Multi-scale tensorization - the blessing of dimensions

R. Schneider (TUB Matheon)

John von Neumann Lecture – TU Munich, 2012

Anouncement

International Focus Workshop on
Entanglement Based Approaches in Quantum Chemistry
to be held from 04 - 06 September 2012 in Dresden.
Max-Planck-Institut für Physik komplexer Systeme
http://events.mpipks-dresden.mpg.de/register/EBAQC12/registration
Information: http://www.mpipks-dresden.mpg.de/ ebaqc12

Sparsity versus rank sparsity

So far, concepts of adaptivity are based on sparsity.

1. Sparsity depends on the underlying bases (or frame, dictionary).
2. But, how to find and construct apppropriate bases (hp FE, Fourier, Gaussians, splines, wavelets, *dreamlets* ...)
3. adaptive concepts (best N-term approx.) are well understood, but often difficult to implement

Rank sparsity (in $2D$) is provided low rank approximation, e.g. SVD (singular value decomposition)

$$\boxed{A(x_1, x_2) \approx \sum_{k=1}^{r} \sigma_k \big(u_k(x_1) \otimes v_k(x_2)\big)}$$

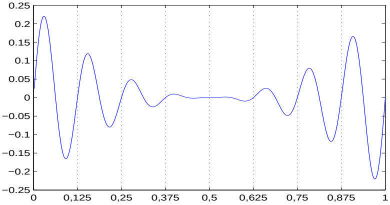
In this case, the non-zero entries σ_k together with the bases $u_k \otimes v_k$ are adaptively computed (*dreamlets*).
Examples: matrix completion, reduced basis functions, etc.

In the present talk, we pursue the second kind of adaptivity

I.

Best one term approximation!

Does it make sense?



yes-

each function has its own multi-scale transform !

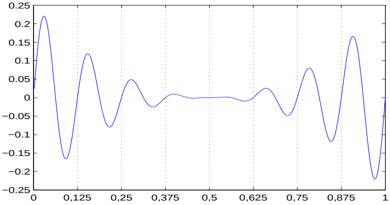
by adaptive multi-wavelet transformation

(subdivision scheme bored from Alperts multi-wavelets)

I.

Best one term approximation!

Does it make sense?



yes-

each function has its own multi-scale transform !

by adaptive multi-wavelet transformation

(subdivision scheme bored from Alperts multi-wavelets)

Polynomial Multi-Wavelets - revisited

Two scale relations, subdivision scheme for discont. multiwavelets (Alpert et al.), orth. polynomials

$$f^0_{k_0}(x) = \sum_{\mu_0=0,1} \sum_{k_1=1}^{r_1} U_0(k_0, \mu_0, k_1) f^1_{k_1}(2x + \mu_0) \ , \mu_0 = 0, 1 \ .$$

$$f^1_{k_1} = \sum_{\mu_1=0,1} \sum_{k_2}^{r_2} U_1(k_1, , \mu_1, k_2) f^2_{k_2}(2x + \mu_1) \ , \ \mu_1 = 0, 1 \ .$$

$$f^i_{k_i} = \sum_{\mu_i=0,1} \sum_{k_i}^{r_i} U_i(k_i, , \mu_i, k_{i+1}) f^{i+1}_{k_{i+1}}(2x + \mu_i) \ , \ \mu_i = 0, 1 \ .$$

$$\dots$$

$$f^{d-1}_{k_{d-1}} = \sum_{\mu_{d-1}=0,1} \sum_{k_{d-1}}^{r_{d-1}} U_{d-1}(k_{d-1}, \mu_{d-1}, k_d) f^d_{k_d}(2x + \mu_{d-1})$$

Filter coefficients $U_i(k_i, , \mu_i, k_{i+1})$, $i = 0, \dots, d-1$ defines the scaling functions $f^0_{k_0}(x)$

i.e. the function f will be encoded by its filter coefficients

Subdivision - Discrete version

$$f_{k_0}(x) = \sum_{n=0}^{2^d-1} F(k_0, n) f^d_{k_d}(2^d x - n) \ ,$$

f^d ON scaling function. binary representation $n = 0 \ldots, 2^d - 1$,
$n = \sum_{j=0}^{d-1} \mu_j 2^{d-j-1}$, $\mu_j = 0, 1$,

$$\boxed{F(k_0, n) = F(k_0, \sum_{j=0}^{d-1} \mu_j 2^{d-j-1}) =: U(k_0, \mu_0, \ldots, \mu_{d-1}) \ .}$$

$$\begin{aligned} (k_0, \boldsymbol{\mu}) \ \mapsto \ & U(k_0, \mu_0, \ldots, \mu_{d-1}) \ = \\ = \ & \sum_{k_0=1}^{r_0} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} \\ & U_0(k_0, \mu_0, k_1) U_1(k_1, \mu_1, k_2) \cdots U_{d-1}(k_{d-1}, \mu_{d-1}, k_d) \end{aligned}$$

Discrete version $f^d_{k_d}(x) \simeq \delta_{0,n}$.

