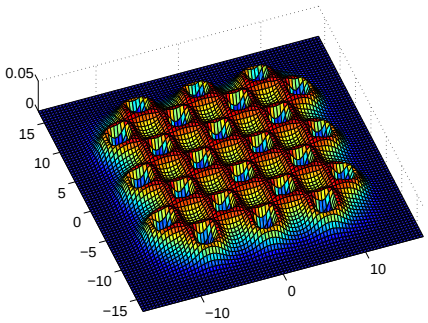


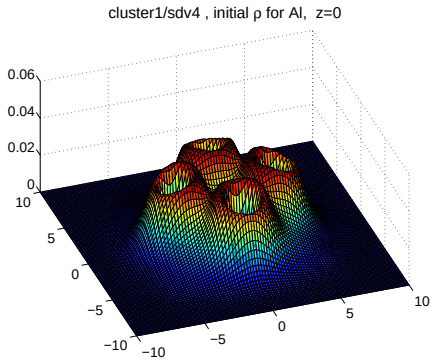


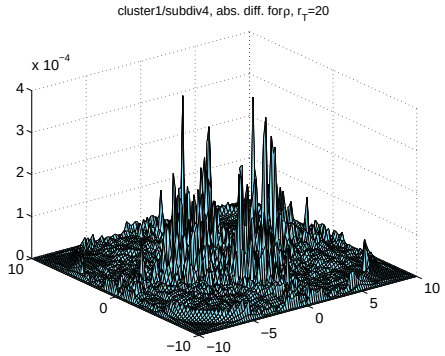
Full size tensor with number of entries 512^3, times for Matlab.

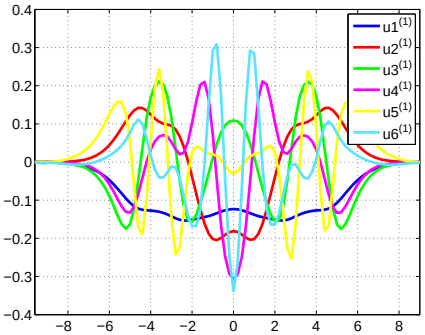
MGA BTA for full format tensors (Aluminium Cluster)

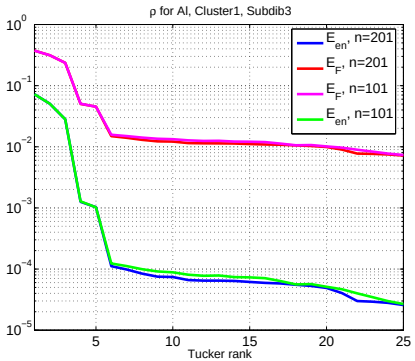
[T. Blesgen, V. Gavini and VKH, '09]











FE (coarse graining) $\Rightarrow$ computed on large supercomputer clusters*.

FE data* $\Rightarrow$ 3D Cartesian grid $\Rightarrow$ Tucker decomposition

* [V. Gavini, J. Knap, K. Bhattacharya, and M. Ortiz,'07]

The Hartree-Fock equation

The Hartree-Fock equation

$$\mathcal{F}\varphi_i(x) = \lambda_i \varphi_i(x), \quad i = 1, ..., N_{orb}, \quad x \in \mathbb{R}^3$$

$$\mathcal{F} := -\frac{1}{2}\Delta - V_c + V_H - \mathcal{K},$$

the Hartree potential

$$V_H(x) := \int_{\mathbb{R}^3} \frac{\rho(y)}{|x-y|}\, dy, \quad \rho(y) = 2\sum_{i=1}^{N_{orb}} (\varphi_i(y))^2$$

where $\varphi_i = \sum_{k=1}^{R_0} c_{i,k} g_k(x)$.

