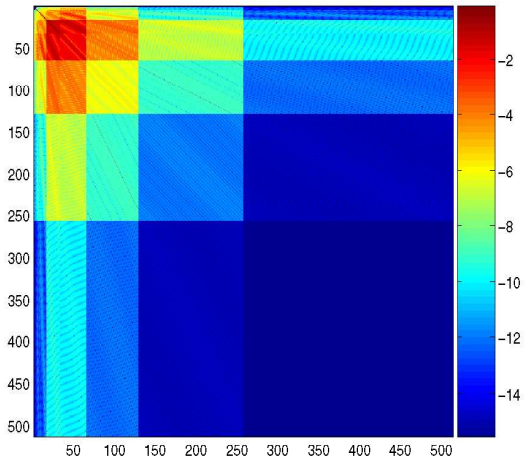
Operator calculus of density matrices and sparse wavelet representation

Reinhold Schneider

Christian Albrechts University Kiel
joint work and project with H.J. Flad,
and W. Hackbusch, MIS MPG Leipzig



Efficiency

Minimize number of basis functions +
minimize computational complexity

$\Leftrightarrow$

Maximize the accuracy

in electronic structure calculation

Analysis of convergence rates,
development of error estimators,
employment of adaptivity and multigrid methods etc.

Best N-Term Approximation

Consider a Riesz (quasi-orthogonal) basis $\Psi = (\psi_\lambda)_{\lambda \in I}$ in some Hilbert space H.
For $v \in \mathrm{H}$, we consider the **error** of the **best n-term approximation**

$$\sigma_{n,\mathrm{H},\Psi}(v) = \min\{\|v - \sum_{\lambda \in \mathcal{T}} w_\lambda \psi_\lambda\| : w_\lambda \in \mathbb{C}\ ,\ \mathcal{T} \subset I\ ,\ \sharp\mathcal{T} = n\}.$$

Best N-Term Approximation

Consider a Riesz (quasi-orthogonal) basis $\Psi = (\psi_\lambda)_{\lambda \in I}$ in some Hilbert space H.
For $v \in \mathrm{H}$, we consider the **error** of the **best n-term approximation**

$$\sigma_{n,\mathrm{H},\Psi}(v) = \min\{\|v - \sum_{\lambda \in \mathcal{T}} w_\lambda \psi_\lambda\| : w_\lambda \in \mathbb{C}\ ,\ \mathcal{T} \subset I\ ,\ \sharp\mathcal{T} = n\}.$$

$$\mathrm{A}^s(\mathrm{H},\Psi) := \{v \in \mathrm{H} : \sigma_{n,\mathrm{H},\Psi}(v) \lesssim n^{-s}\} \qquad \text{with} \qquad \|v\|_{\mathrm{A}^s} := \sup_{n \in \mathbb{N}} n^s \sigma_{n,\mathrm{H},\Psi}(v) + \|v\|$$

is a (quasi-)normed space.

Best N-Term Approximation

Consider a Riesz (quasi-orthogonal) basis $\Psi = (\psi_\lambda)_{\lambda \in I}$ in some Hilbert space H.
For $v \in \mathrm{H}$, we consider the **error** of the **best n-term approximation**

$$\sigma_{n,\mathrm{H},\Psi}(v) = \min\{\|v - \sum_{\lambda \in \mathcal{T}} w_\lambda \psi_\lambda\| : w_\lambda \in \mathbb{C}\ ,\ \mathcal{T} \subset I\ ,\ \sharp\mathcal{T} = n\}.$$

$$\mathrm{A}^s(\mathrm{H},\Psi) := \{v \in \mathrm{H} : \sigma_{n,\mathrm{H},\Psi}(v) \lesssim n^{-s}\} \qquad \text{with} \qquad \|v\|_{\mathrm{A}^s} := \sup_{n \in \mathbb{N}} n^s \sigma_{n,\mathrm{H},\Psi}(v) + \|v\|$$

is a (quasi-)normed space.

Approximation strategy: Keep the n largest coefficients w_λ in $v = \sum_{\lambda \in I} w_\lambda \psi_\lambda$.

