Lectures on Tensor Numerical Methods for Multi-dimensional PDEs

Lecture 4-5. Rank-structured tensor decompositions: canonical and Tucker formats

Venera Khoromskaia & Boris Khoromskij,
Shanghai Jiao Tong University, April 2017

Max-Planck-Institute for Mathematics in the Sciences, Leipzig MAX-PLANCK-GESELLSCHAFT

Lect. 4-5. Rank-structured tensor decompositions: canonical and Tucker formats.

Outline of Lectures 4-5.

1. Tensor product of finite-dimensional Hilbert spaces (d-dimensional Euclidean vectors).

2. Vectorization and matrix unfolding. Contracted product of tensors.

3. Linear and multilinear algebra (MLA) on tensors: $O(n^d)$ complexity.

4. Tensor rank, canonical format, little analogy between cases $d = 2$ and $d \geq 3$.

5. Linear and multilinear operations on "formatted tensors": $O(d)$ complexity.

6. Rank decomposition can be useful in linear algebra:

$O(n^{\log_2 7})$- Strassen algorithm of matrix-matrix multiplication.

7. Orthogonal Tucker and mixed Tucker-canonical models. Main properties (proofs).

8. Towards best (nonlinear) approximation in canonical/Tucker tensor formats.

9. Tucker approximation of tensors generated by aysymptotically "smooth" functions. Exponential convergence in r for Green's kernels.

10. Computing canonical decomposition by Greedy algorithms.

Tensor product of finite dimensional Hilbert spaces

Tensor-product Hilbert space (TPHS), $\mathbb{V}_{\mathbf{n}} = V_1 \otimes ... \otimes V_d$, $\mathbf{n} = (n_1, ..., n_d)$, $n_\ell = \dim V_\ell$.

- Euclidean vector space $\mathbb{V}_{\mathbf{n}} = \mathbb{R}^{n_1 \times \ldots \times n_d}$, $V_\ell = \mathbb{R}^{n_\ell}$ $(\ell = 1, ..., d)$,

$$\mathbf{A} = [a_{\mathbf{i}}] \in \mathbb{V}_{\mathbf{n}} : \quad \langle \mathbf{B}, \mathbf{A} \rangle = \sum\nolimits_{\mathbf{i}} b_{\mathbf{i}} a_{\mathbf{i}}, \quad \mathbf{i} = (i_1, ..., i_d) : i_\ell \in I_\ell = \{1, ..., n_\ell\}.$$

Rem. 4.1. d-th order tensor $\mathbf{A} \in \mathbb{V}_{\mathbf{n}}$ of size $\mathbf{n} = (n_1, ..., n_d)$ is a function of d discrete arguments (multi-dimensional array/vector over $\mathcal{I} := I_1 \times ... \times I_d$, i.e.,

$$\mathbf{A} : I_1 \times ... \times I_d \to \mathbb{R}, \quad \text{with} \quad dim(\mathbb{V}_{\mathbf{n}}) = |\mathbf{n}| := n_1 \cdots n_d.$$

Tensors over $\mathcal{I} \times \mathcal{J}$ can be associated with a matrix that maps $\mathbb{V}(\mathcal{J}) \mapsto \mathbb{V}(\mathcal{I})$.

Notations for the coordinate representation of $\mathbf{A}$,

$$\mathbf{A} := [a_{\mathbf{i}}] = [a_{i_1 \ldots i_d}] = [a(i_1, ..., i_d)] \in \mathbb{R}^{\mathcal{I}}.$$

The *Euclidean scalar product* of tensors induces the Euclidean (Frobenius) norm

$$\|\mathbf{A}\|_F := \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}.$$

The dimension directions $\ell = 1, ..., d$, are called the *modes*.

Tensor is a union of *ℓ-mode fibers*, $a(i_1, ..., i_{\ell-1}, : , i_{\ell+1}, ..., i_d) \in \mathbb{R}^{n_\ell}$, of size n_ℓ.

Separable representation of (discrete) functions in a TPHS

Separable representation in $\mathbb{V}_{\mathbf{n}}$. rank-1 tensors $\Rightarrow$ univariate $O(d)$-operations

$$\mathbf{V} = [v_{i_1 \ldots i_d}] = v^{(1)} \otimes \ldots \otimes v^{(d)} \in \mathbb{V}_{\mathbf{n}}, \quad v_{i_1 \ldots i_d} = \prod\nolimits_{\ell=1}^{d} v_{i_\ell}^{(\ell)}.$$

- The scalar product

$$\langle \mathbf{W}, \mathbf{V} \rangle = \langle w^{(1)} \otimes \ldots \otimes w^{(d)}, v^{(1)} \otimes \ldots \otimes v^{(d)} \rangle = \prod\nolimits_{\ell=1}^{d} \langle w^{(\ell)}, v^{(\ell)} \rangle_{V_\ell}.$$

- Storage: $Stor(\mathbf{V}) = \sum_{\ell=1}^{d} n_\ell \ll \dim \mathbb{V}_{\mathbf{n}} = \prod_{\ell=1}^{d} n_\ell$.

- $O(d)$ bilinear operations: addition, Hadamard product, contraction, convolution, ...

- We often simplify to $n_1 = ... = n_d = n$. Denote by $V^{\otimes d}$ the d-fold tensor product $\mathbb{V}_{\mathbf{n}} = V \otimes ... \otimes V = V^{\otimes d} \equiv \mathbb{R}^{n \times \ldots \times n}$.

Rem. Separation of variables applies to d-variate functions, $f : [-1, 1]^d \mapsto \mathbb{R}$,

$$f(x_1, \ldots, x_d) = f_1(x_1) \ldots f_d(x_d).$$

Vectorization of a tensor

Def. 4.1. For a matrix $A \in \mathbb{R}^{n\times m}$ we use the *vector representation* (vectorization or concatenation) $A \to vec(A) \in \mathbb{R}^{mn}$, where $vec(A)$ is an $nm \times 1$ vector obtained by "stacking" A's columns (the FORTRAN-style ordering)

$$vec(A) := [a_{11}, ..., a_{n1}, a_{12}, ..., a_{n2}, a_{13}, ..., a_{nm}]^T = \begin{bmatrix} [a(i,1)] \\ [a(i,2)] \\ \vdots \\ [a(i,m)] \end{bmatrix} \in \mathbb{R}^{nm\times 1}.$$

$vec(A)$ is a rearranged version of A, implemented by *reshape* in Matlab.

Def. 4.2. The vectorization of a tensor $\mathbf{A} = [a(\mathbf{i})] \in \mathbb{R}^{I_1\times\cdots\times I_d}$ is recursively defined by

$$vec(\mathbf{A}) = \begin{bmatrix} vec([a(i_1, ..., i_{d-1}, 1)]) \\ vec([a(i_1, ..., i_{d-1}, 2)]) \\ \vdots \\ vec([a(i_1, ..., i_{d-1}, n_d)]) \end{bmatrix} \in \mathbb{R}^{|\mathbf{n}|\times 1}.$$

The tensor element $a(i_1, ..., i_d)$ maps to vector entry $(j, 1)$, where

$$j = 1 + \sum_{k=1}^{d} (i_k - 1) \prod_{\ell=1}^{k-1} n_\ell.$$

Application to matrix Lyapunov/Sylvester equations

▶ The matrix Sylvester equation for $X \in \mathbb{R}^{m\times n}$

$$AX + XB^T = G \in \mathbb{R}^{m\times m} \qquad (1)$$

with $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$, can be written in vector form

$$(I_n \otimes A + B \otimes I_m) vec(X) = vec(G).$$

Now the solvability conditions and certain solution methods can be derived. The matrix Sylvester equation (1) is uniquely solvable if (see Thm. 2.7)

$$\lambda_j(A) + \mu_k(B) \neq 0.$$

▶ In the case $A = B$ we arrive at the Lyapunov equation.