Two-scale decomposition-similarity rank

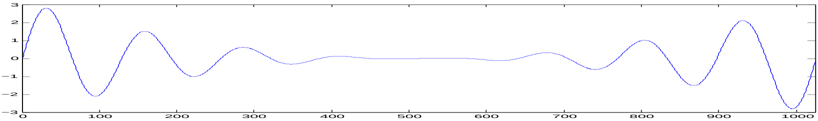
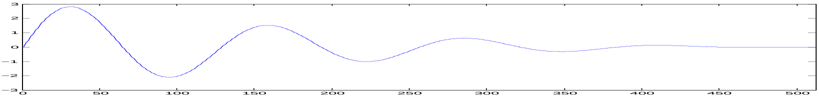
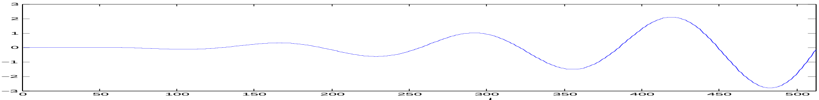


Figure: Function$f(x) = x^2 \sin \frac{8x}{2\pi}$ and functions f_k^1 , $k = 0, 1$



fist half interval f_0^1



second half interval f_1^1

Similarity rank and Multi-scale decomposition

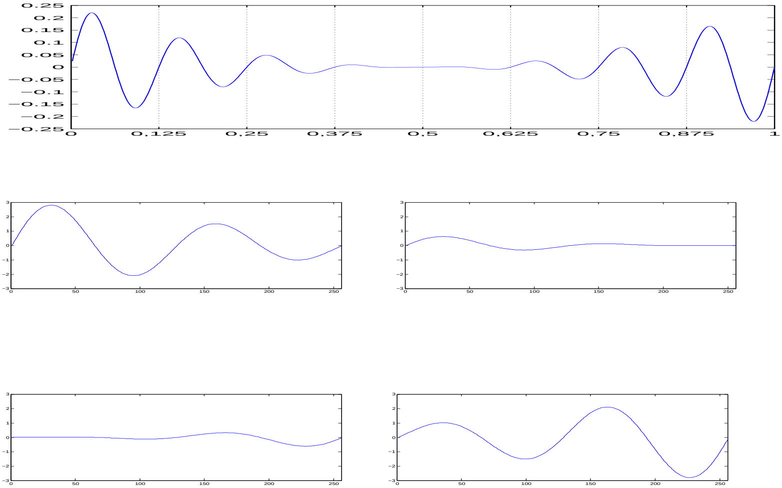


Figure: $f_k^2, k = 0, 1, 2, 3$

POD (proper orthogonal decomposition (SVD) $\Rightarrow$

Similarity rank and Multi-scale decomposition

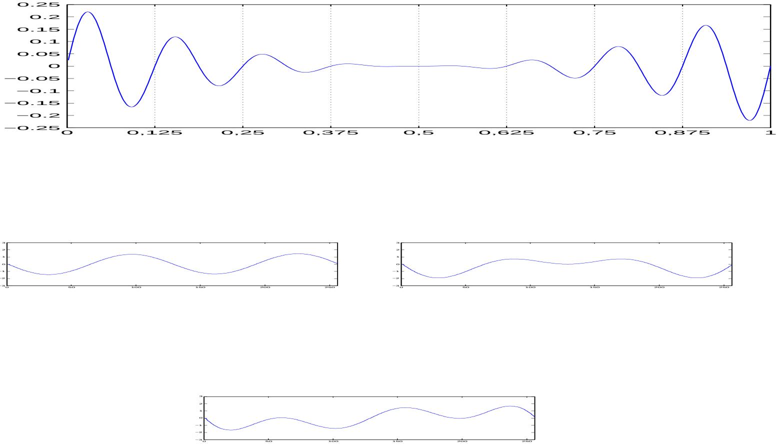


Figure: $f_k^2, k = 0, 1, 2, 3 \ , \ (f_3^2 = 0)$

Recursively performed SVD from level to level $\Rightarrow$

Two-scale decomposition-similarity rank

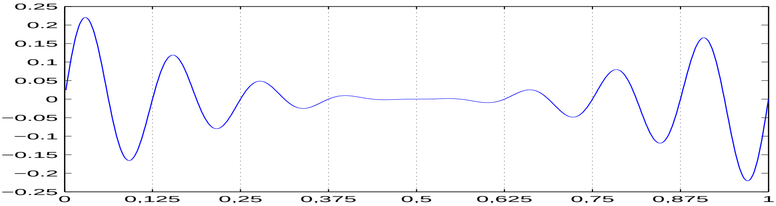


Figure: Function$f(x) = x^2 \sin \frac{8x}{2\pi}$ and translated functions f_k^3

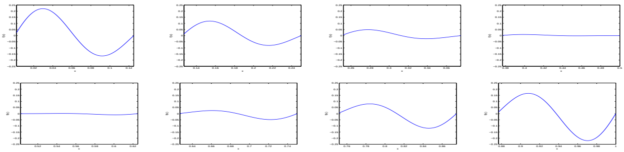
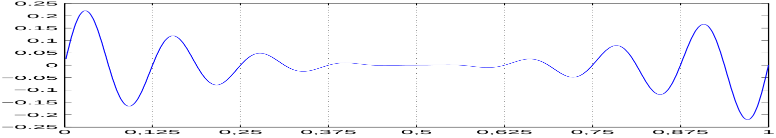


Figure: functions f_k^3, $k = 0, \ldots 7$

Similarity rank and Multi-scale decomposition



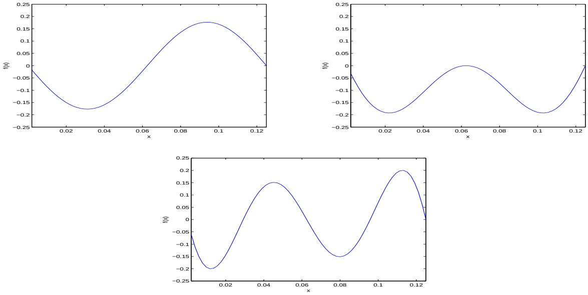


Figure: POD (SVD) functions f_k^3, $k = 0, \ldots, 2, (\ldots, 7)$

Recursively performed SVD from level to level $\Rightarrow$ multiscale decomposition, i.e. a TT representation of U.

