Polynomial Approximation in Hierarchical Tucker Format by Vector–Tensorization

Lars Grasedyck*

Bericht Nr. 308 **April 2010**

Key words: Hierarchical Tucker, tensorization, tensor rank, tensor approximation, tensor train, TT.

MSC: 15A69, 65F99

Institut für Geometrie und Praktische Mathematik
RWTH Aachen

Templergraben 55, D–52056 Aachen (Germany)

* Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 52056 Aachen, Tel.: +49-241-80-93951, Tel.: +49-241-80-92317, lgr@igpm.rwth-aachen.de. Financial support from the DFG under grant GRA2179/2-1 gratefully acknowledged.

Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization

Lars Grasedyck *

April 12, 2010

We analyze and characterize the possibility to represent or approximate tensors that stem from a tensorization of vectors, matrices, or tensors by low (hierarchical) rank. Our main result is that for vectors that stem from the evaluation of a polynomial f of degree p on an equispaced grid, the (hierarchical) rank is bounded by $1+p$. This is not true for the canonical rank, and we prove this by a small counterexample. We extend our result to functions with (few) singularities that are otherwise analytic: for an asymptotically smooth function with m singularities the rank required to achieve a point-wise accuracy of ε is of the size $k \leq C + \log_2(1/\varepsilon) + 2m$. The storage requirements for a tensorized vector of length n are $\mathcal{O}(\log(n)k^3)$ and arithmetic operations (e.g., truncated addition) in the format are of $\mathcal{O}(\log(n)k^4)$ complexity.

Keywords: Hierarchical Tucker, Tensorization, Tensor Rank, Tensor Approximation, Tensor Train, TT.
MSC: 15A69, 65F99

1 Introduction

In this article we construct tensors

$$A \in \mathbb{R}^{n_1 \times \cdots \times n_d}, \qquad n_1, \ldots, n_d \in \mathbb{N},$$

from vectors

$$x \in \mathbb{R}^N, \qquad N := n_1 \cdots n_d.$$

The aim for such a "tensorization" is to find a data-sparse representation of the tensor A and thus reduce, among others, the storage complexity.

This idea was first formulated by Oseledets [10] in the context of matrix tensorization. In numerical experiments he found out that the TT-rank of a tensorization of the 1d and 2d

*Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 52056 Aachen, Tel.: +492418093951, Tel.: +492418092317, lgr@igpm.rwth-aachen.de. Financial support from the DFG under grant GRA2179/2-1 gratefully acknowledged.

Laplacian as well as the Hilbert matrix has its rank bounded by a small constant. We will prove that the rank is indeed bounded. Khoromskij [7] was able to prove that vectors of the exponential form

$$x_i = c \cdot d^i, \qquad i = 1, \ldots, N,$$

allow for a canonical rank one representation after tensorization. Thus, exponential sums and trigonometric sums have tensorizations of small canonical and hierarchical rank proportional to the number of addends. Khoromskij and Oseledets [11] have then successfully applied the tensorization for the solution of PDEs in high dimension and with very large mode sizes.

In this paper, we will address the question for which vectors (respectively matrices and tensors) such a tensorization is possible with low (hierarchical) rank. In Section 2 we define the tensorization of vectors, in Section 3 we introduce the hierarchical Tucker format [5, 4] and the corresponding hierarchical rank. In Section 4 we characterize the class of vectors whose tensorization has hierarchical rank k. Then, we prove that the tensorization of vectors that stem from the evaluation of polynomials of degree at most p on an equispaced grid has a hierarchical rank of at most $p + 1$. A simple example shows that the canonical rank is larger than $p + 1$ and also that a structured grid is essential. In Section 5 we prove that any asymptotically smooth, i.e., piecewise analytic, function with m singularities or discontinuities can be discretized and tensorized such that the rank for an ε-approximation in the $\|\|_\infty$-norm is bounded by $C + \log_2(1/\varepsilon) + 2m$. In Section 6 we consider the tensorization of matrices and tensors and extend the results to arbitrary (e.g., prime) numbers N.

The important point is that the existence can be proven independently of the construction of an approximation. Thus, considering high-order polynomials and additional locally singular functions is reasonable. The strongest requirement is the structure of the underlying grid.

2 Tensorization of Vectors

For a vector $x \in \mathbb{R}^{n \cdot m}$ we define the tensorization (or $\mathcal{F}$olding) of x into a matrix by

$$(\mathcal{F}(x))_{i,j} := x_{i+(j-1)n}, \qquad \mathcal{F}(x) \in \mathbb{R}^{n \times m}.$$

The vector x thus contains the columns of the matrix $\mathcal{F}(x)$ one after the other. The same ordering is used in the standard LAPACK (`http://www.netlib.org/lapack`) matrix format for the representation of dense general matrices. This can be generalized to tensors as follows.

Definition 1 (Vector-Tensorization) *For vectors $x \in \mathbb{R}^{n_1 \cdots n_d}$ we define the tensorization*

$$\mathcal{F} : \mathbb{R}^{n_1 \cdots n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$$

for all indices $i_\mu \in \{1, \ldots, n_\mu\}, \mu = 1, \ldots, d$, by

$$(\mathcal{F}(x))_{i_1,\ldots,i_d} := x_\ell, \qquad \ell := i_1 + \sum_{\mu=2}^{d} (i_\mu - 1) \prod_{\nu=1}^{\mu-1} n_\nu.$$

Example 2 *Consider the vector $x = (1, 2, 3, 4, 5, 6, 7, 8) \in \mathbb{R}^{2 \cdot 2 \cdot 2}$. The tensorization of x is*

$$\mathcal{F}(x) = \left[\begin{array}{cc|cc} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{array} \right]$$

or index-wise

$$\mathcal{F}(x)_{0,0,0} = 1, \quad \mathcal{F}(x)_{0,1,0} = 3, \quad \mathcal{F}(x)_{0,0,1} = 5, \quad \mathcal{F}(x)_{0,1,1} = 7,$$
$$\mathcal{F}(x)_{1,0,0} = 2, \quad \mathcal{F}(x)_{1,1,0} = 4, \quad \mathcal{F}(x)_{1,0,1} = 6, \quad \mathcal{F}(x)_{1,1,1} = 8.$$

Remark 3 *The tensorization $\mathcal{F}$ is an isometric isomorphism for the $\|\|_2$-norm on $\mathbb{R}^{n_1 \cdots n_d}$ and the so-called Frobenius norm $\|A\| := \sqrt{\sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} A_{i_1,\ldots,i_d}^2}$ on $\mathbb{R}^{n_1 \times \cdots \times n_d}$.*