The exchange operator

$$(\mathcal{K}\varphi)(x) := \int_{\mathbb{R}^3} \frac{\tau(x,y)}{|x-y|}\, \varphi(y) dy, \quad \tau(x,y) = \sum_{i=1}^{N_{orb}} \varphi_i(x)\varphi_i(y).$$

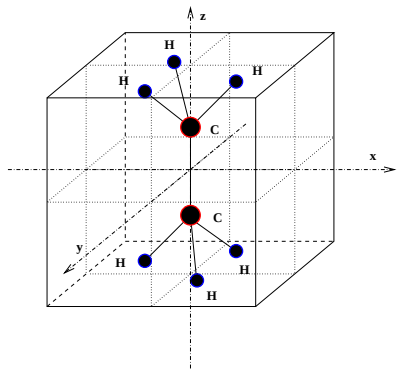
$O(n)$ computation of the Hartree potential on huge 3D grids

V_H is computed on large $n \times n \times n$ Cartesian grids, with $n \sim 10^3 \div 10^4$

$$\rho(x) = 2\sum_{i=1}^{N_{orb}} (\varphi_i)^2, \quad \varphi_i = \sum_{k=1}^{R_0} c_{i,k} g_k(x), \quad x = (x_1, x_2, x_3) \in \mathbb{R}^3$$

$$g_k(x) = g_{k,1}(x_1)\, g_{k,2}(x_2)\, g_{k,3}(x_3),$$
$$g_{k,\ell} = (x_\ell - A_{k,\ell})^{\beta_{k,\ell}} e^{-\lambda_k (x_\ell - A_{k,\ell})^2},$$

$$g_{k,\ell} \to u_k^{(\ell)}; x_\ell \in [-b, b],$$
$$b = 6 \div 10 \ \mathring{A}.$$



$$R_0(CH_4) = 55, \quad R_0(C_2H_6) = 96, \quad R_0(H_2O) = 41, \quad R_\rho = \frac{R_0(R_0+1)}{2}.$$

$$\rho \approx F_\rho = \sum_{k=1}^{R_\rho} c_k\, u_k^{(1)} \otimes u_k^{(2)} \otimes u_k^{(3)}, \quad c_k \in \mathbb{R}, \quad u_k^{(\ell)} \in \mathbb{R}^n.$$

$\Rightarrow \rho$ is presented in the computation box $[-b, b]^3$ with $R_\rho \sim 10^3 \div 10^4$.

$O(n)$ computation of the Hartree potential on huge 3D grids

$R_\rho(CH_4) = 1540, \quad R_\rho(C_2H_6) = 4656, \quad R_\rho(H_2O) = 861$

Next : MGA C2T + T2C transformations to reduce the rank:

$$R_{\rho_r} \sim 10^2.$$

$R_N \sim 20 \div 30$ canonical tensor G for the Newton kernel $\frac{1}{\|x\|}$ (using *sinc*-quadratures) [C. Bertoglio, B. Khoromskij '08]).

The rank-R_{ρ_r} canonical tensor F for ρ:

$$V_H = F * G = \sum_{m=1}^{R_{\rho r}} \sum_{k=1}^{R_N} c_k b_m \left(u_1^k * v_1^m\right) \otimes \left(u_2^k * v_2^m\right) \otimes \left(u_3^k * v_3^m\right)$$

Coulomb matrix: tensor inner products in GTO basis $\{g_k\}_{k=1}^{R_0}$

$$J_{km} := \int_{\mathbb{R}^3} g_k(x) g_m(x) V_H(x) dx, \quad k, m = 1, \ldots R_0.$$

Tensor-product convolution vs. 3D FFT

The cost of computation of V_H by tensor product convolution (1D FFT):

$$\mathcal{N}_{C*C} = O(R_{\rho_r} R_N n \log n)$$

instead of $O(n^3 \log n)$ for 3D FFT.