Rem. 4.1. Since $I_n \otimes A$ and $B \otimes I_m$ commute, we can apply methods based on *sinc*-quadr., to represent the inverse

$$(I_n \otimes A + B \otimes I_m)^{-1} = \int_{\mathbb{R}_+} e^{-t(I_n\otimes A + B\otimes I_m)} dt \approx \sum_{k=-M}^{M} c_k e^{-t_k A} \otimes e^{-t_k B}.$$

If A and B represent the discrete elliptic operators in $\mathbb{R}^d$, we obtain the low-rank tensor-product approx. to the Sylvester solution operator.

Matrix unfolding of a tensor

▶ *Unfolding of a tensor* into a matrix (matricization) is a way to map high order tensor into two-fold arrays by rearranging (reshaping) it for some $\ell \in \{1, ..., d\}$, $\mathbb{R}^{\mathcal{I}} \mapsto \mathbb{R}^{I_\ell \times I_{(-\ell)}}$, and then vectorizing the tensors in $\mathbb{R}^{i_\ell \times I_{(-\ell)}}$ for each $i_\ell \in I_\ell$.

The single hole index set is defined by $I_{(-\ell)} := I_1 \times ... \times I_{\ell-1} \times I_{\ell+1} \times ... \times I_d$.

Def. The unfolding of a tensor $\mathbf{A} \in \mathbb{R}^{I_1 \times \dots \times I_d}$ w.r.t. the index ℓ (along mode ℓ) is defined by a matrix $mat_\ell(\mathbf{A}) := A_{(\ell)}$ of size $n_\ell \times \bar{n}_\ell$, so that the tensor element $a(i_1, ..., i_d)$ maps to matrix element $a(i_\ell, j)$, $i_\ell \in I_\ell$, where

$$A_{(\ell)} = [a_{i_\ell j}], \text{ with } \quad j \in \{1, \dots, \bar{n}_\ell\}, \ \bar{n}_\ell := n_1 \cdots n_{\ell-1} n_{\ell+1} \cdots n_d,$$

$$j = 1 + \sum_{k=1, k\neq\ell}^{d} (i_k - 1) J_k, \quad J_k = \prod_{m=1, m\neq\ell}^{k-1} n_m.$$

Exer. 4.2. ($mat_\ell(\mathbf{A})$ by recursion over $vec(\mathbf{A})$). Derive the representation

$$mat_\ell(\mathbf{A}) = [vec([a(i_1, ..., i_{\ell-1}, 1, i_{\ell+1}, ..., i_d)]), ..., vec([a(i_1, ..., i_{\ell-1}, n_\ell, i_{\ell+1}, ..., i_d)])]^T .$$

The latter representation can be used as the equivalent definition of $mat_\ell(\mathbf{A})$.

Example of matrix unfolding of a tensor; ℓ-rank of a tensor

Rem. 4.2. Kolmogorow's decomposition is a particular way for unfolding of the multivariate function into "one-dimensional" representation (univariate function).

Example. Define a tensor $\mathbf{A} \in \mathbb{R}^{3\times2\times3}$ by two slices,

$$[a_{i1k}] = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \\ 2 & 0 & 2 \end{bmatrix}, \quad [a_{i2k}] = \begin{bmatrix} 2 & 2 & 0 \\ 2 & -2 & 4 \\ 4 & 0 & 4 \end{bmatrix}.$$

The matrix unfolding along mode 1, $A_{(1)}$, is given by

$$A_{(1)} = \begin{bmatrix} 1 & 1 & 0 & 2 & 2 & 0 \\ 1 & -1 & 2 & 2 & -2 & 4 \\ 2 & 0 & 2 & 4 & 0 & 4 \end{bmatrix}.$$

Def. 4.3. (ℓ-rank of a tensor.) The ℓ-rank of $\mathbf{A}$, $r_\ell = rank_\ell(\mathbf{A})$, is the dimension of the vector space spanned by ℓ-mode vectors (fibers), i.e.

$$r_\ell = rank_\ell(\mathbf{A}) = rank(A_{(\ell)}), \quad \ell = 1, ..., d.$$

Matrix unfolding of a tensor is rank deficient

Matrix unfolding contains the full set of ℓ-fibers (redundancy), which can be represented (approximated) using a few orthogonal vectors in $U^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$, provided that $r_\ell = rank_\ell(\mathbf{A}) < n_\ell$ is small: mode-ℓ dominating subspace $span(U^{(\ell)})$ of columns in $U^{(\ell)}$.

This is the basic idea of the so-called Tucker decomposition of a tensor, based on the ortho-projection onto the TPHS, $\mathbb{T} = \bigotimes_\ell span(U^{(\ell)})$, $(\ell = 1, ..., d)$.

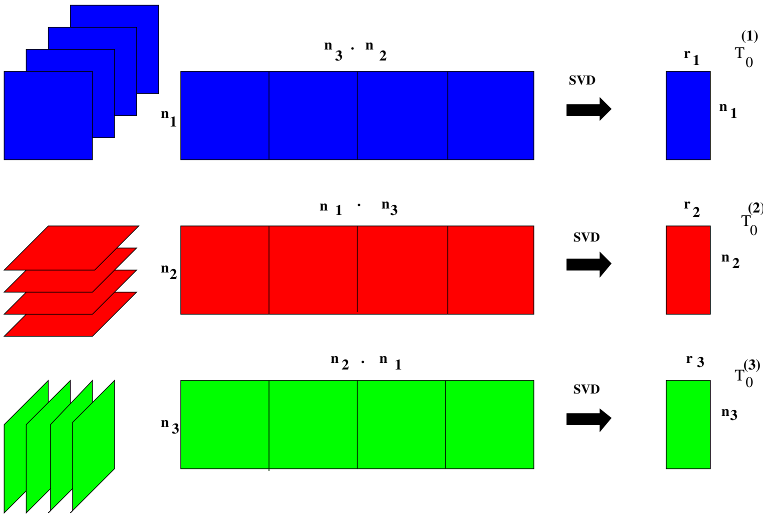


Figure: Visualization of the matrix unfolding for $d = 3$.

Directional tensor ranks and the bilinear tensor operations

- The major difference with the matrix case, however, is the fact that the different ℓ-ranks of a higher-order tensor are not necessarily the same.
- The main tensor-tensor operations:

Scalar product: $\mathbb{V}_\mathbf{n} \times \mathbb{V}_\mathbf{n} \mapsto \mathbb{R}$,

$$\langle \mathbf{W}, \mathbf{V} \rangle = \sum\nolimits_{\mathbf{i}} w_\mathbf{i} v_\mathbf{i}.$$

Hadamard product: $\mathbb{V}_\mathbf{n} \times \mathbb{V}_\mathbf{n} \mapsto \mathbb{V}_\mathbf{n}$,

$$\mathbf{W} \odot \mathbf{V} = \mathbf{Y}, \quad y_\mathbf{i} = w_\mathbf{i} v_\mathbf{i}.$$

Addition: $\mathbb{V}_\mathbf{n} \times \mathbb{V}_\mathbf{n} \mapsto \mathbb{V}_\mathbf{n}$,

$$\mathbf{W} + \mathbf{V} = \mathbf{Y}, \quad y_\mathbf{i} = w_\mathbf{i} + v_\mathbf{i}.$$

Convolution product: $\mathbb{V}_\mathbf{n} \times \mathbb{V}_\mathbf{n} \mapsto \mathbb{V}_{\mathbf{2n-1}}$,

$$\mathbf{W} \star \mathbf{V} = \mathbf{Y}, \quad y_\mathbf{j} = \sum\nolimits_{\mathbf{i}} w_\mathbf{i} v_{\mathbf{j-i+1}}, \quad \mathbf{1} \leq \mathbf{j} - \mathbf{i} + \mathbf{1} \leq \mathbf{n}.$$

- The elements of larger space $\mathbb{V}_{\mathbf{2n-1}}$ can be restricted to $\mathbb{V}_\mathbf{n} \subset \mathbb{V}_{\mathbf{2n-1}}$.