The idea why one could be interested in a tensorization of vectors is that the tensor might allow for a low rank representation and correspondingly for a low rank arithmetic. For dimension $d > 2$ there exist several notions of rank. The most data-sparse is the canonical rank which is based on outer products of vectors:

$$(a_1 \otimes \cdots \otimes a_d)_{i_1,\ldots,i_d} := \prod_{\mu=1}^{d} (a_\mu)_{i_\mu}.$$

Definition 4 (Canoncial rank, CP representation) *Let $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. The minimal number $k \geq 0$ such that*

$$A = \sum_{i=1}^{k} A_i, \qquad A_i = a_{i,1} \otimes \cdots \otimes a_{i,d}, \qquad a_{i,\mu} \in \mathbb{R}^{n_\mu}, \tag{1}$$

is the tensor rank or canonical rank of A. A sum of the form (1) with arbitrary (not necessarily minimal) $k \geq 0$ is called a representation with representation rank k. In the literature such a data-sparse representation is called CANDECOMP [2] or PARAFAC [6] or simply CP model.

For $d = 30$ and $n = 2$ a vector $x \in \mathbb{R}^{\prod_{\mu=1}^{d} n}$ has more than one billion entries (possibly all non-zero), whereas a CP representation (1) of $\mathcal{F}(x)$ with rank k requires only $60k$ parameters. Of course, the three main questions that arise are:

1. Is it possible to approximate $\mathcal{F}(x)$ for a vector x in low rank format ?
2. Can one find such an approximation efficiently ?
3. Is it possible to perform standard arithmetic operations (e.g. linear combinations) with tensors of low rank ?

The last two questions have been answered in [4] for the hierarchical Tucker format, which will be introduced in the next section. The first question is of particular interest here because it allows us to understand in which cases the tensorization makes sense.

The canonical rank of a tensorized vector is typically very large. This is an observation that is difficult to prove or underline by numerical experiments, since there is no way to determine the rank of a high-dimensional tensor. A simple example follows where one can determine the rank.

Example 5 *Consider the vector* $x = (1,2,3,4,5,6,7,8) \in \mathbb{R}^{2\cdot 2\cdot 2}$. *The tensorization of* x *is*

$$\mathcal{F}(x) = \left[\begin{array}{cc|cc} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{array}\right]$$

with front- and back-matrices ($i_3 = 0$ *or* $i_3 = 1$, *respectively)*

$$A := \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \qquad B := \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}.$$

The multilinear rank [12] of $\mathcal{F}(x)$ *is* $(2,2,2)$ *and the hyperdeterminant is*

$$\Delta = \left(\frac{\det(A+B) - \det(A-B)}{2}\right)^2 - 4\det(A)\det(B) = 0.$$

Both together prove that the canonical rank of $\mathcal{F}(x)$ *must be three [3], which is the maximal possible rank [8].*

3 The hierarchical Tucker format

In the hierarchical Tucker format the sparsity of the representation is determined by the hierarchical rank which is the rank of certain matricizations of the tensor. Throughout this section we use the notation

$$\mathcal{I} := \mathcal{I}_1 \times \cdots \times \mathcal{I}_d, \qquad \mathcal{I}_\mu := \{1, \ldots, n_\mu\}, \quad \mu = 1, \ldots, d.$$

3.1 Definition of the $\mathcal{H}$-Tucker format

Definition 6 (Matricization) *For a tensor* $A \in \mathbb{R}^{\mathcal{I}}$, *a collection of dimension indices* $t \subset \{1, \ldots, d\}$ *and the complement* $s := \{1, \ldots, d\} \setminus t$ *the matricization*

$$A^{(t)} \in \mathbb{R}^{\mathcal{I}_t \times \mathcal{I}_s}, \qquad \mathcal{I}_t := \bigtimes_{\mu \in t} \mathcal{I}_\mu, \quad \mathcal{I}_s := \bigtimes_{\mu \in s} \mathcal{I}_\mu,$$

is defined by its entries

$$\left(A^{(t)}\right)_{(i_\mu)_{\mu\in t}, (i_\mu)_{\mu \in s}} := A_{i_1, \ldots, i_d}.$$

Based on the matricization of a tensor A with respect to several sets $t \subset \{1, \ldots, d\}$ one can define the hierarchical rank and the hierarchical Tucker format. In order to be able to perform efficient arithmetics, we require the sets t to be organized in a tree.

Definition 7 (Dimension tree) *A dimension tree or mode cluster tree* $T_{\mathcal{I}}$ *for dimension* $d \in \mathbb{N}$ *is a tree with root* $\mathrm{Root}(T_{\mathcal{I}}) = \{1, \ldots, d\}$ *and depth* p *such that each node* $t \in T_{\mathcal{I}}$ *is either*

1. *a leaf and singleton* $t = \{\mu\}$ *or*

2. *the union of two disjoint successors* $S(t) = \{t_1, t_2\}$:

$$t = t_1 \,\dot\cup\, t_2. \qquad (2)$$

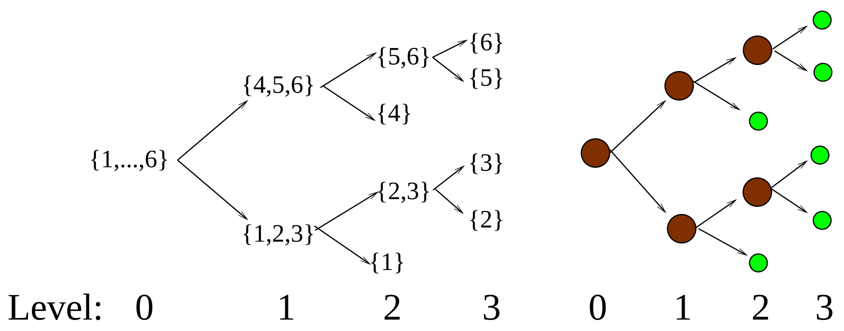

Figure 1: Left: A dimension tree for $d = 6$. Right: The interior nodes $\mathcal{I}(T_\mathcal{I})$ are colored dark (brown), the leaves $\mathcal{L}(T_\mathcal{I})$ are light (green).

The level ℓ of the tree is defined as the set of all nodes having a distance of exactly ℓ to the root, cf. Figure 1 and Example 9. We denote the level ℓ of the tree by

$$T_\mathcal{I}^\ell := \{t \in T_\mathcal{I} \mid \mathrm{level}(t) = \ell\}.$$

The set of leaves of the tree is denoted by $\mathcal{L}(T_\mathcal{I})$ and the set of interior (non-leaf) nodes is denoted by $\mathcal{I}(T_\mathcal{I})$. A node of the tree is a so-called mode cluster (a union of modes).