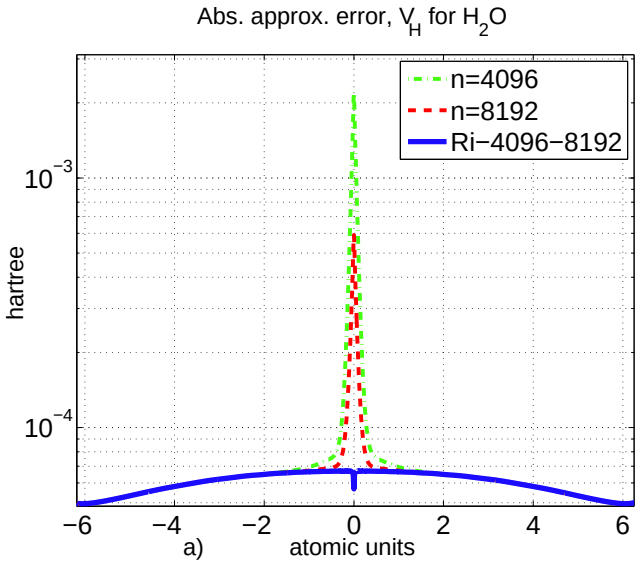
n^3	128^3	256^3	512^3	1024^3	2048^3	4096^3	8192^3	16384^3
FFT_3	4.3	55.4	582.8	~ 6000	–	–	–	~ 2 years
$C * C$	0.2	0.9	1.5	8.8	20.0	61.0	157.5	299.2
C2T	4.2	4.7	5.6	6.9	10.9	20.0	37.9	86.0

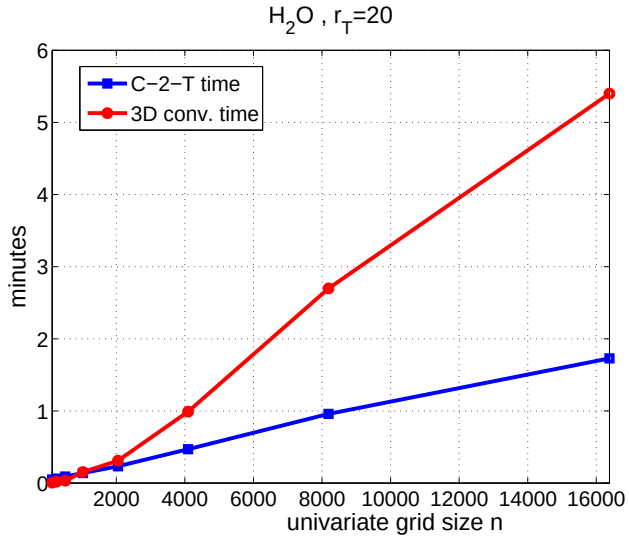
CPU time (in sec) for the computation of V_H for H_2O.

(3D FFT time for $n \geq 1024$ is obtained by extrapolation).

High accuracy computations of V_H

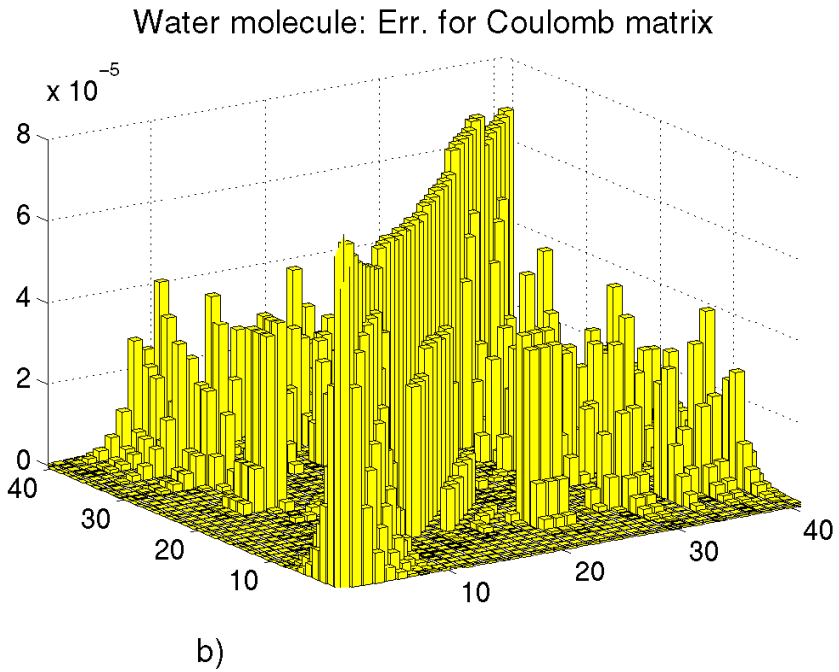
(Calculation over the number of entries $N \approx 4.3 \cdot 10^{12}$).