Higher order SVD (HOSVD)

- Vidal (2003), Oseledets (2009), Grasedyck (2009), Kühn (2012)

1. Matriziation - unfolding
   (μ_0, k_0 i row index, $\mu_1, \ldots, \mu_{d-1}$ - column index)
   $$U(k_0, \mu_0, \ldots, \mu_{d-1}) \approx \mathbf{U}_{k_0,\mu_0}^{\mu_1,\ldots,\mu_{d-1}} .$$
   decompose by SVD, with orthogonal matrices $(\mathbf{u}_{k_0,\mu_0}^{k_1})$,
   $$\mathbf{U}_{k_0,\mu_0}^{\mu_1,\ldots,\mu_{d-1}} = \sum_{k_1=0}^{r_0} \mathbf{u}_{k_0,\mu_0}^{k_1} \mathbf{v}_{k_1}^{\mu_1,\ldots,\mu_{d-1}} ,$$
   $$U_0(k_0, \mu_0, k_1) := \mathbf{u}_{k_0,\mu_0}^{k_1} , \quad k_1 = 1, \ldots, r_0 \ \ (\leq 2\ ) .$$
2. Decompose $V(k_1, \mu_1, \ldots, \mu_{d-1})$ via matricisation
   $$\mathbf{V}_{k_1,\mu_1}^{\mu_2,\ldots,\mu_{d-1}} = \sum_{k_2=1}^{r_1} \mathbf{u}_{k_1,\mu_1}^{k_2} \mathbf{v}_{k_2}^{\mu_2,\ldots,\mu_{d-1}}$$
   $$U_1(k_1, \mu_1, k_2) := \mathbf{u}_{k_1,\mu_1}^{k_2} .$$
3. repeat with $V(k_{d-1}, \mu_{d-2}, \mu_{d-1})$ until one ends with
   $\left[\mathbf{v}_{k_{d-1},\mu_{d-1}}^{k_d}\right] \mapsto U_{d-1}(k_{d-1}, \mu_{d-1}, k_d)$.

What do we gain?

For $n := \sharp\{0, , 1\} = 2$, $r := \max\{r_i : 0 \leq i \leq d-2\}$, storage complexity is bounded by

$$\boxed{\text{DOFs: } \leq 2dr^2,} \quad \text{instead of } 2^d$$

Compression: Whenever $r = \mathcal{O}(1)$ then complexity $N = 2^d$ is reduced to $\mathcal{O}(\log N)$!

The optimal ranks r_i, $\underline{r} = (r_1, \ldots, r_{d-1})$ (multi-rank), of the above decomposition are equal to the

$$\underline{\textit{i-th separation rank } r_i \text{ of } U},$$

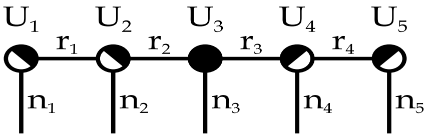
i.e. the ranks of the following matrices, $i = 1, \ldots, d-1$

$$\text{rank of } \mathbf{A}_i = U^{\mu_{i+1},\ldots,\mu_{d-1}}_{\mu_0,\ldots,\mu_i} = r_{i+1}.$$

I.

Vector tensorization and tensor product approximation

Can we work with individual filter coefficients ?



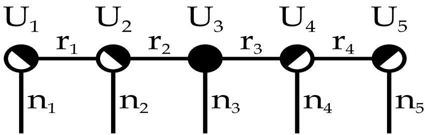
yes-

in novel tensor formats

I.

Vector tensorization and tensor product approximation

Can we work with individual filter coefficients ?



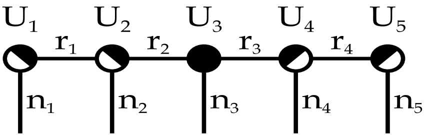
yes-

in novel tensor formats

I.

Vector tensorization and tensor product approximation

Can we work with individual filter coefficients ?



yes-
in novel tensor formats

Setting - Tensors

$$V_\nu := \mathbb{R}^n \ , \ \ \mathcal{H}_d = \mathcal{H} := \bigotimes_{\nu=1}^{d} V_\nu$$ d-fold tensor product Hilbert-s.,

$$\mathcal{H} \simeq \{(x_1, \ldots, x_d) \mapsto U(x_1, \ldots, x_d) \in \mathbb{R} : x_i = 1, \ldots, n_i\} \ .$$

The function $U \in \mathcal{H}$ will be called an order d-tensor.
For notational simplicity, we often consider $n_i = n$. Here $x_1, \ldots, x_d \in \{1, \ldots, n\}$ will be called variables or indices.

$$\mathbf{k} \mapsto U(\mathbf{k}) = U(k_1, \ldots, k_d) = (U_{k_1,\ldots,k_d}) \ , \ \ k_i = 1 \ldots, n_i \ .$$

Or in index (vectorial) notation

$$\mathbf{U} = \left(U_{k_1,\ldots,k_d}\right)_{k_i=0, 1\leq i\leq d}^{n_i}$$

$$\dim \mathcal{H} = n^d$$ curse of dimensions!!!

E.g. wave function $\Psi(\mathbf{r}_1, s_1, \ldots, \mathbf{r}_N, s_N)$

Vector-Tensorization - e.g. Binary coding

1D example: vector, e.g. signal

$$k \to f(k)\ ,\ \Big(\text{ or } \ g(\frac{k}{2^d})\Big)\ ,\ k = 0, \ldots, 2^d - 1\ .$$

Labeling of indices $k \simeq \mu \in \mathcal{I}$ by an binary string of length d,

$$\boldsymbol{\mu} = \mu(k) \quad = \quad (0,0,1,1,0,\ldots) \simeq \sum_{j=0}^{d-1} \mu_i 2^j = k(\boldsymbol{\mu})\ ,\ \mu_i = 0,1\ .$$

Tensorization

$$\boldsymbol{\mu} \mapsto U(\boldsymbol{\mu}) := f(k(\boldsymbol{\mu})) \in \bigotimes_{j=0}^{d-1} \mathbb{R}^2 \quad , \quad \text{or } \bigotimes_{j=0}^{d-1} \mathbb{C}^2.$$

This provides an isomorphm $T : \mathbb{R}^{2^d} \leftrightarrow \bigotimes_{j=0}^{d-1} \mathbb{R}^2$ by $Tf := U$.
So far no information is lost, $N = 2^d$ or $d = \log_2 N$.