Definition 8 (Hierarchical rank, $\mathcal{H}$-Tucker) *Let $T_\mathcal{I}$ be a dimension tree. The* hierarchical rank $(k_t)_{t\in T_\mathcal{I}}$ *of a tensor $A \in \mathbb{R}^\mathcal{I}$ is defined by*

$$\forall t \in T_\mathcal{I} : \quad k_t := \mathrm{rank}(A^{(t)}).$$

The set of all tensors of hierarchical rank (node-wise) at most $(k_t)_{t\in T_\mathcal{I}}$, the so-called $\mathcal{H}$-Tucker *tensors, is denoted by*

$$\mathcal{H}\text{-Tucker}((k_t)_{t\in T_\mathcal{I}}) := \{A \in \mathbb{R}^\mathcal{I} \mid \forall t \in T_\mathcal{I} : \ \mathrm{rank}(A^{(t)}) \le k_t\}.$$

In the hierarchical format only some of the possible subsets t of all modes appear. A special case if the so-called TT-format [9] with corresponding TT-rank and TT-tree, where all nodes are of the form

$$t = \{1, \dots, q\} \text{ or } t = \{q + 1\}, \qquad q = 1, \dots, d.$$

This tree form provides a tree of maximal depth (which may be disadvantageous [4, Section 5]), but we will later see that it may provide a lower rank bound and it is easier to analyse.

In the canonical case the tree is of minimal depth $p := \lceil \log_2(d) \rceil := \min\{i \in \mathbb{N}_0 \mid i \ge \log_2(d)\}$ with mode clusters of the form

$\{1, \dots, d\}$
$\{1, \dots, \lceil d/2 \rceil\}, \{1 + \lceil d/2 \rceil, \dots, d\},$
$\{1, \dots, \lceil d/4 \rceil\}, \{1 + \lceil d/4 \rceil, \dots, \lceil 2d/4 \rceil\}, \{1 + \lceil 2d/4 \rceil, \dots, \lceil 3d/4 \rceil\}, \{1 + \lceil 3d/4 \rceil, \dots, d\},$
etc.

Example 9 *For the dimension indices* $\{1,\dots,4\}$ *the canonical tree is*

$$\begin{array}{lcl} \textit{level } 0 & : & \{1,\dots,4\} \\ \textit{level } 1 & : & \{1,2\}, \quad \{3,\dots,4\} \\ \textit{level } 2 & : & \{1\}, \quad \{2\}, \quad \{3\}, \quad \{4\} \end{array}$$

and the TT-tree is

$$\begin{array}{ll} \textit{level } 0: & \{1,\dots,4\} \\ \textit{level } 1: & \{1,\dots,3\}, \quad \{4\} \\ \textit{level } 2: & \{1,2\}, \quad \{3\} \\ \textit{level } 3: & \{1\}, \quad \{2\}. \end{array}$$

The matricizations of the tensor $\mathcal{F}(x)$, $x_i = i, i \in \{1,\dots,16\}$ *in* $\mathbb{R}^{2\times2\times2\times2}$ *are*

$$\mathcal{F}(x)^{\{1\}} = \begin{bmatrix} 1 & 3 & 5 & 7 & 9 & 11 & 13 & 15 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 \end{bmatrix}, \mathcal{F}(x)^{\{2\}} = \begin{bmatrix} 1 & 2 & 5 & 6 & 9 & 10 & 13 & 14 \\ 3 & 4 & 7 & 8 & 11 & 12 & 15 & 16 \end{bmatrix},$$

$$\mathcal{F}(x)^{\{3\}} = \begin{bmatrix} 1 & 2 & 3 & 4 & 9 & 10 & 11 & 12 \\ 5 & 6 & 7 & 8 & 13 & 14 & 15 & 16 \end{bmatrix}, \mathcal{F}(x)^{\{4\}} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \end{bmatrix},$$

$$\mathcal{F}(x)^{\{1,2,3\}} = (\mathcal{F}(x)^{\{4\}})^T, \quad \mathcal{F}(x)^{\{1,2\}} = \begin{bmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{bmatrix}, \mathcal{F}(x)^{\{3,4\}} = (\mathcal{F}(x)^{\{1,2\}})^T.$$

3.2 Arithmetics in the $\mathcal{H}$-Tucker format

The storage complexity for a tensor in the CP-model representation with rank k, mode size $n \sim k$ and order d is $\mathcal{O}(dk^2)$. In the $\mathcal{H}$-Tucker format the storage complexity is one factor k higher. Apart from the mathematical definition of the $\mathcal{H}$-Tucker format we also require an efficient representation.

Lemma 10 (Hierarchical Tucker format, [4]**)** *Let* $T_{\mathcal{I}}$ *be a dimension tree and* $A \in \mathcal{H}\text{-Tucker}((k_t)_{t\in T_{\mathcal{I}}})$. *Then* A *can be represented by transfer tensors* $(B_t)_{t\in\mathcal{I}(T_{\mathcal{I}})}$ *(for interior nodes) and mode frames* $(U_t)_{t\in\mathcal{L}(T_{\mathcal{I}})}$ *(for leaves), where* $B_t \in \mathbb{R}^{k_t\times k_{t_1}\times k_{t_2}}$ *for* $S(t) = \{t_1, t_2\}$ *and* $U_t \in \mathbb{R}^{\mathcal{I}_t\times k_t}$.

The storage complexity for B_t, U_t from the previous lemma is

$$\text{Storage}((B_t)_{t\in\mathcal{I}(T_{\mathcal{I}})}, (U_t)_{t\in\mathcal{L}(T_{\mathcal{I}})}) \ \leq \ (d-1)k^3 + k\sum_{\mu=1}^{d} n_\mu, \qquad k := \max_{t\in T_{\mathcal{I}}} k_t, \tag{3}$$

i.e. linear in the order d (provided that k is uniformly bounded) [4].

Basic arithmetic operations like linear combinations of $\mathcal{H}$-Tucker tensors can be performed exact, but the representation rank $(k_t)_{t\in T_{\mathcal{I}}}$ will be proportional to the sum of the representation ranks. It is therefore necessary to reduce (truncate) the rank of a tensor A by finding (almost) best approximations with prescribed rank, $\mathcal{T}_k(A)$, or (almost) minimal rank approximations with prescribed truncation accuracy, $\mathcal{T}_\varepsilon(A)$. Such a truncation is possible in the

$\mathcal{H}$-Tucker format. The details are not relevant here, we just summarize the main result from [4].