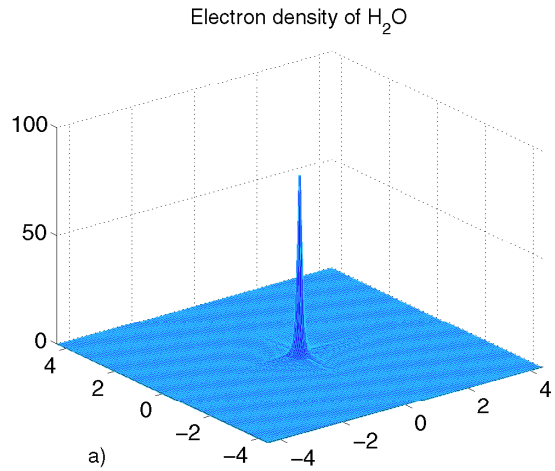




High accuracy computations of the Coulomb matrix for H_2O

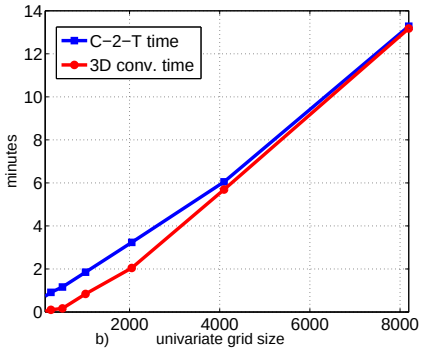
(for calculation over the number of entries $N \approx 4.3 \cdot 10^{12}$). Approx. error $O(h^3)$, $h = 2b/n$ ($h \approx 0.0008$ $\mathring{A}$).

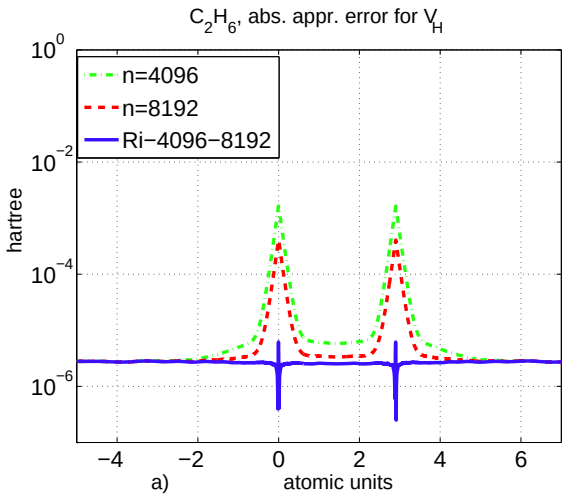


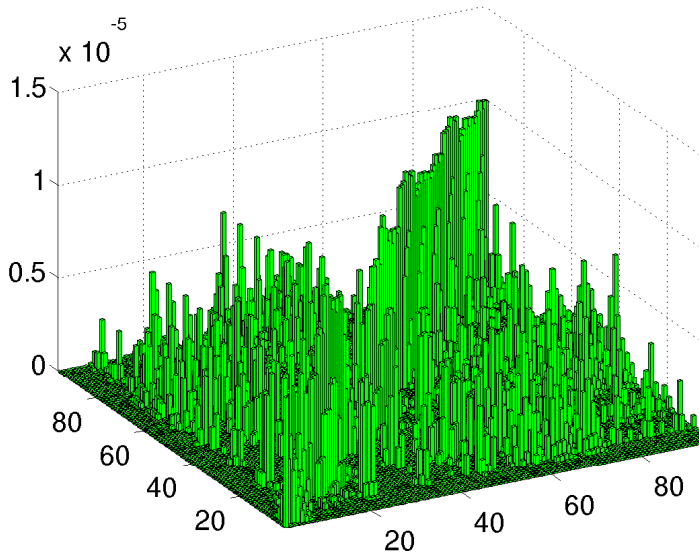


High accuracy computations of the Coulomb matrix for C_2H_6

a) CPU times for V_H : $O(n \log n)$ on $n \times n \times n$ grids with max.: $n = 8192$.
b) Absolute approximation error of the tensor-product computation of the Coulomb matrix for C_2H_6 molecule.