Hierarchical Tucker (HT) format

Hackbusch & Kühn (2009), Grasedyck (2010)

Hierarchical MCTDH *Mayer et al.* (2000) Matrix product states (1992) – (Tree) Tensor networks: e.g. *Vidal, Schollwöck etc. (2003)*

Noteable special case of HT: TT format , Oseledets & Tyrtyshnikov

TT- or matrix product representation of U

$$\boxed{U(\mathbf{x}) = \mathbf{U}_1(x_1)\cdots\mathbf{U}_i(x_i)\cdots\mathbf{U}_d(x_d)}$$

$$= \sum_{k_1=1}^{r_1} .. \sum_{k_{d-1}=1}^{r_{d-1}} U_1(x_1,k_1)U_2(k_1,x_2,k_2)...U_{d-1}(k_{d-2}x_{d-1},k_{d-1})U_d(k_{d-1},x_d,k_d)$$

- component tensors $U_i(k_{i-1},x_i,k_i) \in \mathbb{R}^{r_{i-1}\times n_i\times r_i}$,

if $r := \max\{r_1,\ldots,r_{d-1}\}$, here $n = 2$, (or $4,\ldots$ small!)

storage complexity DOFs: $\leq\ ndr^2$,

Binary coding - signal compression - 1 D functions

Quantized TT - Oseledets (2009), Khoromskij (2009) :

TT approximation of U

- Storage complexity N is reduced to $2r^2 \log_2 N$! (linear in $d = \log_2 N$)
- Allow extreme fine grid size $h = o(\epsilon) = 2^{-d} = \frac{1}{N}$. Example: $d = 50$, then $h \leq 10^{-15}$

Basic question: when is r small or moderate?

(Grasedyck (2010), Hackbusch (2010), Oseledets (2010))

Examples:

1. For Kronecker $\delta_{i,j}$ (Dirac function) is $r = 1$.
2. For plane wave (fixed $k = \sum_{j=1}^{d} \nu_j 2^{j-1}$)

$$e^{2\pi ik} = e^{2\pi i \sum_{j=1}^{d} \nu_j 2^{j-1}} = \Pi_{j=1}^{d} e^{2\pi i \nu_l 2^{j-1}} \ , \ \nu_j = 0, 1,$$

again (complex) $r = 1$, or (real $r = 2$) .

TT representation of tensorized f

The TT ranks r_i, $\underline{r} = (r_1, \ldots, r_{d-1})$, of the TT decomposition of $U = T(f)$ are the ranks of the following matrices, $i = 1, \ldots, d-1$

$$r_i = \text{rank of } \mathbf{A}_i = U^{\mu_{i+1},\ldots,\mu_d}_{\mu_1,\ldots,\mu_i} = U^{(1,0,\ldots)}_{(0,1,\ldots)} .$$

or

$$\mathbf{A}_i = U^q_p \ , p = p(\mu_1, \ldots, \mu_i) \text{ and } q = q(\mu_{i+1}, \ldots, \mu_d)$$

consider $g : [0,1] \to \mathbb{R}$, $x = \frac{k}{2^d}$, $k = 0, \ldots, 2^d - 1$, $f(k) = g(\frac{k}{2^d})$, then k_1 is the number of the subinterval

$$I^i_{k_1} := [\frac{k_1 - 1}{2^i}, \frac{k_1}{2^i}]$$

and $\{\frac{k_1-1}{2^i} + \frac{1}{2^d}[0, \ldots, k_2, \ldots, 2^{d-i} - 1] : k_2 = 0, \ldots, 2d - i - 1\}$ the corresponding grid.

Two-scale decomposition-similarity rank

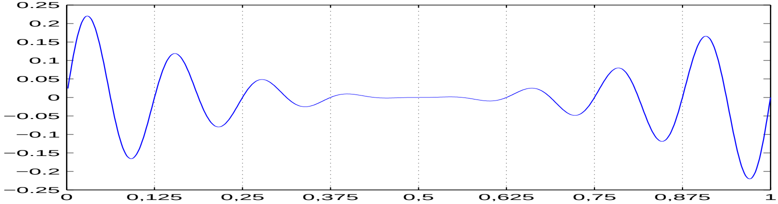
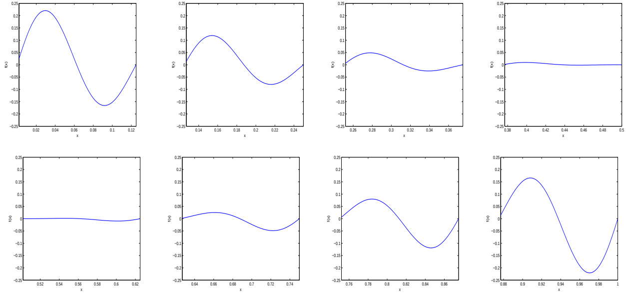
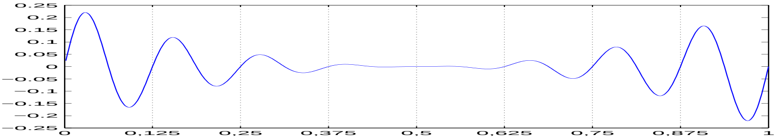


Figure: Function$g(x) = x^2 \sin \frac{8x}{2\pi}$ and translated functions g_k^3



Similarity rank and Multi-scale decomposition



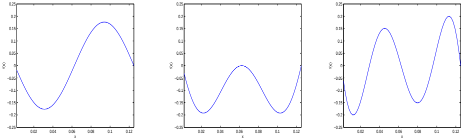


Figure: PCA principal component analysis (SVD) functions g_k^3, $k = 1, \dots 3$

Recursively performed SVD from level to level (Vidal-decomposition) $\Rightarrow$ multiscale decomposition, i.e. a TT representation of U.