Hilbert space: $\mathrm{H} = L^2$ or $\mathrm{H} = H^1$, space of functions in L^2 with bounded kinetic energy $\langle Tv, v\rangle < \infty$.

Basis functions

Candidates:

- Canonical orbitals (=eigen functions of the Fock operator), not known a priori
- Local orbitals (Wannier functions), not known a priori
- atomic type orbitals: Slater type orbitals, Gaussian type orbitals
- ”systematic” basis functions: plane waves, finite elements, multiscale basis functions

Model problem

- effective one particle model (Hartree-Fock, density functional theory, hybrid models):

$$\mathcal{H}_\rho = -\frac{1}{2}\Delta + \mathcal{V}_{core} + \mathcal{V}_{pseudo} + \mathcal{V}_H(n) + \alpha\mathcal{V}_{XC}(n) - \frac{1}{2}\beta\mathcal{W}(\rho),$$

where

$$\rho(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{N/2} \varphi_i(\mathbf{x})\overline{\varphi_i(\mathbf{y})} \qquad (\textbf{density matrix}), \qquad n(\mathbf{x}) = 2\rho(\mathbf{x},\mathbf{x}) \qquad (\textbf{electron density})$$

- aim: linear scaling methods
  - Linear scaling methods using Gaussian type basis functions are limited to ”nondiffusive” basis functions. Furthermore, scaling with respect to the number of basis functions is not linear!
  - Competing with discretizations like plane waves or Finite Differences, which do not (easily) admit linear scaling approaches.

Multiscale Analysis

Practical and/or analytical tool

Multiscale hierarchy: $V_j = \text{span}\{\phi_{j,k} : k \in \Delta_j\}$

$$(V_{-l} \subset V_{-l+1} \subset \ldots) V_0 \subset V_1 \subset \ldots \subset V_j \subset V_{j+1} \subset \ldots \subset L^2$$

Multiscale Analysis

Practical and/or analytical tool

Multiscale hierarchy: $V_j = \operatorname{span}\{\phi_{j,k} : k \in \Delta_j\}$

$$(V_{-l} \subset V_{-l+1} \subset \ldots)V_0 \subset V_1 \subset \ldots \subset V_j \subset V_{j+1} \subset \ldots \subset L^2$$

Multiscale decomposition and wavelets

- Decomposition: $V_{j+1} = V_j \oplus W_j, \qquad V_J = \bigoplus_{j=-1}^{J-1} W_j, \qquad W_{-1} := V_0$
- Wavelets: $W_j = \operatorname{span}\{\psi_{j,k} : k \in \nabla_j := \Delta_{j+1} \setminus \Delta_j\} = \operatorname{span} \Psi_j$
- local supports: $\operatorname{diam} \operatorname{supp} \psi_{j,k} \sim 2^{-j} \;\; , \;\; \operatorname{diam} \operatorname{supp} \phi_{j,k} \sim 2^{-j}$
- Orthonormality: $\langle \psi_{j,k}, \psi_{j',k'} \rangle = \delta_{(j,k),(j',k')}$
- *d* vanishing moments: $|\langle \psi_{j,k}, x^\alpha \rangle| = 0 \;\; , \;\; |\alpha| < d.$

Theorem (Norm equivalences for Sobolev spaces):

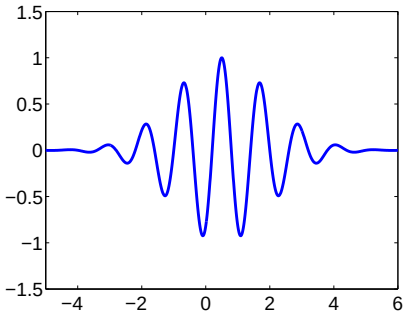
For all $0 \leq s < \sup\{s \in \mathbb{R} : \phi \in H^s(\mathbb{R})\}$ there holds ($\lambda = (j,k)$, $|\lambda| = j$)