TS computation of the exchange matrix

[VKH '09]

$$K_{k,m} := \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} g_k(x)\frac{\tau(x,y)}{|x-y|}g_m(y)dxdy, \quad k,m = 1,\dots R_0$$

1. Compute the discrete tensor product convolution

$$W_{a,m}(x) = \int_{\mathbb{R}^3} \frac{\varphi_a(y)g_m(y)}{|x-y|}dy, \quad \text{cost: } O(R_N R_0 n \log n) \text{ per element } (\times R_0 N)$$

2. Entries of the Exchange matrix for every orbital,

$$V_{km,a} = \int_{\mathbb{R}^3} \varphi_a(x)g_k(x)W_{a,m}(x)dx$$

with the cost $O(R_N R_0^4 n_f)$.

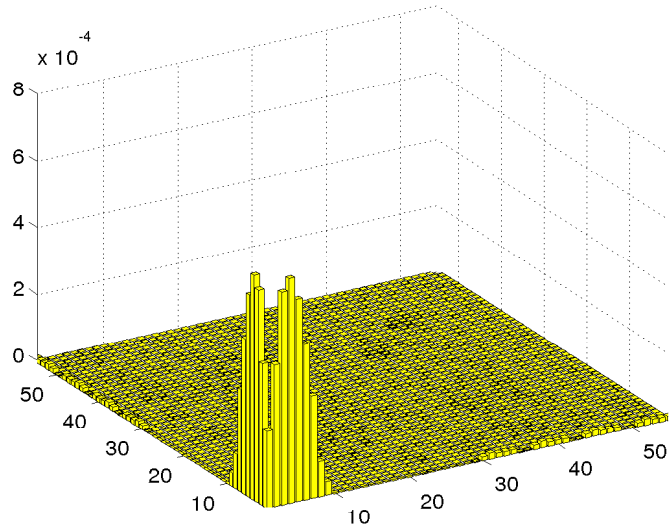
3. Entries of the Exchange matrix

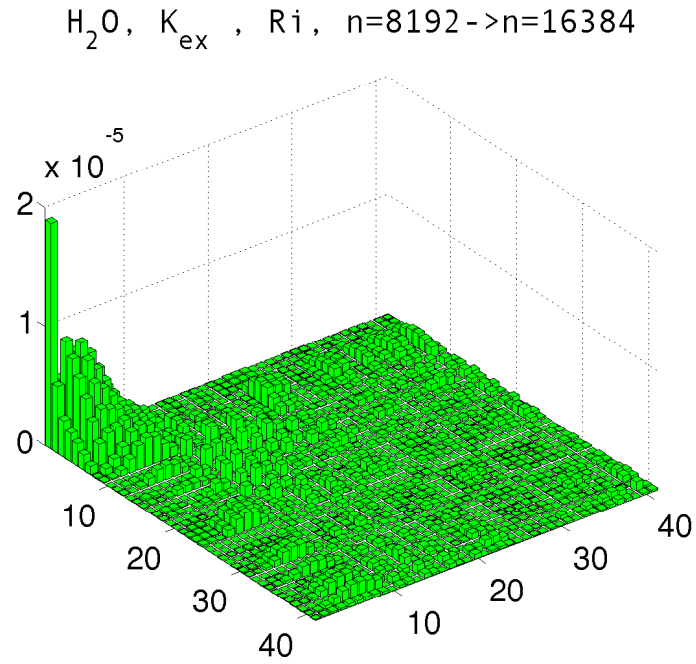
$$K_{k,m} = \sum_{a=1}^{N_{orb}} V_{km,a}$$

TS computation of the exchange matrix on huge 3D grids

Abs. error for the Exchange matrices K_{ex} for CH_4 (Richardson extrapolation for $n = 4096$)

and H_2O (Rich. for $n = 16384$).