Examples for admitting low rank representations

Low ranks

1. (Piecewise) polynomials of degree $p+1$.
2. $e^{i\langle \mathbf{k},\mathbf{x}\rangle}$, i.e. Fourier polynomials
3. Splines, wavelets etc.
4. adaptive - best n-term approximation
   For homogenization:
5. Let $\phi : \mathbb{R} \to \mathbb{K}$ be 2^{-j}-periodic, then $\phi|_\Omega$ is if rank ${}_i\phi = 1$.
6. Let $f : [0,1] \to \mathbb{K}$ has sim rank$f = p+1$, and ϕ 1-periodic, then the modulated function $x \mapsto g(x) := f(x)\phi(\frac{x}{2^{-j}})$ has ranks rank ${}_i g \leq p+1$.

see e.g. appendix.

Low TT rank approximation

Theorem

1. *$f \in H^s(\Omega_k^j)$, $k \in \Delta_j$, ($\Omega \subset \mathbb{R}^D$). There exists f_ϵ with rank r_j on level j satisfying*

$$\|f - f_\epsilon\|^2 \lesssim r_j^{-s/D} \sum_{k \in \Delta_j} |f|_{H^s}^2 \ . \tag{1}$$

2. *f analytic on the domains Ω_k^j, $k \in \Delta_j$*

$$\|f - f_\epsilon\| \lesssim e^{-\alpha r_j} \ , \quad \textit{for some } \alpha > 0 \ . \tag{2}$$

3. *Let f be a piecewise analytic function on $\Omega \setminus \{x_0\}$ satisfying*

$$|\partial^{\boldsymbol{\alpha}} f(\mathbf{x})| \leq c_{\boldsymbol{\alpha}} |\mathbf{x} - \mathbf{x}_0|^{\gamma - |\boldsymbol{\alpha}|} \boldsymbol{\alpha}!$$

then there exists $\alpha > 0$ s.t.

$$\|f - f_\epsilon\| \lesssim e^{-\alpha r_j} \ . \tag{3}$$

Examples for admiting low rank representations

- From this perspective, the approach can be easily extended to 2D and 3D Finite Elements with uniform refinement.
- Black-box algorithm: The multi-level scheme corresponds to a (multi-) wavelet packet decomposition.
- The two-scale relations are not fixed, but optimized in each component (such that no wavelets appears).
- ranks $r_j \leq cN$ for best N-term approximation (e.g. Fourier, wavelets etc.), even with $c << 1$ but scaling is $\mathcal{O}(r^2)$!

Two-scale decomposition - matricisation

Reference domain Ω_0 and $\Omega = \bigcup_{k\in\Delta_j} \Omega_k^j$, above example $\Omega = \Omega_0$, $\Omega_k^j = [k2^{-j}, (k+1)2^{-j}]$ together with isomorphisms (renormalization group)

$$\boxed{T_k^j : \Omega_k^j \to \Omega_0 \quad , \Omega = \bigcup_{k\in\Delta_j} \Omega_k^j \ .}$$

$f_l := \sum_{k\in\Delta_j} f_l|_{\Omega_k^j} : \Omega \to \mathbb{K}$, $l = 1, \ldots, m$ simultanously,

$$f_{k,l}^j : \Omega_k^j \to \mathbb{K}\ ,\ x \mapsto f_{k,l}^j(x) := f_l\big((T_k^j)^{-1}x\big) = f_l \circ (T_k^j)^{-1}(x)\ ,\ x \in \Omega_0\ .$$

The *similarity rank* of f at level j

$$\boxed{\text{sim rank}_j f := \dim \operatorname{span}\{f_{k,l}^j : k \in \Delta_j\ l = 1, \ldots, n\} = r_j\ .}$$

is equal to the TT Rank r_j of U.

Multi-scale decomposition by recursion

$$\Omega_0^j = \bigcup_{\mu_j=0}^{n_j-1} \Omega_{\mu_j}^{j+1} \quad , \quad T_{\mu_j}^j : \Omega_{\mu_j}^{j+1} \to \Omega_0^{j+1} \ , \ \mu_j = 0, \ldots, n_j$$

$k \in \Delta_{j+1}$ can be encoded by a multi-index
$\boldsymbol{\mu} = \boldsymbol{\mu}(k) = (\mu_0, \ldots, \mu_j) \in \mathcal{I}_0 \times \cdots \times \mathcal{I}_j, \mathcal{I}_j = 0, \ldots, n_j - 1,$

We define multi-scale scaling functions (see multi -wavelets)

$$x \mapsto \varphi_{\alpha_{j-1}}^j(x), x \in \Omega_0 \ .$$

Two-scale relation

$$\varphi_{\alpha_{j-1}}^j(x) = \sum_{\mu_j=1}^{n_j} \sum_{\alpha_j=1}^{r_d} U_j(\alpha_{j-1}, \mu_j, \alpha_j) \varphi_{\alpha_j}^{j+1}(T_{\mu_j}^j x) \ , \ x \in \Omega_0 \ .$$

Multi-scale decomposition by recursion

Iterating

$$T_k^{j+1} = T_{k(\boldsymbol{\mu})}^{j+1} = T_{\mu_0}^0 \circ \cdots \circ T_{\mu_j}^j \ .$$

yields nonlinear (multi wavelet (packets)-) subdivison scheme reconstructing A vector, resp. function f corresponds to a tensor U

$$x \mapsto f(x), x \in \Omega_0 \ , \ (\boldsymbol{\mu}) \mapsto U(\boldsymbol{\mu}) = U(\mu_0, \ldots, \mu_d) \ ,$$

$$\Rightarrow \quad f(x) = \sum_{\boldsymbol{\mu}:k(\boldsymbol{\mu}) \in \Delta_d} U(\boldsymbol{\mu}) \varphi^d(T_{k(\boldsymbol{\mu})}^d x) \ , \ x \in \Omega_0 \ .$$

$$U(\boldsymbol{\mu}, \alpha_d) = \sum_{\alpha_0=1}^{r_0} \cdots \sum_{\alpha_{d-1}}^{r_{d-1}} U_i(\alpha_{i-1}, \mu_i, \alpha_i) \ .$$