Tensor-matrix contracted product

Def. A *tensor-matrix contracted product* (TMCP) along mode ℓ: Given $\mathbf{V} \in \mathbb{R}^{I_1 \times \ldots \times I_d}$, and $M \in \mathbb{R}^{J_\ell \times I_\ell}$, the mode-$\ell$ TMCP is defined by the tensor

$$\mathbf{U} = \mathbf{V} \times_\ell M \in \mathbb{R}^{I_1 \times \ldots \times I_{\ell-1} \times J_\ell \times I_{\ell+1} \ldots \times I_d},$$

$$u_{i_1,\ldots,i_{\ell-1},j_\ell,i_{\ell+1},\ldots,i_d} = \sum_{i_\ell=1}^{n_\ell} v_{i_1,\ldots,i_{\ell-1},i_\ell,i_{\ell+1},\ldots,i_d} m_{j_\ell,i_\ell}, \quad j_\ell \in J_\ell.$$

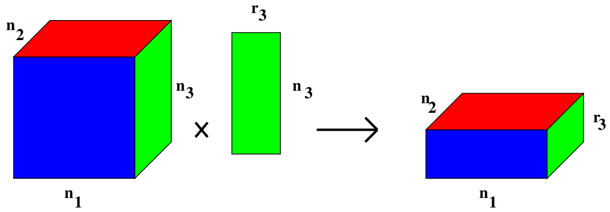


Figure: Contracted product of a third-order tensor with a matrix.

- The TMCP is a generalization of the matrix-matrix multiplication:

$$M_{(n,m)} \times_2 M_{(p,m)} = M_{(n,m)} M_{(p,m)}^T \to M_{(n,p)}.$$

Rank-R matrix decomposition is a prototype of tensor formats

- $d = 2$: rank-R $m \times n$ matrices,

$$V = \sum_{k=1}^{R} u_k v_k^T \equiv UV^T \equiv \overline{U}D\overline{V}^T,$$

$$U = [u_1, \ldots, u_R]^T \in \mathbb{R}^{m \times R} \quad V, \overline{U}, \overline{V} \in \mathbb{R}^{n \times R}, \quad D \in \mathbb{R}^{R \times R}.$$

- Most commonly used tensor formats generalize the rank-R decomposition of matrices:

– **Canonical (CP)** tensor decomposition (multidimensional rank-R representation).

– The **orthogonal Tucker** decomposition – multidimensional analogy of the matrix SVD.

– Matrix product states (**MPS**) – d-term factorization by "univariate factors".

Special cases of MPS: tensor train (**TT**), hierarchical Tucker (**HT**)

– New **range separated (RS)** tensor formats:
sum of localized CP tensors combined with the global CP/Tucker/TT decomposition.

RS formats provide the efficient tool for approximation of many-particle systems.

A sum of rank-1 elements: canonical format, pro and contra

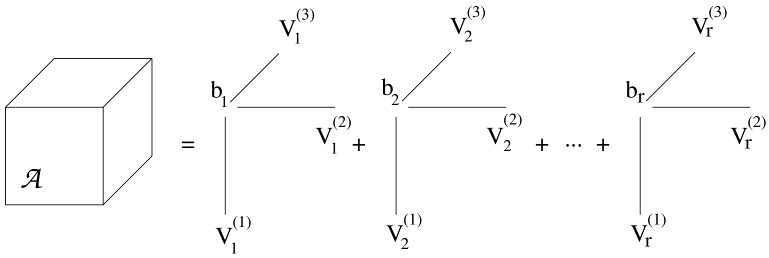
Def. Canonical R-term representation: $\mathbf{V} \in \mathcal{C}_R(\mathbb{V}_\mathbf{n})$, if [Hitchcock '27, ...]

$$\mathbf{V} = \sum_{k=1}^{R} \mathbf{v}_k^{(1)} \otimes \ldots \otimes \mathbf{v}_k^{(d)}, \quad \mathbf{v}_k^{(\ell)} \in V_\ell.$$

- Advantages: Storage $= dRN$, simple multilinear algebra.

Analytic methods of low-rank approximation for Green's kernels, etc.

- The minimal R is called a tensor rank of $\mathbf{V}$, $R = rank(\mathbf{V})$.
- $V^{(\ell)} = [\mathbf{v}_1^{(\ell)}, \ldots, \mathbf{v}_R^{(\ell)}] \in \mathbb{R}^{n_\ell \times R}$ is called a side matrix.



- Limitations: $\mathcal{C}_R(\mathbb{V}_\mathbf{n})$ is the non-closed set $\Rightarrow$ lack of stable approximation methods.

Exact rank-R representation is N-P hard.

Example. $f(x) = x_1 + \ldots + x_d$. $rank_{Can}(f) = d$? But it can be approximated by rank-2 element

$$f(x) = \lim_{\varepsilon \to 0} \frac{\prod_{\ell=1}^{d}(1+\varepsilon x_\ell) - 1}{\varepsilon}.$$

Little analogy between the cases $d = 2$ and $d \geq 3$

If $d > 2$, the situation changes dramatically.

I. Set of tensors of rank not larder than r,

$$\widehat{\mathcal{C}}_r(d) := \{\mathbf{T} \in V_1 \otimes \ldots \otimes V_d : \ rank(\mathbf{T}) \leq r\},$$

is closed when $d = 2$ (matrices) [Minsky], or if $r = 1$ (rank-1 tensors), [Golub, Zhang '01].

II. For $d \geq 3$, $r \neq 1$, the set $\widehat{\mathcal{C}}_r(d)$ is non-closed [Book: Ladsberger '12].

Example. $f(x) = x_1 + \ldots + x_d$. $rank_{Can}(f) = d$? It can be approximated by rank-2 elem.

$$f(x) = \lim_{\varepsilon \to 0} \frac{\prod_{\ell=1}^{d}(1+\varepsilon x_\ell) - 1}{\varepsilon}.$$

Exer. 2.2. Let x, y be two linearly independent vectors in V (say, $dim(V) = 2$). Consider the tensor $\mathbf{T} \in V \otimes V \otimes V = V^{\otimes 3}$,

$$\mathbf{T} := x \otimes x \otimes x + x \otimes y \otimes y + y \otimes x \otimes y.$$

Prove that: (a) $rank(\mathbf{T}) = 3$; (b) $\mathbf{T}$ has no best rank-2 approximation.

Comments to the proof of Exercise 2.2

Proof of (a): By contradiction to the condition $dim(V) = 2$.

Proof of (b): Consider a sequence $\{\mathbf{S}_k\}_{k=1}^{\infty}$ in $V^{\otimes 3}$,

$$\mathbf{S}_k := x \otimes x \otimes (x - ky) + (x + \frac{1}{k}y) \otimes (x + \frac{1}{k}y) \otimes ky.$$

Clearly that $rank(\mathbf{S}_k) \leq 2$ for all k. By multi-linearity of $\otimes$,

$$\mathbf{S}_k = \mathbf{T} + \frac{1}{k} y \otimes y \otimes y.$$

Hence, for any choice of norm on $V \otimes V \otimes V$,

$$\|\mathbf{S}_k - \mathbf{T}\| = \frac{1}{k}\|y \otimes y \otimes y\| \to 0 \quad as\ k \to \infty.$$

III. For $d \geq 3$ we do not know any finite algorithm to compute $r = rank(\mathbf{A})$, except simple bounds:

$$0 \leq rank(\mathbf{A}) \leq n^{d-1}.$$

Compare with the matrix case $d = 2$.