Solution of the Hartree-Fock on a sequence of 3D grids

[Flad, Khoromskij, VKH '09]

Discretized GTO basis, $\varphi_i(x) \approx \sum_{k=1}^{R_0} c_{ki} \overline{g}_k(x)$, $x \in \mathbb{R}^3$, $i = 1, ..., N_{orb}$

$$\mathcal{F}\varphi_i(x) = \lambda_i \varphi_i(x), \quad \Rightarrow \quad \mathbf{F}U = \lambda \mathbf{S}U$$

$$\mathbf{F} = H_0 + J(C) - K(C), \quad C = \{c_{ki}\} = [U_1 U_2 \dots U_{N_{orb}}] \in \mathbb{R}^{R_0 \times N_{orb}}$$

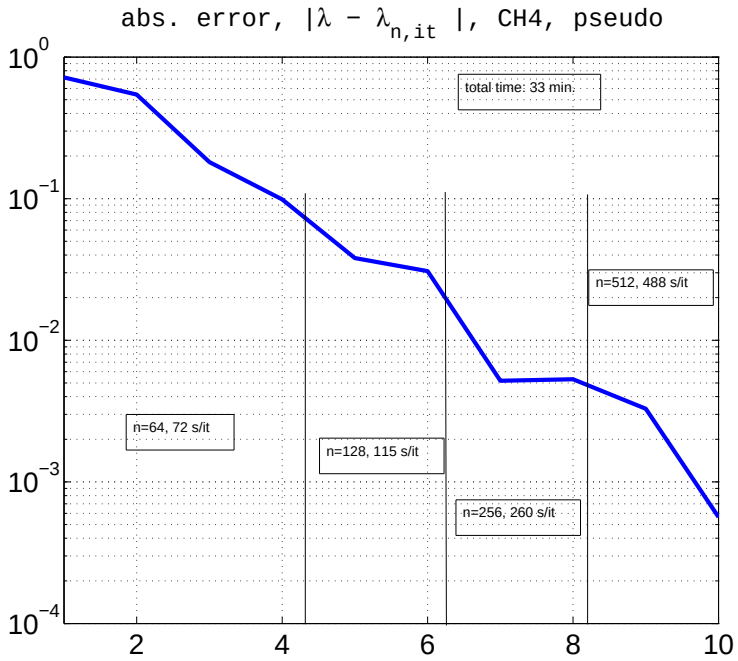
H_0 is the stiffness matrix of the core Hamiltonian $-\frac{1}{2}\Delta + V_c$

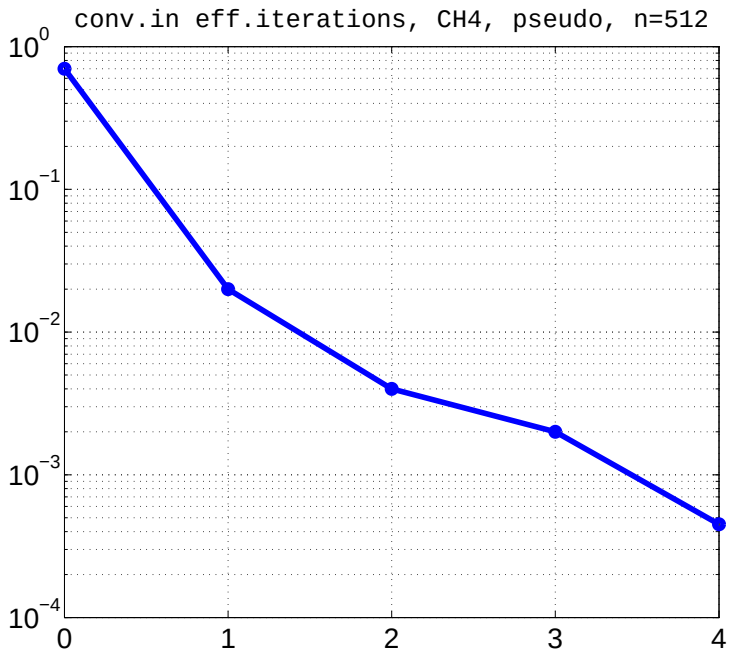
Our iterative scheme

- Initial guess for $m = 0$: $J = 0$, $K = 0$ $\Rightarrow$ $\mathbf{F}_0 = H_0$.
- On a sequence of refined grids $n = n_0, 2n_0, \dots, 2^p n_0$, solve EVP $[H_0 + J_m(C) - K_m(C)]U = \lambda S U$.
- Fast update of $J_m(C)$ and $K_m(C)$.
- Grid dependent termination criteria $\varepsilon_p = \varepsilon_0 \cdot 4^{-p}$.
- Use DIIS [Pulay '80] to provide fast convergence.