U_i computed e.g. sequentially by SVDs (opt. - black box)

Binary coding - linear operators and matrices

$\mathbf{A} = \left[a(k_1, k_2)\right] : \bigotimes_{j=1}^{d} \mathbb{K}^2 \to \bigotimes_{j=1}^{d} \mathbb{K}^2 \simeq \mathbb{K}^{2^d}$. coding
$\boldsymbol{\theta} = (\theta_1, \ldots \theta_{2d}) := \big((\mu_1, \nu_1), \ldots, (\mu_d, \nu_d)\big)$.

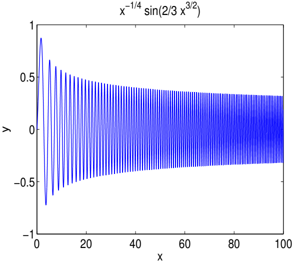
$$\big((\mu_1, \nu_1), \ldots, (\mu_d, \nu_d)\big) \mapsto A(\boldsymbol{\theta}) := a\big(\mathbf{k}(\boldsymbol{\theta})\big) \in \bigotimes_{j=1}^{d} \mathbb{K}^{2\times 2} \ ,$$

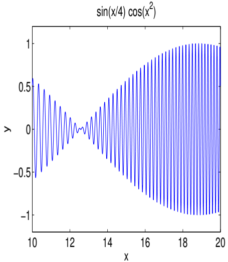
Example:

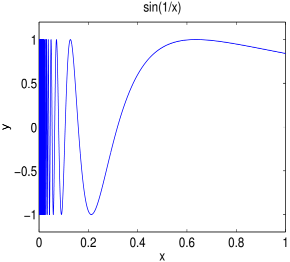
1. Bit reversal is a permutation of tensor variables.
2. Identity matrix
$$\mathbf{I} = \mathbf{I}_{2\times 2} \otimes \cdots \otimes \mathbf{I}_{2\times 2} \ .$$
3. The Hadamar-Walsh transform is given by
$$\mathbf{W} = \bigotimes_{j=1}^{d} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$
4. finite differences, e.g. $\frac{d^2 f}{dx^2} \approx \frac{f(x_{k-1}) - 2f(x_k) + f(x_{l+1})}{2h^2}$, $h = 2^{-d}$, has ranks $r_j \leq 3$.
5. (quantum) Fourier transform

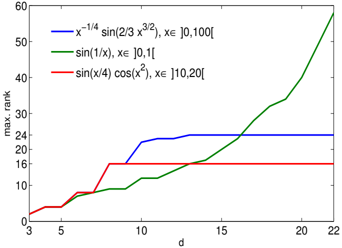
Examples: TT approximation of tensorized functions (QTT)

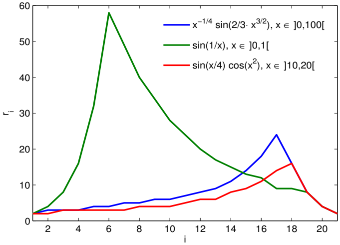
Airy function: $f(x) = x^{1/4} \sin \frac{2x^{2/3}}{3}$, chirp: $f(x) = \sin \frac{x}{4} \cos(x^2)$ and $f(x) = \sin \frac{1}{x}$









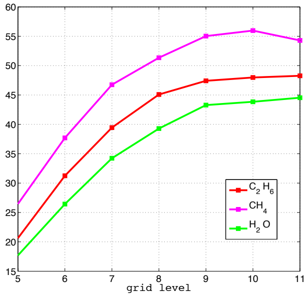


TT ranks of the tensorized Hartree potential V_H.

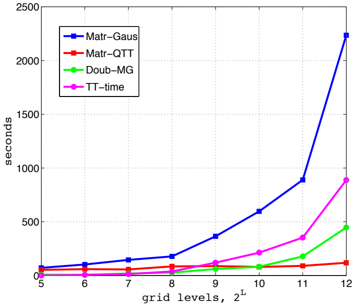
[Khoromskaia, Khoromskij, R.Schneider '11]

$$\bar{r} = \sqrt{\frac{1}{D}\sum_{\ell=1}^{D} r_{\ell-1} r_{\ell}}, \; D = 3L.$$

The average QTT ranks $\bar{r}$ of V_H for molecules CH_4, H_2O and C_2H_6.



Comparison of times: QTT vs. canonical format (CH_4).



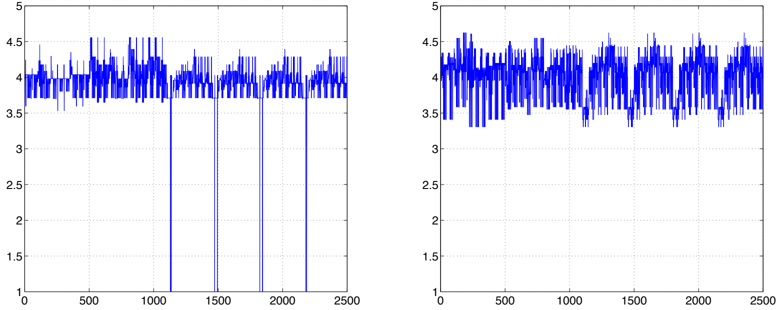
Grid-based calculations on $n \times n \times n$ ($n^{\otimes 3}$) 3D Cart. grids, $n = 2^L$, $L = 5, \ldots, 12$, (in Matlab). For $L = 12$, $n^3 = 6.8 \cdot 10^{10}$.