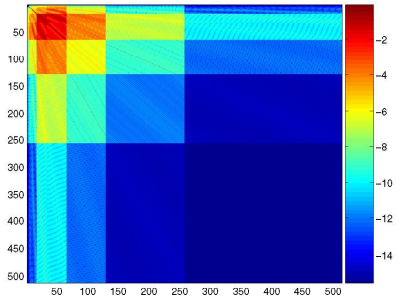
$$\|u\|_s^2 \sim \sum_{\lambda} |\langle u, \psi_\lambda \rangle|^2 2^{2|\lambda| s}$$

Sparse grid - hyperbolic cross

$$\psi_{\mathbf{j},\mathbf{k}}(\mathbf{x}) = \psi_{j_1,k_1}(x_1) \cdot \ldots \cdot \psi_{j_N,k_N}(x_N) \ , \ \Psi = \Psi \otimes \ldots \otimes \Psi = \bigotimes_{i+1}^{N} \Psi \ ,$$

$$\|\mathbf{j}\|_1 = \sum_{i=1}^{N} j_i \leq L \ , \ \sharp\Psi \sim L^{N-1} 2^L \sim n \log^{N-1} \text{ instead of } n^N \ !$$



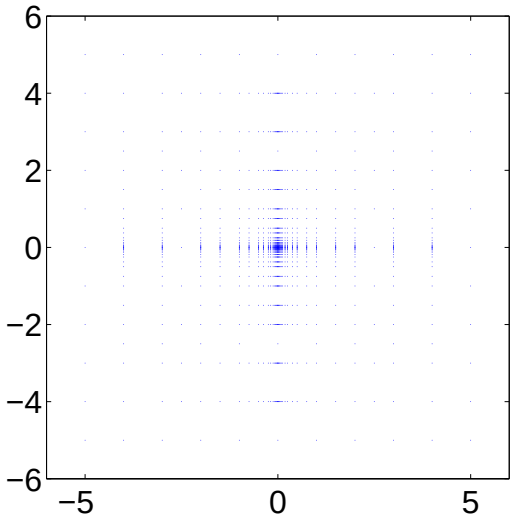


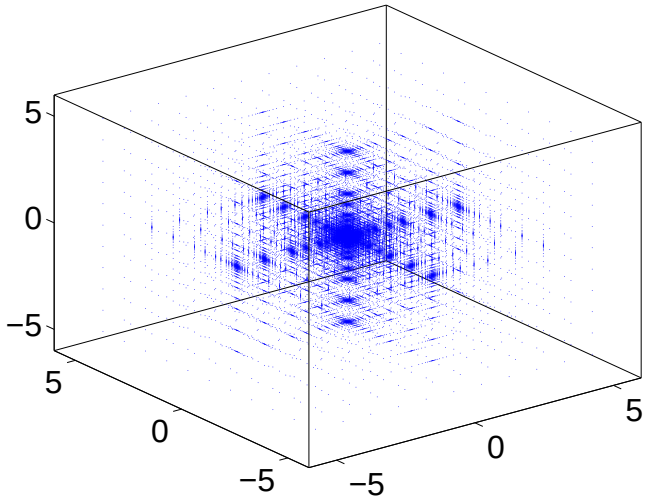
Theorem (Flad, Hackbusch, Schneider '05) :

- For every $0 < s < d - 1$ (d vanishing moments) the singularity Φ of the electron-nuclear cusp and therefore the canonical orbitals (Lieb-Simon) φ_i belong to $\mathrm{A}^s(H^1(\mathbb{R}^3), \bigotimes_{i=1}^{3} \Psi)$, (instead of $\mathrm{A}^{s/3}$, but GTO bases are exponentially convergent).

- For every $0 < s < d - 1$ the density matrix ρ belongs to $\mathrm{A}^s(H^1(\mathbb{R}^3) \otimes H^1(\mathbb{R}^3), \bigotimes_{i=1}^{6} \Psi)$, (instead of $\mathrm{A}^{s/6}$).