Little analogy between the cases $d = 2$ and $d \geq 3$

IV. For fixed $d \geq 3$ and n we do not know the exact value of $\max\{rank(\mathbf{A})\}$.

J. Kruskal '77 proved that:

– for any $2 \times 2 \times 2$ tensor we have $\max\{rank(\mathbf{A})\} = 3 < 4$;

– for $3 \times 3 \times 3$ tensors there holds $\max\{rank(\mathbf{A})\} = 5 < 9$.

V. "Probabilistic" properties of $rank(\mathbf{A})$:
In the set of $2 \times 2 \times 2$ tensors there is about 79% of rank-2 tensors and 21% of rank-3 tensors, while rank-1 tensors appear with probability 0, (J. Kruskal).

Clearly, for $n \times n$ matrices we have (why?)

$$\mathcal{P}\{rank(A) = n\} = 1.$$

VI. $rank(\mathbf{A})$ depends on the number field (say, $\mathbb{R}$ or $\mathbb{C}$): $rank_{\mathbb{C}}(\sin(x_1 + ... + x_d)) = 2$.

VII. Prop. Multiplicity of rank decompositions: The trigonometric identity

$$d \geq 2: \quad f(x) := \sin(\sum_{j=1}^{d} x_j) = \sum_{j=1}^{d} \sin(x_j) \prod_{k \in \{1,...,d\} \setminus \{j\}} \frac{\sin(x_k + \alpha_k - \alpha_j)}{\sin(\alpha_k - \alpha_j)}$$

holds $\forall\ \alpha_k \in \mathbb{R}$, s.t. $\sin(\alpha_k - \alpha_j) \neq 0$ for all $j \neq k$. [Mohlenkamp, Monzón '05; Dolgov, Khoromskij, Savostianov '12]

Strassen algorithm of fast matrix-matrix multiplication

Finding the tensor rank can be a useful concept even in the classical linear algebra.

Historical remarks on the Strassen algorithm of fast matrix-matrix multiplication in $O(n^{\log_2 7})$ complexity.

$O(n^{2+\varepsilon})$ algorithm to multiply two $n \times n$ matrices provides $O(n^{2+\varepsilon})$ method for solving system of n linear equations. [Strassen 1969].

Best known result: $O(n^{2.376})$ [Copperesmith-Winograd 1987].

Lloyd Nick Trefethen bets Peter Alfred (25 June 1985) that a method will have been found to solve

$$Ax = b$$

in $O(n^{2+\varepsilon})$ operations for any $\varepsilon > 0$ (numerical stability is not an issue).

Details at personal homepage by Prof. L.N. Trefethen (Uni. Oxford).

Many related results can be found in the monograph by J.M. Landsberg (2012).

Strassen algorithm via rank decomposition

In the block form

$$\begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \cdot \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$$

with

$$C_k = \sum_{i=1}^{4} \sum_{j=1}^{4} \gamma_{ijk} A_i B_j, \quad k = 1, ..., 4,$$

where for the 3-rd order coefficients tensor $[\gamma_{ijk}]$ of size $4 \times 4 \times 4$ we have (slice-wise)

$$[\gamma_{ijk}] = \lhd_1 \begin{bmatrix} 1000 \\ 0010 \\ 0000 \\ 0000 \end{bmatrix} \lhd_2 \begin{bmatrix} 0100 \\ 0001 \\ 0000 \\ 0000 \end{bmatrix} \lhd_3 \begin{bmatrix} 0000 \\ 0000 \\ 1000 \\ 0010 \end{bmatrix} \lhd_4 \begin{bmatrix} 0000 \\ 0000 \\ 0100 \\ 0001 \end{bmatrix}.$$

Here $\lhd_i$ means that the related matrix corresponds to the slice number $i \leq 4$.

Strassen algorithm via rank decomposition

Suppose that we have rank-R expansion

$$\gamma_{ijk} = \sum_{t=1}^{R} u_{it} v_{jt} w_{kt}.$$

Then

$$C_k = \sum_{t=1}^{R} w_{kt} \sum_{i=1}^{4} \sum_{j=1}^{4} u_{it} A_i v_{jt} B_j = \sum_{t=1}^{R} w_{kt} \left(\sum_{i=1}^{4} u_{it} A_i \right) \left(\sum_{j=1}^{4} v_{jt} B_j \right).$$

Precompute $\Sigma_t = \sum_{i=1}^{4} u_{it} A_i$, $\Delta_t = \sum_{j=1}^{4} v_{jt} B_j$, $t = 1, \ldots, R$, and reduce the initial task to R matrix-matrix products of size $n/2 \times n/2$.

We have $R \leq 8$ (**why ?**), but there are representations (infinitely many) of rank 7: (Strassen's result) $\mapsto O(n^{\log_2 7})$. Indeed, let $n = 2^p$, then

$$n^3 \mapsto R 8^{-1} n^3 \mapsto \ldots \mapsto R^p 8^{-p} n^3 = 2^{p \log_2 R} = n^{\log_2 R}.$$

Open problem: Is it possible to construct rank decompositions with $R < 7$?
If yes, then the Strassen's result can be improved.

Orthogonal Tucker model

Def. Rank $\mathbf{r} = [r_1, \ldots, r_d]$ Tucker tensors: $\mathbf{U} \in \mathcal{T}_{\mathbf{r}}(\mathbb{V}_{\mathbf{n}})$ if [Tucker '66, De Lathawer et. al. '2000]

$$\mathbf{U} = \sum_{k_1,\ldots,k_d=1}^{\mathbf{r}} b_{k_1 \ldots k_d} \mathbf{u}_{k_1}^{(1)} \otimes \ldots \otimes \mathbf{u}_{k_d}^{(d)} \in T_1 \otimes \ldots \otimes T_d, \quad T_\ell = span\{\mathbf{u}_{k_\ell}^{(\ell)}\}_{k_\ell=1}^{r_\ell} \subset \mathbb{R}^{n_\ell}.$$

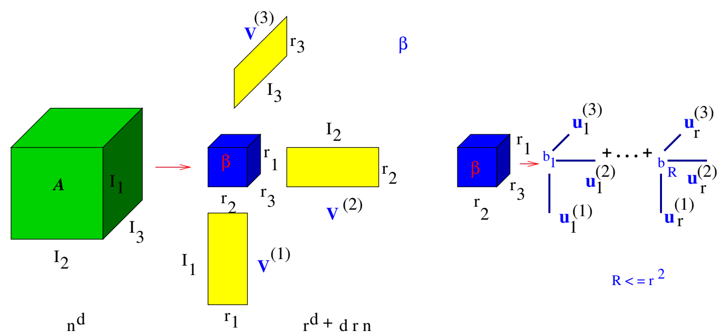
$\mathbf{B} = [b_{\mathbf{k}}] \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ is the core tensor. Using contracted product notation

$$\mathbf{U} = \mathbf{B} \times_1 U^{(1)} \times_2 \ldots \times_d U^{(d)}, \text{ with side matrices } U^{(\ell)} = [\mathbf{u}_1^{(\ell)}, \ldots, \mathbf{u}_{r_\ell}^{(\ell)}] \in \mathbb{R}^{n_\ell \times r_\ell}.$$

- $d = 2$: SVD of a rank-r matrix, $A = UDV^T = D \times_1 U \times_2 V$, $U \in \mathbb{R}^{n \times r}$, $D \in \mathbb{R}^{r \times r}$.
- Storage: $drN + r^d$, $r = \max r_\ell \ll N$ (efficient if $d = 3$, e.g. the Hartree-Fock equation).