Let $T_{\mathcal{I}}$ be a dimension tree and $A \in \mathcal{H}\text{-Tucker}(k_t)_{t\in\mathcal{I}}$. Let A^{best} denote the best approximation of A in $\mathcal{H}\text{-Tucker}((\tilde{k}_t)_{t\in T_{\mathcal{I}}})$ and $\mathcal{T}_{\tilde{k}}(A)$ the truncation of A to rank $(\tilde{k}_t)_{t\in T_{\mathcal{I}}}$. Then the truncation is quasi-optimal,

$$\|A - \mathcal{T}_{\tilde{k}}(A)\| \leq \sqrt{2d-3}\|A - A^{\text{best}}\|,$$

and it can be computed in

$$\mathcal{O}(d \max_{t\in T_{\mathcal{I}}} k_t^4 + \sum_{\mu=1}^{d} n_\mu k_\mu^2).$$

For the proof we refer to [4, Theorem 3.11, Remark 3.12, Lemma 4.9].

We conclude that the $\mathcal{H}$-Tucker format is almost as data-sparse as the CP-model, and additionally it allows for a formatted (truncated) arithmetic in quasi-optimal complexity (one additional factor k) and with quasi-optimal accuracy (a factor $\sqrt{2d-3}$). The crucial question is: why should the rank k be small ? This question will be answered in the following.

4 Matricization of Tensorized Vectors

We recall that the tensorization $\mathcal{F}(x) \in \mathbb{R}^{n_1\times\cdots\times n_d}$ of a vector $x \in \mathbb{R}^{n_1\cdots n_d}$ is of the form

$$(\mathcal{F}(x))_{i_1,\ldots,i_d} := x_\ell, \qquad \ell := i_1 + \sum_{\mu=2}^{d}(i_\mu - 1)\prod_{\nu=1}^{\mu-1} n_\nu.$$

We will first look at a special case, the TT-format, where nodes of the tree are of a simple form.

Theorem 11 (Hierarchical rank of tensorized vectors in TT-format) *Let $\mathcal{F}(x) \in \mathbb{R}^{n_1\times\cdots\times n_d}$ be the tensorization of x and for $q = 1,\ldots,d$ let $N_q := \prod_{\mu=1}^{q} n_\mu$ and $N_q' := \prod_{\mu=q+1}^{d} n_\mu$. Then the rank of the matricization $\mathcal{F}(x)^{(t)}$, $t := \{1,\ldots,q\}$, is*

$$\text{rank}(\mathcal{F}(x))^{(t)} = \dim\left(\text{span}\{x|_{I_{\ell,q}} \mid \ell = 1,\ldots,N_q'\}\right), \quad I_{\ell,q} := \{1 + (\ell-1)N_q, \ldots, \ell N_q\}.$$

In particular, if for given $t = \{1,\ldots,q\}$ all sub-vectors fulfill

$$\forall \ell = 1,\ldots,N_q' : \qquad x|_{I_{\ell,q}} \in V_q \subset \mathbb{R}^{N_q},$$

then the rank of the matricization $(\mathcal{F}(x))^{(t)}$ is bounded by $\dim V_q$.

Proof: The matricization $(\mathcal{F}(x))^{(t)}$ has entries of the form

$$(\mathcal{F}(x))^{(t)}_{(i_1,\ldots,i_q),(i_{q+1},\ldots,i_d)} = x_i.$$

We consider a single column of the matricization. For this, let $(i_{q+1},\ldots,i_d)$ be fixed. The parameters for the row indices are $(i_1,\ldots,i_q)$. In the vector x the corresponding entries of the column are consecutively arranged from

$$\ell_1 := 1 + \sum_{\mu=q+1}^{d}(i_\mu - 1)\prod_{\nu=1}^{\mu-1} n_\nu = 1 + \left(\sum_{\mu=q+1}^{d}(i_\mu - 1)\prod_{\nu=q+1}^{\mu-1} n_\nu\right) N_q =: 1 + (\ell-1)N_q$$

to

$$\ell_2 := N_q + \sum_{\mu=q+1}^{d} (i_\mu - 1) \prod_{\nu=1}^{\mu-1} n_\nu = N_q - 1 + \ell_1 = \ell N_q.$$

Thus, the column is the sub vector $x|_{\ell_1,\dots,\ell_2} = x|_{I_{\ell,q}}$, which proves the theorem. ∎

The assumption on the sub-vectors of x is a bit abstract. In the following corollary we consider a concrete and interesting example, where the vector x is the discretization of a function on an equispaced grid. Prior to that we consider general mode clusters t.

Theorem 12 (Hierarchical rank of tensorized vectors in $\mathcal{H}$-Tucker-format) *Let $\mathcal{F}(x) \in \mathbb{R}^{n_1\times\cdots\times n_d}$ be the tensorization of x and for $t = \{r+1,\dots,s\}$ let $N_r := \prod_{\mu=1}^{r} n_\mu$, $N_s := \prod_{\mu=1}^{s} n_\mu$, and $N_s' := \prod_{\mu=s}^{d} n_\mu$. Then the rank of the matricization $\mathcal{F}(x)^{(t)}$ is*

$$\begin{aligned} \operatorname{rank}(\mathcal{F}(x))^{(t)} &= \dim\left(\operatorname{span}\{x|_{I_{\ell,t}} \mid \ell = 1,\dots,N_r N_s'\}\right), \\ I_{\ell_1+(\ell_2-1)N_r,t} &:= \{\ell_1 + (j-1)N_r + (\ell_2-1)N_s \mid j = 1,\dots,N_s/N_r\}. \end{aligned}$$

Proof: The matricization $(\mathcal{F}(x))^{(t)}$ has entries of the form

$$(\mathcal{F}(x))^{(t)}_{(i_{r+1},\dots,i_s),(1,\dots,i_r,i_{s+1},\dots,i_d)} = x_i.$$

We consider a single column of the matricization. For this, let $(1,\dots,i_r,i_{s+1},\dots,i_d)$ be fixed. The parameters for the row indices are $(i_{r+1},\dots,i_s)$. The corresponding entries in x are

$$\underbrace{1 + \sum_{\mu=1}^{r} (i_\mu - 1) \prod_{\nu=1}^{\mu-1} n_\nu}_{\in[1,2,\dots,N_r]} + \underbrace{\sum_{\mu=r+1}^{s} (i_\mu - 1) \prod_{\nu=1}^{\mu-1} n_\nu}_{\in N_r[0,1,\dots,N_s/N_r-1]} + \underbrace{\sum_{\mu=s+1}^{d} (i_\mu - 1) \prod_{\nu=1}^{\mu-1} n_\nu}_{\in N_s[0,1,\dots,N_s'-1]}.$$

∎

In the TT-format every column of $\mathcal{F}(x)^{(t)}$ is a contiguous sub-vector of x. In the $\mathcal{H}$-Tucker-format the columns are 'staggered' sub-vectors of x.