Ranks of the tensorized two-electron integrals (TEI) tensor.

[Khoromskaia, Khoromskij, R.Schneider '12]

Given the "basis sampling" tensor $\mathbf{G} = [\mathbf{G}_{\mu\nu}] \in \mathbb{R}^{N_b \times N_b \times n^{\otimes 3}}$ and the convolution $\mathbf{H} = [\mathbf{P}_{\mathcal{N}} *_{n^{\otimes 3}} \mathbf{G}_{\kappa\lambda}] \in \mathbb{R}^{N_b \times N_b \times n^{\otimes 3}}$, the TEI tensor is computed by

$$\mathbf{B} = [b_{\mu\nu\kappa\lambda}] \approx \mathbf{G} \times_{n^{\otimes 3}} (\mathbf{P}_{\mathcal{N}} *_{n^{\otimes 3}} \mathbf{G}) = \langle \mathbf{G}, \mathbf{P}_{\mathcal{N}} *_{n^{\otimes 3}} \mathbf{G} \rangle_{n^{\otimes 3}} = \langle \mathbf{G}, \mathbf{H} \rangle_{n^{\otimes 3}}.$$



Pseudopotential of CH_4: average QTT ranks of product basis functions, $\mathbf{G}_{\mu\nu}$, (left) and their Newton potential, $\mathbf{H}_{\kappa\lambda}$, (right), $\varepsilon = 10^{-6}$,
$\mu, \nu, \kappa, \lambda = 1, ..., N_b$, $N_b = 50$, $n = 8192$.

Examples for linear operators - matrices

$$\big((\mu_0, \nu_0), \ldots, (\mu_{d-1}, \nu_{d-1})\big) \mapsto A(\boldsymbol{\theta}) \in \bigotimes_{j=1}^{d} \mathbb{K}^{2\times 2} \ ,$$

1. Identity matrix
$$\mathbf{I} = \mathbf{I}_{2\times 2} \otimes \cdots \otimes \mathbf{I}_{2\times 2} \ .$$
2. The Hadamar-Walsh transform is given by
$$\mathbf{W} = \bigotimes_{j=1}^{d} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$
3. 1D- Laplace operator (2nd order) is of multi-ranks $r_i \leq 3$
4. Convolutions (*Hackbusch 2011*)
5. The discrete Fourier transform is a $3d$-fold product of rank ≤ 2 operators - Fast Fourier transform (Quantum Fourier Transform)

Numerics on QTT-Fourier transform versus FFT in 1D

$$t \mapsto f(t) = \Pi(t) := \chi_{[-\frac{1}{2},\frac{1}{2}]}(t) \text{ char. function}$$

$$\hat{f}(\xi_j) = 2\Re \int_0^{+\infty} f(t)\exp(-2\pi t \xi_j) dt \approx 2\Re \sum_{k=0}^{n-1} f(t_k)\exp(-2\pi t_k \xi_j) h_t,$$

Table: Time for QTT–FFT (in milliseconds) w.r.t. size $n = 2^d$ and accuracy . $time_{QTT}$ is the runtime of Alg. QTT–FFT, is the runtime of the FFT from the FFTW library, and $\hat{f}$ = sinc is the effective QTT–rank of the Fourier image. In courtesy of B. Khoromskij

$f = \Pi(t)$		$= 10^{-4}$		$= 10^{-8}$		$= 10^{-12}$	
d		$\hat{f}$		$\hat{f}$		$\hat{f}$	
16	1.7	4.66	7.9	6.85	13.8	8.85	20.0
18	8.9	4.70	9.7	6.86	16.7	8.82	23.4
20	42.5	4.75	11.3	6.85	19.8	8.86	30.6
22	180	4.77	13.1	6.83	23.3	8.89	36.4
24	810	4.74	15.0	6.72	26.3	8.94	41.7
26	4100	4.62	17.0	6.76	30.0	8.89	46.5
28	26300	4.57	18.9	6.80	33.0	8.88	51.2
30	—	4.72	20.3	6.78	36.2	8.84	57.0
40	—	4.20	29.1	6.59	53.6	8.78	83.2
50	—	3.96	39.3	6.45	70.5	8.48	109
60	—	3.69	50.0	6.25	87.6	8.32	133

Comparison: quantum information theory - $\bigotimes_{i=1}^{d} \mathbb{C}^2$

$$\mathcal{H} = \mathbb{C}^{2^d} \eqsim L_2(\{1,\ldots,2^d\},\mathbb{C}) \eqsim L_2(\{0,1\}^d,\mathbb{C}) \eqsim \bigotimes_{i=1}^{d} \mathbb{C}^2$$

all equipped with L_2 resp. Froebenius norms.

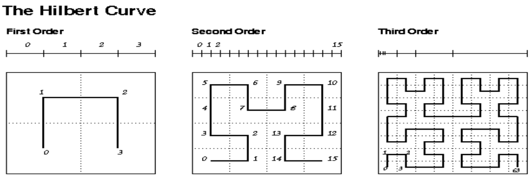
- here we have the same configuration space
- here we do not confine to $U \in H$ with $\|U\| = 1$
- not necessarily a probabilistic interpretation
- we do not confine to unitary operator $A : \mathcal{H} \to \mathcal{H}$
- But: TT ranks must be moderate!

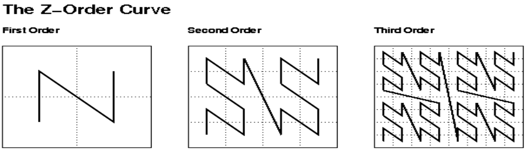
The treatment of quantum mechanical problems with MPS willl be presented later!