Adaptive sparse grid H-atom:





Nonlocal Operators – Wavelet Matrix Compression

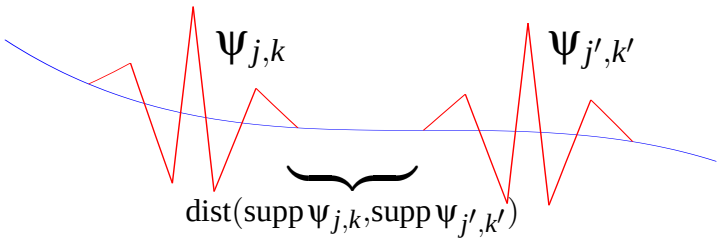
Operators $\mathcal{A} : H^t(\mathbb{R}^3) \to H^{-t}(\mathbb{R}^3)$ with a Schwartz kernel satisfying a Calderón Zygmund type estimate

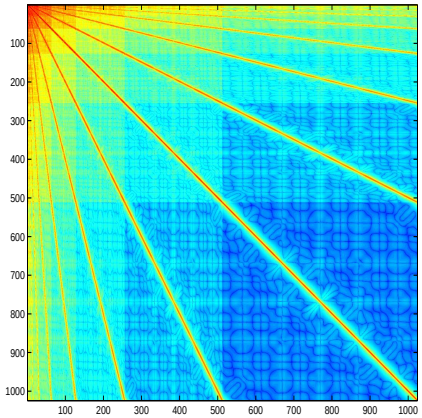
$$|\partial_{\mathbf{x}}^{\alpha}\partial_{\mathbf{y}}^{\beta}K(\mathbf{x},\mathbf{y})| \lesssim \frac{1}{|\mathbf{x}-\mathbf{y}|^{3+2t+|\alpha|+|\beta|}} \quad , \quad \mathbf{x} \neq \mathbf{y} \, ,$$

have matrix coefficients

$$|\langle \mathcal{A}\psi_{j,k}, \psi_{j',k'}\rangle| \leq c \, \frac{2^{(j+j')(\tilde{d}+3/2)}}{\operatorname{dist}(\operatorname{supp}\psi_{j,k}, \operatorname{supp}\psi_{j',k'})^{3+2t+2\tilde{d}}}$$

1. Compression:
(Dahmen-Prößdorf-Schneider,
von Petersdorff-Schwab)





$\Rightarrow$ with these arguments we can proof that, for distant pairs the energy contribution in MP2 calculation decays

$$O(R^{-6})$$

Basic observation:

$$\langle \varphi_a \varphi_j | r_{1,2}^{-1} | \varphi_b \varphi_l \rangle \lesssim R^{-3}$$

Theorem (Flad, Hackbusch, Schneider '05) :

- **Linear scaling complexity**: The density matrix ρ, and the ground state energy of an effective single particle model (HF, LDA, etc.), can be computed with an accuracy $\varepsilon \leq n^{-s}$, $s < d-1$, within (log-)linear complexity $O(n(\log^2 n))$.

- For every $0 < s < 1/2$ the singularity $\Theta \approx |\mathbf{x}-\mathbf{y}|$ of an electron-electron cusp (Kato, Hoffmann-Ostenhoff et al.) belongs to $\mathrm{A}^s(H^1(\mathbb{R}^6), \Psi_x \otimes \Psi_y)$, (instead of $\mathrm{A}^{s/2}$, but the same as for GTO bases).

- **Claim**: The wave function Ψ belongs to $\mathrm{A}^s(H^1(\mathbb{R}^{3N}), \bigotimes_{i=1}^{N} \Psi)$, $0 < s < \frac{1}{2}$.

Kronecker tensor product

- Separation of variables:

$$F(x_1,\ldots,x_N,y_1,\ldots,y_N) \approx \sum_{k=1}^{p} f_{1,k}(x_1,y_1)\cdot\ldots\cdot f_{N,k}(x_N,y_N)$$

- Kronecker product decomposition of tensors:

$$\mathbf{A} = \sum_{k=1}^{p} \mathbf{U}_{1,k}\otimes\ldots\otimes\mathbf{U}_{N,k}$$