HF for the pseudopotential case of CH_4

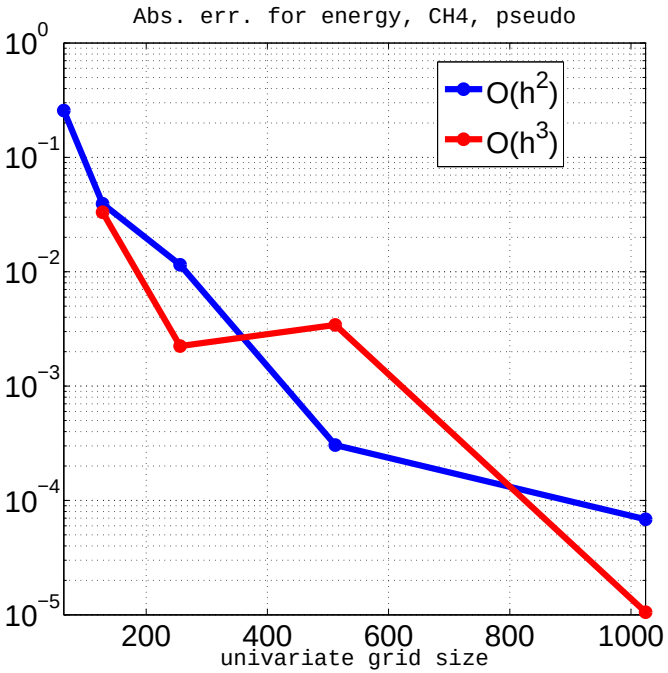
Left: multilevel convergence of iterations for the pseudopotential case of CH_4.
Right: convergence in effective iterations for finest grid.

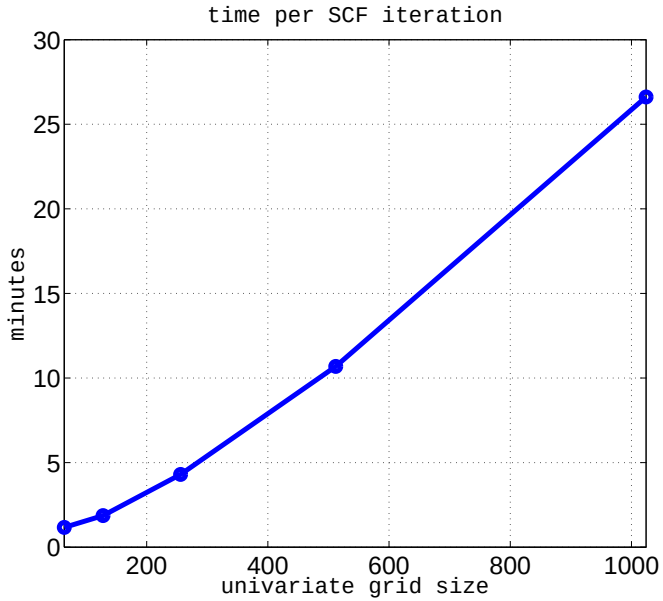




Convergence in HF energy

Left: convergence in HF energy.
Right: linear scaling of the iteration time in the univariate grid size.