Ordering of indices - space filling curves

There are various possibilities for higher-dimensional functions

1. lexigraphical ordering $(k_1, k_2, \ldots) \to \bigotimes_{i=1}^{D} \otimes_{l=1}^{n_i} \mathbb{R}^2$
2. Z-curve ordering, multi-level octree ordering
   $(k_1, k_2, \ldots) \to \bigotimes \mathbb{R}^{2^D}$
3. Hilbert space filling curves





Appendix: Translation invariant operators and spaces

$$v \mapsto L^{(p)}v := v^{(p)} + a_{p-1}v^{(p-1)} + \cdots + a_0 \ .$$

The kernel of the linear differential operator $L^{(p)}$

$$V_p := \{v \in H^p_{loc}(\mathbb{R}) : L^{(p)}v = 0\} \ .$$

The space V_p is translation invariant, i.e.

$$f(\cdot) \in V_p \Rightarrow f(\cdot - a) \in V_p \ , \forall a \in \mathbb{R} \ , \ \ \text{dim } V_p = p + 1 \ .$$

Let $L^{(p)}$ be a linear partial differential operator of order p,

$$v \mapsto L^{(p)}v := \sum_{|\alpha| \leq p} a_\alpha D^\alpha v.$$

Then $V_p := \{v \in H^p_{loc}(\mathbb{R}) : L^{(p)}v = 0\}$, is translation invariant.

$$S^p = \bigoplus_{k \in \Delta_j} V_p|_{\Omega^j_k} \ \Rightarrow \text{sim rank} S^p = \text{dim } V_p \ .$$

Examples

1. $L^{(2)}v = v''$ then V_p contains polynomials of degree $p = 2$.
2. $L^{(1)}v = v' + a$, $a \in \mathbb{C}$ then $v(x) = ce^{-ax}$.
3. $L^{(2)}v = v'' + a^2$, $a \in \mathbb{R}$, then $v(x) = c_1 \cos ax + c_2 \sin ax$ or $v(x) = ce^{i\sqrt{a}x}$.
4. harmonic polynomials for the Laplacian $L = \Delta$,
5. $e^{i\langle \mathbf{k},\mathbf{x}\rangle}$ or Fourier Bessel function for the Helmholtz operator, (Treftz method)
6. advection equation (Demkovich et al.)

Corollary

- $S^p := \bigoplus S^p(\Omega_j^k)$, $S^p(\Omega_j^k) = \{f : \Omega_k^j : \mathbb{R} : f \textit{ is a polynomial of degree } \leq p\}$ $p_n = \sum_{k=0}^{p} a_k x^k$ *are of similarity rank* $\leq (p+1)^D$.
- *Assume* $f = \sum_{|k|\leq p} \hat{f}_k e^{2\pi ikx}$, *is piecewise bandlimited, and* $l \leq j$ *sufficiently large. Then* f *has rank* $r_l \leq 2p + 1$.

Appendix - examples

Let $T_k : \mathbb{R}^D \to \mathbb{R}^D$. We will call a function space V_T of functions T_k *-invariant* if $\forall\, k \in \Delta_j$ and transformations T_k
$f \in V_T \Rightarrow f \circ T_k \in V_T$.

1. If V_T is a T_k invariant space of dimension $\dim V_T$ then the similarity rank of a function $f \in V_T$ is $\leq \dim V_T$.
2. T_k-translation invariant space V_j generated by $\varphi_k(x) = \varphi(x - k)$ has similarity rank $p \leq \sharp\{\text{supp } \varphi_k \cap \Omega_0^j \neq \emptyset\}$.
3. Multiresolution spaces $V^l = V^0 \oplus \bigoplus_{i=0}^{l-1} W^i$, where $l \leq j$ have similarity rank $\leq p$.

Homogenization

1. Let $\phi : \mathbb{R} \to \mathbb{K}$ be 2^{-J}-periodic, $J \geq j$ then $\phi|_\Omega$ is if sim rank $\phi = 1$.
2. Let $f : [0, 1] \to \mathbb{K}$ has sim rank$f = p$, and ϕ 1-periodic, than $x \mapsto g(x) := f(x)\phi(\frac{x}{2^{-j}})$ has similarity rank sim rank $g = p$.

Two-scale decomposition

Theorem

1. *Let us consider*

$$f = \sum_{j=0}^{d-1} \sum_{k \in \mathcal{I}^j} \sum_{i=1}^{p} a^j_{k,i} \varphi^j_{k,i}$$

to be a piecewise polynomial function with polynomial degree $\leq p-1$*. Let* N_j *be the number of domains* Ω^j_k *where* $f|_{\Omega^j_k}$ *is not a polynomial. Then the similarity rank of f at level j satisfy*

$$\textit{sim rank}_j\ f \leq N_j + p\ .$$

2. *(Best N-term representation) Let f be expanded by multiwavelets of degree p within N terms, then the similarity rank is bounded by*

$$\textit{sim rank}_j\ f \leq \max\{N, p+1\}\ .$$

Thank you
for your attention.

References:

Hackbusch W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus* to appear in Springer Verlag 2012

I. Oselsedets & B. Khoromskij, Quantized TT format – QTT (Oseledets (09), Tyrtishnikov, Khoromskij)

I. Oseledets, *Constructive Representation of Functions in Tensor Formats*,

L. Grasedyck, *Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorisation* , (2010)

W. Hackbusch, *Tensorisation of Vectors and their Efficient Convolution* ,

W. Hackbusch, L. Grasedyck, R. Schneider *Multilevel tensorisation* (in preparation)

B. Khoromskij, *Tensors-structured Numerical Methods in Scientific Computing: Survey on Recent Advances.* Chemometrics and Intellingent Laboratory Systems, 110 (2012), 1-19.

V. Khoromskaja, B. Khoromskij and R. Schneider QTT approximation of two-electron integrals Preprint