Def. Mixed Tucker-canonical tensors. [Khoromskij, Khoromskaia '08] Storage: $drN + dRr$.

$$\mathcal{T}_{\mathcal{C}_R, \mathbf{r}} := \{\mathbf{U} \in \mathcal{T}_{\mathbf{r}}(\mathbb{V}_{\mathbf{n}}) : \mathbf{B} = [b_{\mathbf{k}}] \in \mathcal{C}_R(\mathbb{R}^{r_1 \times \cdots \times r_d})\}.$$



Summary: Additive tensor formats in the index (coordinate) representation

- **Canonical tensors, V** $\in \mathcal{C}_R = \mathcal{C}_R(\mathbb{V}_{\mathbf{n}})$. Storage: dRN.

$$v(i_1, \ldots, i_d) = \sum\nolimits_{k=1}^{R} v_k^{(1)}(i_1) \ldots v_k^{(d)}(i_d), \quad \mathbf{v}_k^{(\ell)} \in \mathbb{R}^{n_k}.$$

- **Tucker tensors, U** $\in \mathcal{T}_{\mathbf{r}} = \mathcal{T}_{\mathbf{r}}(\mathbb{V}_{\mathbf{n}})$. Storage: $drN + r^d$, $r = \max r_\ell$.

$$u(i_1, \ldots, i_d) = \sum_{k_1,\ldots,k_d=1}^{\mathbf{r}} b_{k_1,\ldots,k_d} u_k^{(1)}(i_1) \ldots u_k^{(d)}(i_d).$$

The Tucker representation is unique up to rotation in mode ℓ via orthogonal matrix $S^{(\ell)}$,

$$U^{(\ell)} \to S^{(\ell)} U^{(\ell)}, \quad \mathbf{B} \to \mathbf{B} \times_1 (S^{(1)})^T \times_2 \ldots. \times_d (S^{(d)})^T.$$

Exer. Prove: maximal *canonical rank* of the Tucker tensor is $R = (\prod_{\ell=1}^{d} r_\ell)/\max_\ell r_\ell$.

- **Mixed Tucker-canonical tensors, U** $\in \mathcal{T}_{\mathcal{C}_R,\mathbf{r}} \subset \mathcal{T}_{\mathbf{r}}(\mathbb{V}_{\mathbf{n}})$, $\mathbf{B} \in \mathcal{C}_R$. Storage: $drN + dRr$.

$$\mathbf{B} \in \mathcal{C}_R(\mathbb{R}^{r_1 \times \ldots \times r_d}) : \qquad b_{k_1,\ldots,k_d} = \sum_{\gamma=1}^{R} b_\gamma^{(1)}(k_1) \ldots b_\gamma^{(d)}(k_d).$$

- Canonical and Tucker tensor formats are useful in electronic structure calculations as well as in the traditional applications in data processing.

The continuous prototype of Tucker approximation

Remark. The tensor product polynomial interpolation of order $\mathbf{r}$ to continuous function in $\mathbb{H} = L^2(I^d)$ provides the Tucker approximation (functional Tucker) by

$$f(x_1, \ldots, x_d) \approx \sum_{\mathbf{j}=1}^{\mathbf{r}} f(\nu_{j_1}, \ldots, \nu_{j_d}) \prod_{\ell=1}^{d} L_{j_\ell}(x_\ell).$$

L_{j_ℓ} is a set of the Lagrange polynomials on $[-1, 1]$ at, say, Chebyshev-Gauss-Lobatto grid, $\nu_{j_\ell} = \cos \frac{\pi j_\ell}{N}$, $j_\ell = 0, \ldots, r_\ell$. The core tensor is given by $[f(\nu_{j_1}, \ldots, \nu_{j_d})]$.

Example. Rank-1 elements: $f = \exp(\sum_{\ell=1}^{d} g_\ell(x_\ell)) = \prod_{\ell=1}^{d} \exp(g_\ell(x_\ell))$.
Example. $f(x) = \sum_{j=1}^{d} x_j$: $rank_{Tuck}(f) = 2$.

Example. $f = \sin\left(\sum_{j=1}^{d} x_j\right)$: $rank_{Tuck}(f) = 2$ since $rank_{CP}(f) = 2$ holds over the field $\mathbb{C}$,

$$\sin(\sum_{j=1}^{d} x_j) == Im(e^{i \sum_{j=1}^{d} x_j}) = (e^{i \sum_{j=1}^{d} x_j} - e^{-i \sum_{j=1}^{d} x_j})/2i.$$

Exer. Compute the canonical, Tucker and ℓ-mode ε-rank of the Hilbert tensor $\mathbf{A} = [a_{ijk}]$, $a_{ijk} = 1/(i + j + k)$, $(i, j, k = 1, \ldots, n)$ with $n = 10^2$, up to approximation error $\varepsilon = 10^{-3}, 10^{-4}, 10^{-5}$. Exponential convergence in r_ε? (Hint: Sinc quadrature).

Nonlinear approximation in tensor formats

Probl. 1. Efficient MLA in tensor classes $\mathcal{S}$ getting rid of the curse of dimensionality.

Probl. 2. Best rank-structured approximation of a high-order tensor $\mathbf{X} \in \mathbb{V}_{\mathbf{n}}$ in the fixed nonlinear set (manifold) $\mathcal{S} \subset \{\mathcal{T}_{\mathbf{r}}, \mathcal{C}_R, \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}}\}$.

Probl. 3. For fixed accuracy $\varepsilon > 0$, efficient approximation in $\mathcal{S}$ of a high-order tensor $\mathbf{X} \in \mathbb{V}_{\mathbf{n}}$ with adaptively chosen rank parameter.

Since both $\mathcal{T}_{\mathbf{r}}$ and $\mathcal{C}_R$ are not linear spaces, we arrive at a nontrivial nonlinear approximation problem on estimation:

- Given $\mathbf{X} \in \mathbb{V}_{\mathbf{n}}$ (more generally, $\mathbf{X} \in \mathcal{S}_0 \subset \mathbb{V}_{\mathbf{n}}$), find

$$T_{\mathbf{r}}(\mathbf{X}) := \operatorname*{argmin}_{\mathbf{A} \in \mathcal{S}} \|\mathbf{X} - \mathbf{A}\|, \quad \text{where} \quad \mathcal{S} \subset \{\mathcal{T}_{\mathbf{r}}, \mathcal{C}_R, \boldsymbol{\mathcal{T}}_{\mathcal{C}_{R,\mathbf{r}}}\}. \tag{2}$$

Principal question: **How to solve the approximation problem (2) efficiently?**

- For algebraic tensors: HOSVD, ALS-type iteration, SVD-based rank reduction, Greedy.
- For function generated tensors: polynomial or *sinc* approximation + rank reduction.