- Important to improve the computation of

$$(\langle \mathcal{V}\psi_\lambda,\psi_{\lambda'}\rangle)$$

- Rank one updates (Grasedyck, Chinnamsetty et al. '04). See **Poster**

- Expanding $|\mathbf{x}-\mathbf{y}|$ (Hackbusch '04)

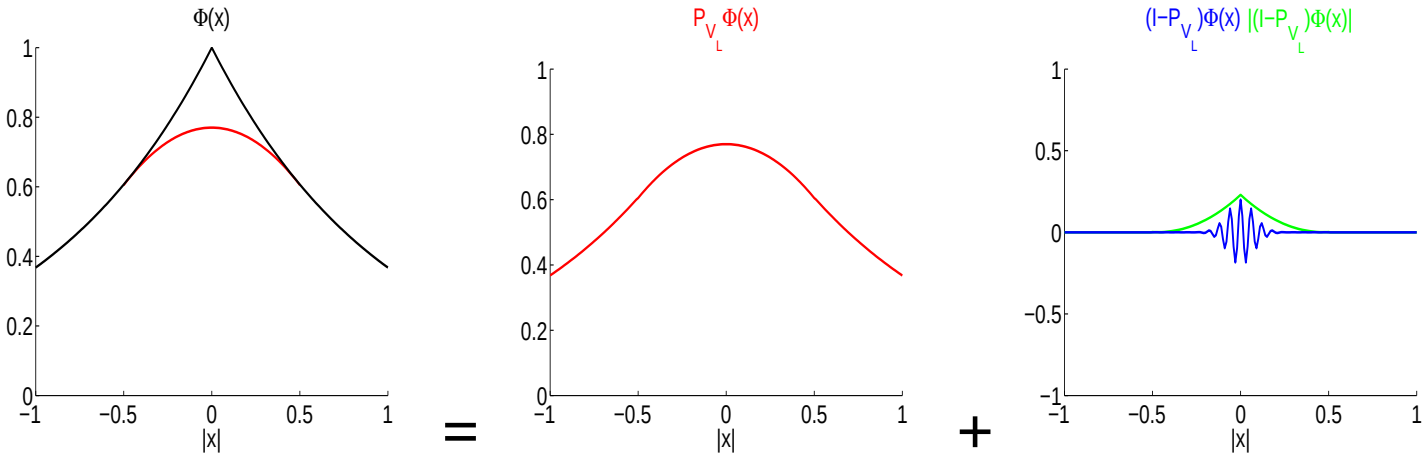
Contracted basis functions and separation of scales

- split orbital Φ into relatively smooth part $P_{V_L}\Phi$ and singular part $(I - P_{V_L})\Phi$:

$$\Phi = P_{V_L}\Phi + (I - P_{V_L})\Phi, \qquad P_{V_L}\Phi \perp (I - P_{V_L})\Phi$$

- molecular scale: smooth part $P_{V_L}\Phi$ by mesh widths $h = 2^{-j} \approx 0.2 \ldots 0.5$

- atomic scale: approximate singular part by contractions of higher-level basis functions with small support – by the projections $(I - P_{V_L})\Phi_{STO/GTO}$ of the core singularities (oszillating and only few per atom)



Error decay rates for the ground state energy

- Multiscale bases and sparse grids: n degrees of freedom

Hyperbolic cross	(Yserentant '04, '05)	$O(n^{-1/3})$
Adaptive sparse grids	(H., F., S.)	$O(n^{-s})$ for all $s < 1$
		$O(n^{-2})$ using $r_{1,2}$ singularity functions
		Demand: linear complexity $O(n)$

- GTO bases: basis set of size $\mathcal{N}$ and n degrees of freedom

CCSD : $n \sim (N^2(\mathcal{N}-N)^2) \sim \mathcal{N}^2$	$O(\mathcal{N}^{-1}) = O(n^{-1/2})$
$r_{1,2}$-Methods CCSD	$O(\mathcal{N}^{-2}) = O(n^{-1})$
Linear scaling CCSD: $n \sim \mathcal{N}$	$O(\mathcal{N}^{-1}) = O(n^{-1})$

Actually complexity is worse.

- Monte-Carlo methods: n sample points $\quad O(n^{-1/2})$