Corollary 13 (Polynomial Approximation) *Let $(x_i)_{i=1}^{n_1\cdots n_d}$, $d \geq 1$, be the discrete evaluation of a polynomial $f(y)$ of degree p on a regular (equispaced) grid of points $\xi_i = (i-1)h$, $h := 1/(N-1)$, $N := \prod_{\mu=1}^{d} n_\mu$, i.e.*

$$x_i = f(\xi_i).$$

Then for every $t = \{r+1,\dots,s\} \subset \{1,\dots,d\}$ the tensorization $\mathcal{F}(x)^{(t)}$ fulfills

$$\operatorname{rank}(\mathcal{F}(x)^{(t)}) \leq p + 1.$$

Proof: We consider in the first part the case $r = 0$ and treat the general case afterwards. For every ℓ the sub-vector $x_{1+(\ell-1)N_q,\dots,\ell N_q}$ is the discrete evaluation of some polynomial (shifted f) of degree p in the same N_q nodes $h\{1,2,\dots,N_q\}$. The polynomials are different, but the nodal points are the same so that the interpolation maps the $p+1$-dimensional space of polynomials of degree less or equal p to an at most $p+1$ dimensional vector space V_q. For the general case, the sub-vectors are of the form

$$x_{\ell_1+0N_r+(\ell_2-1)N_s,\cdots,\ell_1+(N_s/N_r-1)N_r+(\ell_2-1)N_s} = x_{\ell'+(j-1)N_r}, \quad j = 1,\dots,N_s/N_r.$$

Thus, every vector is the interpolation of a (different) polynomial of degree at most p in the same points $H, 2H, \dots, N_s/N_r H$, $H := N_r h$. Again, the interpolation maps into an at most $p+1$-dimensional space V_q. ∎

Example 14 *The tensor $\mathcal{F}(x)$ from Example 9 is the tensorization of the vector x which is the evaluation of the linear function $f(y) = y$ on the grid $\{1, 2, \ldots, 16\}$. The matricizations all have exactly rank $k = 2$. In particular $\mathcal{F}(x)^{\{1,2\}}$ has four columns which are the four parts $x|_{\{\ell,\ell+1,\ell+2,\ell+3\}}, \ell = 1, \ldots, 4$, of the vector x.*

One can easily see that an equidistant (or similarly structured) grid is necessary: for the function $f(y) = y$ the value x_i codes the location of the grid point ξ_i, so if all of them can be chosen independently one has to store all the values independently.

5 Singularities

The result on the polynomial approximation gives some insight into the class of functions that can easily be approximated by low rank. However, we are also interested in the approximation of functions that have a singularity, i.e., where the polynomial approximation deteriorates as the number of grid points increases. In order to treat the problem in more generality, we will allow a finite number of singularities and assume that the function is analytic otherwise.

Definition 15 (Asymptotically smooth) *Let $f : J = [a, b] \to \mathbb{R}$ be a function and let $S := \{s_1, \ldots, s_m\} \subset J$. The function f is said to be asymptotically smooth with singular points S, if for all points $y \in J \setminus S$ holds*

$$|\partial^i f(y)| \leq C_1 \gamma^\sigma \gamma^i i!, \quad \gamma \leq C_2 \mathrm{dist}(y, S)^{-1},$$

where σ is the degree of singularity.

Example 16 *The function $f(x) = x^{-1}$ is asymptotically smooth on any interval $J \subset \mathbb{R}$ with singular point $S = \{0\}$ and degree $\sigma = 1$:*

$$|\partial^i f(y)| = i!|y^{-\sigma-i}| \leq \gamma^1 \gamma^i i!, \quad \gamma := |y|^{-1} = \mathrm{dist}(y, S)^{-1}.$$

We can nicely approximate an asymptotically smooth function in intervals that are bounded away from the singularities. Thus, if we ignore the two closest intervals per singular point, then on all other intervals we can apply a polynomial approximation. The few singularities will then require a rank increased by their number times two. Before we formulate the theorem, we need an auxiliary lemma.

Lemma 17 *Let $N = \prod_{\mu=1}^{d} n_\mu$ and for $t := \{r + 1, \ldots, s\} \subset \{1, \ldots, d\}$ let N_r, N_s, N'_s be defined as in Theorem 12. Then the sets*

$$I_{\ell_1+(\ell_2-1)N_r,t} := \{\ell_1 + (j-1)N_r + (\ell_2 - 1)N_s \mid j = 1, \ldots, N_s/N_r\}$$

form a partition of $\{1, \ldots, N\}$.

Proof: There is no overlap between the sets (N_s is a multiple of N_r and $\ell_1 \leq N_r$), and all indices $\ell = 1, \ldots, N$ can be written in the form $\ell = \ell_1 + (j-1)N_r + (\ell_2 - 1)N_s$ for integers $\ell_1 \leq N_r, j \leq N_s/N_r, \ell_2 \leq N'_s$. As a special case $r = 0$ the TT-format is included, where $N_r = 1$. ∎

Theorem 18 (TT-Rank of tensorizations of asymptotically smooth functions) *Let $(x_i)_{i=1}^N$, $d \geq 1$, be the discrete evaluation of an asymptotically smooth function f with singular points $S = \{s_1, \ldots, s_m\}$ on a regular (equispaced) grid of points $\xi_i = (i-1)h, h := 1/(N-1)$, $N = \prod_{\mu=1}^d n_\mu$, i.e.*

$$x_i = f(\xi_i).$$

Then there exists a tensor $Z \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ that approximates the tensorization $\mathcal{F}(x)$ point-wise with accuracy ε and that has its hierarchical rank bounded by

$$\operatorname{rank}(Z^{(t)}) \leq C + \log_2(1/\varepsilon) + 2m, \qquad t = \{1, \ldots, q\}.$$

The constant C depends on the degree of singularity σ.