Higher order SVD (HOSVD) approximation in Tucker format

Given tensor $\mathbf{A}$, let $U^{(\ell)} \in \mathbb{R}^{n_\ell \times R_\ell}$ be left singular matrix of its matrix unfolding,

$$A_{(\ell)} = U^{(\ell)} D_\ell V^{(\ell)^T}, \quad \ell = 1, ..., d.$$

Def. (HOSVD). [De Lathauwer, et al., '00]. The HOSVD of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{B} \times_1 U^{(1)} \times_2 U^{(2)} ... \times_d U^{(d)},$$

where the ℓ-mode rank of $\mathbf{A}$ is equal to R_ℓ ($\ell = 1, ..., d$), and

$$\mathbf{B} = \mathbf{A} \times_1 U^{(1)^T} \times_2 U^{(2)^T} ... \times_d U^{(d)^T}.$$

Thm. 4.1 (Truncated HOSVD). [De Lathauwer, et al., '00]. For given rank parameter $\mathbf{r} = (r_1, ..., r_d)$, define a tensor $\tilde{\mathbf{A}}_{\mathbf{r}}$ by discarding the smallest ℓ-mode singular values $\sigma^{(\ell)}_{r_\ell+1}, \sigma^{(\ell)}_{r_\ell+2}, ..., \sigma^{(\ell)}_{R_\ell}$ ($\ell = 1, ..., d$), i.e., set the entries of $\mathbf{B}$ with indexes larger than r_ℓ equal to zero.
The approximation error is estimated by

$$\|\mathbf{A} - \tilde{\mathbf{A}}_{\mathbf{r}}\|^2 \leq \sum_{\ell=1}^{d} \varepsilon_\ell^2,$$

where we denote

$$\varepsilon_\ell^2 = \sum_{i_\ell = r_\ell + 1}^{R_\ell} (\sigma^{(\ell)}_{i_\ell})^2.$$

Proof of Thm. 4.1 and useful properties of the Tucker tensors

Proof. Let $U_r^{(\ell)} \in \mathbb{R}^{n_\ell \times r_\ell}$ be the left singular matrix in the truncated SVD of $A_{(\ell)}$. Introduce the orthogonal projection operator

$$P_\ell = I \otimes ... \otimes I \otimes (U_r^{(\ell)} U_r^{(\ell)^T}) \otimes I \otimes ... \otimes I,$$

then the rank-$\mathbf{r}$ Tucker approximation can be represented as $\tilde{\mathbf{A}}_\mathbf{r} = P_1 ... P_d \mathbf{A}$. Using the orthogonality we complete the proof

$$\begin{aligned} \|\mathbf{A} - \tilde{\mathbf{A}}_\mathbf{r}\|^2 &= \|(\mathbf{A} - P_1\mathbf{A}) + (P_1\mathbf{A} - P_1P_2\mathbf{A}) + ... + (P_1...P_{d-1}\mathbf{A} - P_1...P_d\mathbf{A})\| \\ &\le \|\mathbf{A} - P_1\mathbf{A}\|^2 + \|\mathbf{A} - P_2\mathbf{A}\|^2 + ... + \|\mathbf{A} - P_d\mathbf{A}\|^2 \\ &\le \varepsilon_1^2 + ... + \varepsilon_d^2. \end{aligned}$$

Other useful properties.

- Best rank-$(r_1, ..., r_d)$ approximation $\mathbf{A}_\mathbf{r}^*$ exists.

Rem. Denote $\varepsilon = \|\mathbf{A} - \mathbf{A}_\mathbf{r}^*\|$. Truncated HOSVD gets lost only the factor $\sqrt{d}$ compared with "best rank-$(r_1, ..., r_d)$ approximation" since $\varepsilon_\ell \le \varepsilon$ (interesting point !).

- $\|\mathbf{A}\| = \|\mathbf{B}\|$.
- For the canonical rank of the Tucker tensor $R \le \prod_{\ell=1}^d r_\ell / \max\{r_\ell\}$.

Reduced HOSVD (RHOSVD) approximation in canonical format

Rank reduction in the canonical format:

Canonical (Reduced HOSVD) $\mapsto$ Tucker $\mapsto$ canonical (ALS) [Khoromskij, Khoromskaia, SISC '08; '14].

For given $\mathbf{A} \in \mathcal{C}_{R,\mathbf{n}}$, $(n \le R)$, in the rank-R canonical format,

$$\mathbf{A} = \sum\nolimits_{\nu=1}^{R} \xi_\nu \mathbf{u}_\nu^{(1)} \otimes \ldots \otimes \mathbf{u}_\nu^{(d)}, \quad \xi_\nu \in \mathbb{R}, \qquad (3)$$

use its contracted product representation (Tucker with $\mathbf{r} = (R, ..., R)$)

$$\mathbf{A} = \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d U^{(d)}, \quad \boldsymbol{\xi} = diag\{\xi_1, ..., \xi_R\} \in (\mathbb{R}^R)^{\otimes d},$$

via ℓ-mode side matrices $U^{(\ell)} = [\mathbf{u}_1^{(\ell)} ... \mathbf{u}_R^{(\ell)}] \in \mathbb{R}^{n \times R}$ $(\ell = 1, ..., d)$.

Def. (RHOSVD). For given $\mathbf{r}$, let $W^{(\ell)} := Z_0^{(\ell)} D_{\ell,0} V_0^{(\ell)^T} \approx U^{(\ell)}$, be the truncated SVD of the side matrix $U^{(\ell)}$, $(\ell = 1, ..., d)$, where $D_{\ell,0} = \text{diag}\{\sigma_{\ell,1}, \sigma_{\ell,2}, ..., \sigma_{\ell,r_\ell}\}$ and $Z_0^{(\ell)} = [\mathbf{z}_1^{(\ell)}, ..., \mathbf{z}_{r_\ell}^{(\ell)}] \in \mathbb{R}^{n \times r_\ell}$, $V_0^{(\ell)} \in \mathbb{R}^{R \times r_\ell}$, represent the respective orthogonal factors. The RHOSVD approximation of $\mathbf{A}$ in (3) is given by the Tucker tensor

$$\mathbf{A}_{(\mathbf{r})}^0 = \boldsymbol{\xi} \times_1 \left[Z_0^{(1)} D_{1,0} V_0^{(1)^T} \right] \times_2 \cdots \times_d \left[Z_0^{(d)} D_{d,0} V_0^{(d)^T} \right].$$

Error bound for RHOSVD approximation

Thm. (RHOSVD) (a) Given $\mathbf{A} \in \mathcal{C}_{R,\mathbf{n}}$ in (3) and $\mathbf{r} = (r_1, \ldots, r_d)$, the minimization probl.

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

is equivalent to the *dual maximization problem* on the Grassman/Stifel manifold $\mathbb{S}_{r_\ell}$,

$$[Z^{(1)}, ..., Z^{(d)}] = \mathrm{argmax}_{W^{(\ell)} \in \mathcal{G}_\ell[\mathbb{S}_{r_\ell}]} \left\| \sum_{\nu=1}^{R} \xi_\nu \left({W^{(1)}}^T \mathbf{u}_\nu^{(1)} \right) \otimes ... \otimes \left({W^{(d)}}^T \mathbf{u}_\nu^{(d)} \right) \right\|^2 .$$

(b) (Error of RHOSVD). Let $\sigma_{\ell,1} \geq \sigma_{\ell,2} ... \geq \sigma_{\ell,\min(n,R)}$ be the singular values of $U^{(\ell)} \in \mathbb{R}^{n \times R}$ $(\ell = 1, ..., d)$. Then the RHOSVD approx. $\mathbf{A}^0_{(\mathbf{r})}$, exhibits the error bound,

$$\|\mathbf{A} - \mathbf{A}_{(\mathbf{r})}\| \leq \|\mathbf{A} - \mathbf{A}^0_{(\mathbf{r})}\| \leq \|\boldsymbol{\xi}\| \sum_{\ell=1}^{d} \Big(\sum_{k=r_\ell+1}^{\min(n,R)} \sigma^2_{\ell,k} \Big)^{1/2}, \quad \|\boldsymbol{\xi}\| = \sqrt{\sum\nolimits_{\nu=1}^{R} \xi_\nu^2}.$$