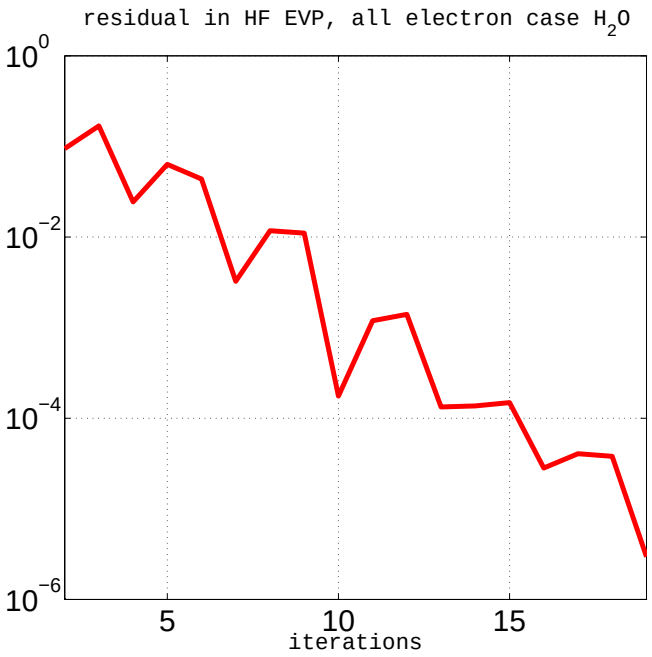
The total energy is calculated by

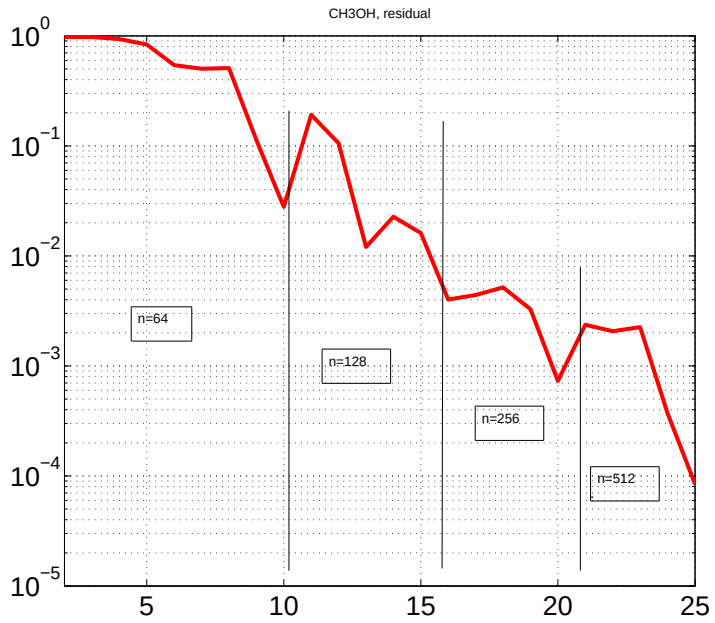
$$E_{HF} = 2\sum_{i=1}^{N_{orb}} \lambda_i - \sum_{i=1}^{N_{orb}} \left(\widetilde{J}_i - \widetilde{K}_i\right)$$

with $\widetilde{J}_i = (\varphi_i, V_H\varphi_i)_{L^2}$ and $\widetilde{K}_i = (\varphi_i, \mathcal{V}_{ex}\varphi_i)_{L^2}$ $(i = 1, \ldots, N_{orb})$ the Coulomb and exchange integrals, computed with respect to the orbitals φ_i.

Iterations of the nonlinear 3D EVP solver

Left: all electron case of H_2O.
Right: the pseudopotential case of CH_3OH .





Thank you for attention

The talk is based on

1. H.-J. Flad, V. Khoromskaia, B. Khoromskij.
   Solution of the Hartree-Fock Equation by the Tensor-Structured Methods.
   Preprint MPI MIS 44/2009.
2. V. Khoromskaia.
   Computation of the Hartree-Fock Exchange by the Tensor-structured Methods.
   Preprint MPI MIS 25/2009
3. T. Blesgen, V. Gavini and V. Khoromskaia.
   Approximation of the Electron Density of Aluminium Clusters in Tensor-product Format,
   Preprint MPI MIS 66/2009
4. V. Khoromskaia and B. N. Khoromskij.
   Multigrid Accelerated Tensor Approximation of Function Related Multi-dimensional Arrays.
   SIAM Journal on Scientific Computing, vol. 31, No. 4, pp. 3002-3026, 2009.
5. B. N. Khoromskij, V. Khoromskaia, S.R. Chinnamsetty and H.-J. Flad.
   Tensor Decomposition in Electronic Structure Calculations on 3D Cartesian Grids.
   Journal of Computational Physics, 228(2009), pp. 5749-5762, 2009.
6. B. N. Khoromskij and V. Khoromskaia.
   Low Rank Tucker-Type Tensor Approximation to Classical Potentials.
   Central European Journal of Mathematics v.5, N.3, 2007, pp.523-550.

http://personal-homepages.mis.mpg.de/vekh/