Proof: We construct the tensor Z by a vector-tensorization $Z := \mathcal{F}(z)$, where the entries of z are given on the same grid as those of x but for a different function $z_i = g(\xi_i)$. The function g is defined piecewise, where it is either a polynomial of degree p or it is the function f itself. Thus, it suffices to bound the rank of Z and to determine the degree of the polynomials necessary to achieve the desired point-wise accuracy $\|f - g\|_\infty < \varepsilon$. The degree p will be specified later.

First, we partition the grid of points hierarchically over all levels $q = 1, \ldots, d$ as follows. At the start the partition is empty, $\mathcal{P} := \emptyset$. For each level $q = 1, \ldots, d$ we add to $\mathcal{P}$ all sets $I_{\ell,q}$ where

1. $I_{\ell,q} \not\subset \cup\mathcal{P}$ and
2. $\operatorname{dist}(x_i, S) > hN_q/2$ or $N_q < 2$.

On each of the sets $I_{\ell,q} \in \mathcal{P}$ we apply a polynomial approximation, except for the small sets where one can simply use the function f.

1. Polynomial approximation on intervals $I_{\ell,q} \in \mathcal{P}, N_q \geq 2$: By definition of the partition the set fulfills $\operatorname{dist}(x_i, S) > hN_q/2$ for all $i \in I_{\ell,q}$. We define the interval $J := [\min_{i \in I_{\ell,q}} \xi_i,\ \max_{i \in I_{\ell,q}} \xi_i]$. The diameter of J can be estimated by $|J| < hN_q$. The distance of J to S is at least $hN_q/2$. According to [1, Lemma 3.13] the $\|\|_\infty$ best approximation of f by a polynomial g of degree p fulfills

$$\begin{aligned} \|f - g\|_{\infty,J} &\leq 4eC_1\gamma^\sigma(1 + \gamma|J|)p(1 + 2/(\gamma|J|))^{-p} \\ &\leq 4eC_1(hN_q/2)^{-\sigma}3p2^{-p}. \end{aligned}$$

   In order to achieve $\|f - g\|_{\infty,J} < \varepsilon$ we thus have to choose

$$p := C + \log_2(1/\varepsilon).$$

   Note that the constant C may depend on the degree of singularity σ.

2. Rank bound for all matricizations: Let $q \in \{1, \ldots, d\}$. According to Theorem 12 we have to determine the dimension of the span of all vectors $(x|_{I_{\ell,q}})_{\ell=1}^{N'_q}$. For all sets on which we have applied the polynomial approximation the span has dimension at most $p + 1$. The number of intervals that do not fulfill the distance criterion, i.e., where $\operatorname{dist}(x_i, S) > hN_q/2$, is at most two times the number m of singular points, i.e., there are at most $2m$ vectors where the polynomial approximation is not applied. Thus the dimension of the total span is at most

$$\operatorname{rank}(Z^{(1,\ldots,q)}) \leq p + 1 + 2m.$$

■

We have to stress again that the result of Theorem 18 depends strongly on the regular structure of the grid.

Remark 19 ($\mathcal{H}$-Tucker **format)** *For the general $\mathcal{H}$-*Tucker *format the nodes are $t = \{r + 1, \ldots, s\}$ and the index sets are of the form*

$$I_{\ell_1+(\ell_2-1)N_r,t} = \{\ell_1 + (j-1)N_r + (\ell_2 - 1)N_s \mid j = 1, \ldots, N_s/N_r\}$$

For all but $2m$ values of ℓ_2 the singularities have a distance of at least hN_s to the nodes corresponding to indices $j \in I_{\ell_1+(\ell_2-1)N_r,t}$. The maximal distance between nodes is $hN_rN_s/N_r = hN_s$, thus the same polynomial approximation as above (the TT case) is applicable. It remains to find a low rank approximation for indices with the $2m$ fixed values of ℓ_2. Here we change the construction from column-wise to row-wise, i.e., we approximate every row-part except the $2m$ close to the singularities, by a polynomial of degree p. Again, for one row-part (ℓ_1 varying) the domain for the interpolation is of size hN_r and the distance to the singularity is at least hN_r. In total the rank is proportional to $4m(p+1)+2m$. One can as well derive the estimate $k \le (C+1+p+2m)^2$ directly by [4, 5.3.2] from Theorem 18.

6 Matrix and Tensor Tensorization

The tensorization has been defined so far for vectors that stem from the discretization of a function f defined on an interval in $\mathbb{R}$. We will now consider the case where $f(x,y)$ is a bivariate function and thus a matrix has to be tensorized. If the function $f(x,y)$ can be approximated by a degenerate expansion

$$f(x,y) \approx \sum_{i=1}^{k} f_i(x) g_i(y)$$

with few terms k, then one can simply proceed as before with the functions f_i, g_i which are now univariate. However, when the coupling between x and y is strong, e.g.

$$f(x,y) = 1/|x-y|, \qquad x, y \in [0,1],$$

then such a type of approximation is worthless since k will grow with the meshwidth. In this case one can simply forbid to ever separate the two directions x, y.

6.1 Tensorization of Matrices

For a matrix $X \in \mathbb{R}^{(n_1 \cdot n_2) \times (m_1 \cdot m_2)}$ we define the tensorization of X by

$$(\mathcal{M}^*(X))_{(i_1,i_2),(j_1,j_2)} := X_{i_1+(i_2-1)n_1, j_1+(j_2-1)m_1}, \qquad \mathcal{M}^*(x) \in \mathbb{R}^{(n_1 \times n_2) \times (m_1 \times m_2)}.$$

This corresponds to a vector-tensorization of all rows and columns of the matrix.