(c) Let the canonical decomposition satisfy the stability condition $\sum\limits_{\nu=1}^{R} \xi_\nu^2 \leq C \|\mathbf{A}\|^2$, then the quasi-optimal RHOSVD approximation is robust in the relative norm

$$\|\mathbf{A} - \mathbf{A}^0_{(\mathbf{r})}\| / \|\mathbf{A}\| \leq C \sum_{\ell=1}^{d} \Big(\sum_{k=r_\ell+1}^{\min(n,R)} \sigma^2_{\ell,k} \Big)^{1/2}.$$

Sketching the proof

Proof. Using the contracted product representations of $\mathbf{A} \in \mathcal{C}_{R,\mathbf{n}}$ and $\mathbf{A}^0_{(\mathbf{r})}$, leads to the following expansion for the approximation error,

$$\begin{aligned}
\mathbf{A} - \mathbf{A}^0_{(\mathbf{r})} &= \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d U^{(d)} \\
&- \boldsymbol{\xi} \times_1 \left[Z_0^{(1)} D_{1,0} {V_0^{(1)}}^T \right] \times_2 \left[Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T \right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T \right] \\
&= \boldsymbol{\xi} \times_1 \left[U^{(1)} - Z_0^{(1)} D_{1,0} {V_0^{(1)}}^T \right] \times_2 \left[Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T \right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T \right] \\
&+ \boldsymbol{\xi} \times_1 U^{(1)} \times_2 \left[U^{(2)} - Z_0^{(2)} D_{2,0} {V_0^{(2)}}^T \right] \cdots \times_d \left[Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T \right] + ... \\
&+ \boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_{d-1} U^{(d-1)} \times_d \left[U^{(d)} - Z_0^{(d)} D_{d,0} {V_0^{(d)}}^T \right].
\end{aligned}$$

Assume that $n \leq R$ and introduce the ℓ-mode residual matrix

$$\Delta^{(\ell)} = U^{(\ell)} - Z_0^{(\ell)} D_{\ell,0} {V_0^{(\ell)}}^T, \quad \{\Delta^{(\ell)}\}_\nu = \sum_{k=r_\ell+1}^{n} \sigma_{\ell,k} z_k^{(\ell)} v_{k,\nu}^{(\ell)}, \quad \nu = 1, ..., R,$$

with notations (and same for $U^{(\ell)}$)

$$V_0^{(\ell)} = [\mathbf{v}_1^{(\ell)}, ..., \mathbf{v}_{r_\ell}^{(\ell)}]^T \in \mathbb{R}^{R \times r_\ell}, \quad \mathbf{v}_k^{(\ell)} = \{v_{k,\nu}^{(\ell)}\}_{\nu=1}^{R} \in \mathbb{R}^R.$$

Sketching the proof

The ℓth summand on the righ-hand side in $\mathbf{A} - \mathbf{A}^0_{(\mathbf{r})}$ takes the form

$$\mathbf{B}_\ell = \boldsymbol{\xi} \times_1 U^{(1)} \cdots \times_{\ell-1} U^{(\ell-1)} \times_\ell \Delta^{(\ell)} \times_{\ell+1} W^{(\ell+1)} \cdots \times_d W^{(d)}.$$

This leads to the error bound (by the triangle inequality)

$$\begin{aligned} \|\mathbf{A} - \mathbf{A}^0_{(\mathbf{r})}\| \leq \sum_{\ell=1}^{d} \|\mathbf{B}_\ell\| &= \|\boldsymbol{\xi} \times_1 \Delta^{(1)} \times_2 W^{(2)} \cdots \times_d W^{(d)}\| \\ &+ \|\boldsymbol{\xi} \times_1 U^{(1)} \times_2 \Delta^{(2)} \cdots \times_d W^{(d)}\| + \ldots \\ &+ \|\boldsymbol{\xi} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_d \Delta^{(d)}\|, \end{aligned}$$

where the ℓth term $\mathbf{B}_\ell$ is represented by

$$\sum_{\nu=1}^{R} \xi_\nu \left[\mathbf{u}_\nu^{(1)} \cdots \times_{\ell-1} \mathbf{u}_\nu^{(\ell-1)} \times_\ell \{\Delta^{(\ell)}\}_\nu \times_{\ell+1} \sum_{k=1}^{r_{\ell+1}} \sigma_{\ell+1,k} \mathbf{z}_k^{(\ell+1)} v_{k,\nu}^{(\ell+1)} \cdots \times_d \sum_{k=1}^{r_d} \sigma_{d,k} \mathbf{z}_k^{(d)} v_{k,\nu}^{(d)} \right],$$

providing the estimate (in view of $\|\mathbf{u}_\nu^{(\ell)}\| = 1$, $\ell = 1, ..., d$, $\nu = 1, ..., R$)

$$\|\mathbf{B}_\ell\| \leq \sum_{\nu=1}^{R} |\xi_\nu| (\sum_{k=r_\ell+1}^{n} \sigma_{\ell,k}^2 (v_{k,\nu}^{(\ell)})^2)^{1/2} \cdot (\sum_{k=1}^{r_{\ell+1}} \sigma_{\ell+1,k}^2 (v_{k,\nu}^{(\ell+1)})^2)^{1/2} \cdots (\sum_{k=1}^{r_d} \sigma_{d,k}^2 (v_{k,\nu}^{(d)})^2)^{1/2}.$$

Finalizing the proof

Recall that $U^{(\ell)}$ ($\ell = 1, ..., d$) has normalized columns, i.e.,

$$1 = \|\mathbf{u}_\nu^{(\ell)}\| = \| \sum_{k=1}^{n} \sigma_{\ell,k} z_k^{(\ell)} v_{k,\nu}^{(\ell)} \|,$$

implying $\sum_{k=1}^{n} \sigma_{\ell,k}^2 (v_{k,\nu}^{(\ell)})^2 = 1$ for $\ell = 1, ..., d$ $\nu = 1, ..., R$.
We finalize the error bound as follows,

$$\begin{aligned} \|\mathbf{A} - \mathbf{A}^0_{(\mathbf{r})}\| &\leq \sum_{\ell=1}^{d} \sum_{\nu=1}^{R} |\xi_\nu| \left(\sum_{k=r_\ell+1}^{n} \sigma_{\ell,k}^2 (v_{k,\nu}^{(\ell)})^2 \right)^{1/2} \\ &\leq \sum_{\ell=1}^{d} \left(\sum_{\nu=1}^{R} \xi_\nu^2 \right)^{1/2} \left(\sum_{\nu=1}^{R} \sum_{k=r_\ell+1}^{n} \sigma_{\ell,k}^2 (v_{k,\nu}^{(\ell)})^2 \right)^{1/2} \\ &= \sum_{\ell=1}^{d} \|\boldsymbol{\xi}\| \left(\sum_{k=r_\ell+1}^{n} \sigma_{\ell,k}^2 \sum_{\nu=1}^{R} (v_{k,\nu}^{(\ell)})^2 \right)^{1/2} = \|\boldsymbol{\xi}\| \sum_{\ell=1}^{d} \left(\sum_{k=r_\ell+1}^{n} \sigma_{\ell,k}^2 \right)^{1/2}. \end{aligned}$$

The case $R < n$ can be analyzed along the same line.

(c) A simple consequence of the general error estimate.

Discussion of RHOSVD

Remark. Extra factor $\|\boldsymbol{\xi}\|$, compared to HOSVD, is relaxed in many applications.