Definition 20 (Preliminary Matrix Tensorization) *We define the (preliminary) tensorization for matrices $X \in \mathbb{R}^{n_1 \cdots n_d \times m_1 \cdots m_d}$ by*

$$(\mathcal{M}^*(X))_{(i_1,\ldots,i_d),(j_1,\ldots,j_d)} := X_{\ell_i,\ell_j}, \qquad \ell_i := i_1 + \sum_{\mu=2}^{d}(i_\mu - 1)\prod_{\nu=1}^{\mu-1} n_\nu,$$
$$\ell_j := j_1 + \sum_{\mu=2}^{d}(j_\mu - 1)\prod_{\nu=1}^{\mu-1} m_\nu,$$
$$\mathcal{M}^*(X) \in \mathbb{R}^{(n_1\times\cdots\times n_d)\times(m_1\times\cdots\times m_d)}.$$

Example 21 *Consider the matrix $X = \mathrm{diag}(1,2,3,4,5,6,7,8) \in \mathbb{R}^{2\cdot 2\cdot 2\times 2\cdot 2\cdot 2}$. The tensorization of X is a diagonal matrix $\mathcal{M}^*(X)$ with diagonal entries*

$$\mathrm{diag}(\mathcal{M}^*(X)) = \left[\begin{array}{cc|cc} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{array}\right]$$

or index-wise

$$\begin{array}{ll} \mathcal{M}^*(X)_{(0,0,0),(0,0,0)} = 1, & \mathcal{M}^*(X)_{(0,0,1),(0,0,1)} = 5, \\ \mathcal{M}^*(X)_{(1,0,0),(1,0,0)} = 2, & \mathcal{M}^*(X)_{(1,0,1),(1,0,1)} = 6, \\ \mathcal{M}^*(X)_{(0,1,0),(0,1,0)} = 3, & \mathcal{M}^*(X)_{(0,1,1),(0,1,1)} = 7, \\ \mathcal{M}^*(X)_{(1,1,0),(1,1,0)} = 4, & \mathcal{M}^*(X)_{(1,1,1),(1,1,1)} = 8. \end{array}$$

The tensorization is an isomorphism

$$\mathcal{M}^* : \mathbb{R}^{\prod_{\mu=1}^{d} n_\mu \times \prod_{\mu=1}^{d} m_\mu} \longrightarrow \mathbb{R}^{(\otimes_{\mu=1}^{d} n_\mu)\times(\otimes_{\mu=1}^{d} m_\mu)}.$$

The matrix structure is thus preserved by the tensorization, or, in other words, only the column (respectively row) vectors are tensorized.

The tensor format $\mathcal{M}^*(X)$ is reasonable for a theoretical treatment concerning the matrix properties, but it is more elegant to reorganize the modes in order to treat the tensor as before in the vector case.

Definition 22 (Matrix Tensorization) *We define the tensorization of a matrix $X \in \mathbb{R}^{n_1 \cdots n_d \times m_1 \cdots m_d}$ by*

$$\begin{aligned} \mathcal{M} \quad &: \quad \mathbb{R}^{n_1\cdots n_d \times m_1 \cdots m_d} \longrightarrow \mathbb{R}^{n_1 m_1 \times \cdots \times n_d m_d}, \\ (\mathcal{M}(X))_{i_1+(j_1-1)n_1),\ldots,i_d+(j_d-1)n_d} \quad &:= \quad X_{\ell_i,\ell_j}, \\ & \qquad \ell_i := i_1 + \sum_{\mu=2}^{d}(i_\mu - 1)\prod_{\nu=1}^{\mu-1} n_\nu, \\ & \qquad \ell_j := j_1 + \sum_{\mu=2}^{d}(j_\mu - 1)\prod_{\nu=1}^{\mu-1} m_\nu. \end{aligned}$$

For the tensor $\mathcal{M}(X)$ we can apply the standard hierarchical Tucker approximation. The ranks for the representation are defined as above by

$$k_t := \mathrm{rank}_t(\mathcal{M}(X)) := \mathrm{rank}(\mathcal{M}(X)^{(t)}).$$

For a node $t = \{1, \dots, q\}$ we obtain a matrix with columns

$$\mathcal{M}(X)^{(1,\dots,q)}_{\ell} = X|_{I_{\ell_1,q} \times J_{\ell_2,q}},$$

$$\ell_1 \in \{1, \dots, N'_q\}, \quad N'_q = \prod_{\mu=q+1}^{d} n_\mu, \qquad \ell_2 \in \{1, \dots, M'_q\}, \quad M'_q = \prod_{\mu=q+1}^{d} m_\mu,$$

where the row and column indices are

$$\begin{aligned} I_{\ell_1,q} &:= \{1 + (\ell_1 - 1)N_q, \dots, \ell_1 N_q\} \\ J_{\ell_2,q} &:= \{1 + (\ell_2 - 1)M_q, \dots, \ell_2 M_q\}. \end{aligned}$$

The same techniques and arguments as for the vector case apply. In the following we give some simple examples.

Example 23 (1d Laplace) *The one-dimensional Laplacian discretized by finite elements or finite differences on an equispaced grid with 2^d grid points leads (up to scaling) to a tridiagonal system matrix X with diagonal entries $X_{i,i} = 2$ and sub- and super-diagonal entries $X_{i,i-1} = X_{i-1,i} = -1$. For any partition of the matrix into $2^q \times 2^q$ tiles there appear only four different patterns: The block can be zero, it can have a single entry in the lower left or upper right corner, or it can be a block on the diagonal with the same tridiagonal structure. For the rank only the latter three block types are relevant, so the rank of a matricization of $\mathcal{M}(X)$ is always three (except when $N_q < 3$ or $N'_q < 3$).*

Example 24 (2d Laplace) *The two-dimensional Laplacian discretized on an equispaced grid with $n \times n$, $n := 2^d$ grid points leads (up to scaling) to a system matrix X with diagonal entries $X_{i,i} = 4$, sub- and super-diagonal entries $X_{i,i-1} = X_{i-1,i} = -1$ and off-diagonal entries $X_{i,i-n} = X_{i-n,i} = -1$. For any partition of the matrix into $2^q \times 2^q$ tiles there appear only five different patterns: The block can be zero, it can have a single entry in the lower left or upper right corner, it can be a block on the diagonal with the same five point stencil structure, or it can be an off-diagonal coupling in the lower or upper half which is the same diagonal matrix block in both cases. For the rank only the latter four block types are relevant, so the rank of a matricization of $\mathcal{M}(X)$ is always four (except when $N_q < 4$ or $N'_q < 4$).*

Example 25 (Hilbert matrix) *The Hilbert matrix has entries*

$$X_{i,j} = 1/(i - j + 0.5)$$

and thus allows for a polynomial approximation in blocks that are separated from the diagonal. On the diagonal the blocks are all identical, and the same holds for the sub- and super-diagonal blocks. Thus the rank is bounded by $3 + p^2$, where $p \sim \log(1/\varepsilon)$ for a point-wise accuracy of ε.

6.2 Tensor Tensorization

In principle, there is no restriction to extend the formalism also to higher order tensors, i.e., to tensorize a tensor

$$X \in \mathbb{R}^{\otimes_{\mu=1}^{d_1} \prod_{\nu=1}^{d_2} n_{\mu,\nu}}$$

into a tensor

$$\mathcal{T}(X) \in \mathbb{R}^{\otimes_{\nu=1}^{d_2} \prod_{\mu=1}^{d_1} n_{\mu,\nu}}$$

which might make sense when d_1 is small compared to d_2, or if an additional structure is present.