▶ The stability condition (c) in Thm. (RHOSVD) is fulfilled if [Khoromskaia, Khoromskij, NLAA '15]

(A) All canonical vectors in (3) are non-negative that is the case for *sinc*-quadrature based approximations to Green's kernels based on integral Laplace transforms, since in this case all skeleton vectors are non-negative and $a_k > 0$.

(B) The partial orthogonality of the canonical vectors holds, i.e. rank-1 tensors $\mathbf{a}_\nu^{(1)} \otimes \ldots \otimes \mathbf{a}_\nu^{(d)}$ $(\nu = 1, ..., R)$ are mutually orthogonal.

▶ Numerical examples in Matlab (in Lect. 6).

Exer. 4.7. A grid-based tensor related to $f(x) = x_1 + ... + x_d$ has the Tucker rank$= 2$.

▶ The Tucker tensor $\mathbf{A}^0_{(\mathbf{r})}$ can be approximated by the low-rank canoncal tensor by applying the ALS iteration to the small-size Tucker core.

▶ RHOSVD shows the simple way (!) to the transform of a canonical tensor with large canonical rank R to the low-rank MPS (TT) tensor. Discussion in Lectures 7 - 8 - 9.

Multilinear operations in canonical/Tuckers tensor formats

Multilinear tensor operations are reduced to operation on 1D vectors.

Given tensors $\mathbf{A}_1$, $\mathbf{A}_2$ in the canonical format

$$\mathbf{A}_1 = \sum_{k=1}^{R_1} c_k \mathbf{u}_k^{(1)} \otimes \ldots \otimes \mathbf{u}_k^{(d)}, \quad \mathbf{A}_2 = \sum_{m=1}^{R_2} b_m \mathbf{v}_m^{(1)} \otimes \ldots \otimes \mathbf{v}_m^{(d)}.$$

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

$$\langle \mathbf{A}_1, \mathbf{A}_2 \rangle := \sum_{k=1}^{R_1} \sum_{m=1}^{R_2} c_k b_m \prod_{\ell=1}^{d} \left\langle \mathbf{u}_k^{(\ell)}, \mathbf{v}_m^{(\ell)} \right\rangle.$$

Def. Given tensors $\mathbf{F} = [f_\mathbf{i}] \in \mathbb{R}^\mathcal{I}, \mathbf{G} = [g_\mathbf{i}] \in \mathbb{R}^\mathcal{I}$, define their discrete *convolution product* by

$$\mathbf{F} * \mathbf{G} := \left[\sum_{\mathbf{i} \in \mathcal{I}} f_\mathbf{i} g_{\mathbf{j}-\mathbf{i}+1} \right]_{\mathbf{j} \in \mathcal{J}}, \quad \mathcal{J} := \{1, ..., 2n-1\}^d,$$

where $\mathbf{j} - \mathbf{i} + 1 \in \mathcal{I}$ ($\mathbf{G}$ can be extended to the index set beyond $\mathcal{I}$ by zeros).

Convolution, for $d = 3$, complexity $O(R_1R_2 n \log n) \ll n^3 \log n$, (for 3D FFT),

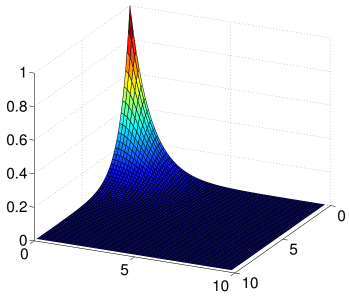
$$\mathbf{A}_1 * \mathbf{A}_2 = \sum_{k=1}^{R_1} \sum_{m=1}^{R_2} c_k b_m (\mathbf{u}_m^{(1)} * \mathbf{v}_k^{(1)}) \otimes (\mathbf{u}_m^{(2)} * \mathbf{v}_k^{(2)}) \otimes (\mathbf{u}_m^{(3)} * \mathbf{v}_k^{(3)}).$$

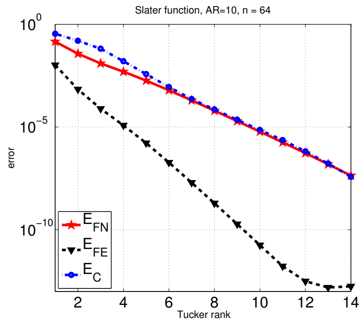
Low-rank approximations in canonical and Tucker formats

▶ Beginning of tensor numerical methods: Exponential convergence in the Tucker/CP rank for classical potentials ($f = e^{-\mathbf{r}}$, $f = e^{-\mathbf{r}}/\mathbf{r}$, $f = 1/\mathbf{r}$). Numerics **[Khoromskij, Khoromskaia '07, ...]**

Theory on the canonical/Tucker approximations to a class of 3D functions **[Khoromskij '06 - '10]**

$$\|\mathbf{A} - \mathbf{A}_R\| \leq Ce^{-\gamma R}.$$





The approximation theory for classical potentials via canonical and Tucker formats is based on the *sinc* methods (see Lect. 7).

Main approaches combine: Analytic methods of approximation + high-order extension(s) of truncated SVD + nonlinear alternating iteration + multigrid.

Computing canonical decomposition: heuristic Pure Greedy Algorithm (PGA)

For $\mathcal{S} = \mathcal{C}_R$, the canonical decomposition can be considered in the framework of the best R-term approximation with regard to a redundant set $\mathcal{D} = \mathcal{C}_1$ (dictionary) of rank-1 functions. **[V. Temlyakov]**.

For $f \in \mathbb{H}$, the best R-term approximation error is defined by

$$\sigma_R(f, \mathcal{D}) := \inf_{s \in \Sigma_R(\mathcal{D})} \|f - s\|, \quad \Sigma_R(\mathcal{D}) = \{s = \sum_{g \in \mathcal{D}} c_g g : \ \#\text{summands} \leq R\}.$$

Let $g = g(f) \in \mathcal{D}$ maximize $|\langle f, g\rangle|$ with $\|g\| = 1$, (best rank-1 nonlinear approx.).
Define

$$G(f) := \langle f, g\rangle g, \quad R(f) := f - G(f).$$

$R_0(f) := f$ and $G_0(f) := 0$.
The *Pure Greedy Algorithm* (PGA) inductively computes an estimate to the best R-term approximation: Given $f \in \mathbb{H}$, introduce

$$R_0(f) := f \quad \text{and} \quad G_0(f) := 0.$$

Then, for all $1 \leq m \leq R$, we inductively define

$$G_m(f) := G_{m-1}(f) + G(R_{m-1}(f)), \quad R_m(f) := f - G_m(f).$$

The decomposition in $\mathcal{C}_R$,

$$f = \sum_{k=1}^{R} a_k v_k, \quad v_k = \phi_k^{(1)}(x_1) \otimes ... \otimes \phi_k^{(d)}(x_d) \in \mathcal{C}_1,$$

is called *completely orthogonal* if

$$\langle \phi_k^{(\ell)}, \phi_m^{(\ell)} \rangle = \delta_{k,m}, \quad \forall \ell = 1, ..., d \;\Leftrightarrow\; \Phi^{(\ell)} = [\phi_1^{(\ell)}, ..., \phi_R^{(\ell)}] - \textit{orthogonal set}.$$

Greedy *completely orthogonal decomposition* (COD) is defined as PGA with the orthogonality constraint on $\Phi^{(\ell)}$.

Lem. (Tucker format with the diagonal core.) **[Golub, Zhang 2001].** Let $f \in \mathbb{H}$ allow a rank-R *COD*. Then the Greedy COD algorithm correctly computes it.

If $a_1 > a_2 > ... > a_R > 0$, then the decomposition is unique.

Limitations. Poor approximation properties of the COD !

Poor convergence properties of PGA on general tensors.