Definition 26 (Tensor Tensorization) *We define the tensorization for tensors* $X \in \mathbb{R}^{\otimes_{\mu=1}^{d_1} \prod_{\nu=1}^{d_2} n_{\mu,\nu}}$ *by*

$$\mathcal{T}^*(X) \in \mathbb{R}^{\otimes_{\nu=1}^{d_2} \otimes_{\mu=1}^{d_1} n_{\mu,\nu}},$$

$$(\mathcal{T}^*(X))_{(i_1^1,\dots,i_1^{d_1}),\cdots,(i_{d_2}^1,\dots,i_{d_2}^{d_1})} := X_{\ell_1,\cdots,\ell_{d_1}}, \qquad \ell_\eta := i_1^\eta + \sum_{\mu=2}^{d_2} (i_\mu^\eta - 1) \prod_{\nu=1}^{\mu-1} n_{\mu,\nu}.$$

The corresponding $\mathcal{H}$-Tucker *compatible tensorization is given by*

$$\mathcal{T}(X) \in \mathbb{R}^{\otimes_{\nu=1}^{d_2} \prod_{\mu=1}^{d_1} n_{\mu,\nu}},$$

$$(\mathcal{T}(X))_{\ell_1,\dots,\ell_{d_2}} := (\mathcal{T}^*(X))_{(i_1^1,\dots,i_1^{d_1}),\cdots,(i_{d_2}^1,\dots,i_{d_2}^{d_1})}, \qquad \ell_\eta = i_\eta^1 + \sum_{\mu=2}^{d_1} (i_\eta^\mu - 1) \prod_{\nu=1}^{\mu-1} n_{\nu,\eta}.$$

6.3 General Vectors

In principle one could be interested in a tensorization of a general vector $x \in \mathbb{R}^N$, where N is not necessarily a product of many small numbers but, e.g., a prime number. Formally, this can be easily overcome by the definition of an extended vector,

$$\bar{x}_i := \begin{cases} x_i & i \le N \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, \dots, \prod_{\mu=1}^{d} n_\mu \ge N.$$

Theorem 27 ($\mathcal{H}$-rank of tensorized extended vectors) *Let* $\mathcal{F}^{ex}(x) \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *be the tensorization of the extension* $\bar{x}$ *of* $x \in \mathbb{R}^N$ *and for* $t = \{1, \dots, q\}$ *let* $N_q := \prod_{\mu=1}^{q} n_\mu$*, and* $N_q' := \prod_{\mu=q+1}^{d} n_\mu$*. Then the rank of the matricization* $\mathcal{F}^{ex}(x)^{(t)}$ *is*

$$\begin{aligned} \operatorname{rank}(\mathcal{F}^{ex}(x))^{(t)} &\in \{k, k+1\}, \\ k &:= \dim\left(\operatorname{span}\{x|_{I_{\ell,q}} \mid \ell = 1, \dots, N_q', \quad I_{\ell,q} \subset \{1, \dots, N\}\}\right), \\ I_{\ell,q} &:= \{1 + (\ell - 1)N_q, \dots, \ell N_q\}.\}. \end{aligned}$$

Proof: The index sets all fulfil $I_{\ell,q} \cap \{1, \dots, N\} \in \{\emptyset, \{1, \dots, N\}\}$ except for a single index ℓ. Thus the column rank is the column rank of the matrix where the ℓ-th column is eliminated, plus one if the eliminated column is linearly independent of the others. ∎

We want to remark that the statement of Theorem 27 is not true for the general $\mathcal{H}$-Tucker format with clusters $t = \{r + 1, \dots, s\}$. There the rank can only be bounded by $2k + 1$.

Corollary 28 (Polynomial approximation) *Let* x *be as in Corollary 13 (or Theorem 18). Then the hierarchcial rank of* $\mathcal{F}(\bar{x})$ *in the TT-format is bounded by* $p + 2$ *(or* $C + \log_2(1/\varepsilon) + 2m$*, respectively) for any extension* $\bar{x}$ *of* x*. In the general* $\mathcal{H}$*-Tucker format the rank is bounded by* $2p + 3$ *(or* $C + 2\log_2(1/\varepsilon) + 4m$*, respectively).*

7 Conclusion

We conclude that a large class of vectors that stem from the evaluation of a polynomial or asymptotically smooth function on an equispaced grid, allows for a tensorization with small hierarchical rank k. A truncated arithmetic for such tensorized vectors is possible in $\mathcal{O}(k^4 \log N)$, where N is the length of the vector. This holds independently of the number N, i.e., whether N can be factorized or not is not relevant. However, the requirement that the grid is equispaced (or similarly structured) is necessary.

References

[1] Börm, S., Löhndorf, M., Melenk, J.M.: Approximation of integral operators by variable-order interpolation. Numerische Mathematik **99**(4), 605–643 (2005)

[2] Carroll, J.D., J.Chang, J.: Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. Psychometrika **35**, 283–319 (1970)

[3] de Silva, V., Lim, L.H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. **30**, 1084–1127 (2008)

[4] Grasedyck, L.: Hierarchical singular value decomposition of tensors. SIAM J. Matrix Anal. Appl. (2010). Accepted

[5] Hackbusch, W., Kühn, S.: A new scheme for the tensor representation. J. Fourier Anal. Appl. **15**, 706–722 (2009)

[6] Harshman, R.A.: Foundations of the PARAFAC procedure. UCLA Working Papers in Phonetics **16**, 1–84 (1970)

[7] Khoromskij, B.: $\mathcal{O}(d \log N)$-Quantics approximation of N-d tensors in high-dimensional numerical modeling. Preprint 55, Max Planck Institute for Mathematics in the Sciences, Leipzig (2009)

[8] Kruskal, J.B.: Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Appl. **18**, 95–138 (1977)

[9] Oseledets, I.: Compact matrix form of the d-dimensional tensor decomposition. Preprint 09-01, Institute of Numerical Mathematics RAS, Moscow, Russia (2009)

[10] Oseledets, I.: Tensors inside of matrices give logarithmic complexity. Preprint 09-04, Institute of Numerical Mathematics RAS, Moscow, Russia (2009)

[11] Oseledets, I., Khoromskij, B.: Quantics-TT approximation of elliptic solution operators in higher dimensions. Preprint 79, Max Planck Institute for Mathematics in the Sciences, Leipzig (2009)

[12] Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**, 279–311 (1966)