Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Black Box Approximation of Tensors in Hierarchical Tucker Format

(revised version: October 2010)

by

Jonas Ballani, Lars Grasedyck, and Melanie Kluge

Preprint no.: 57 2010

Black Box Approximation of Tensors in Hierarchical Tucker Format

Jonas Ballani [*], Lars Grasedyck, Melanie Kluge [†]

October 1, 2010

We derive and analyse a scheme for the approximation of order d tensors $A \in \mathbb{R}^{n\times\cdots\times n}$ in the hierarchical ($\mathcal{H}$-) Tucker format, a dimension-multilevel variant of the Tucker format and strongly related to the TT format. For a fixed rank parameter k, the storage complexity of a tensor in $\mathcal{H}$-Tucker format is $\mathcal{O}(dk^3 + dnk)$ and we present a (heuristic) algorithm that finds an approximation to a tensor in the $\mathcal{H}$-Tucker format in $\mathcal{O}(dk^4 + d\log(d)nk^2)$ by inspection of only $dk^3 + d\log(d)nk^2$ entries. Under mild assumptions, tensors in the $\mathcal{H}$-Tucker format are reconstructed. For general tensors we derive error bounds that are based on the approximability of matrices (matricizations of the tensor) by few outer products of its rows and columns. The construction parallelizes with respect to the order d and we also propose an adaptive approach that aims at finding the rank parameter for a given target accuracy ε automatically.

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

1 Introduction

In this article we provide a heuristic that aims at finding — by inspecting only a few entries of A — an accurate representation of a tensor

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

in a data-sparse low rank tensor format, namely the hierarchical ($\mathcal{H}$-) Tucker format [6, 5]. The tensor A could, e.g., be given by a multivariate function $f : [0,1]^d \to \mathbb{R}$ on a tensor grid:

$$A_{i_1,\ldots,i_d} = f(\xi_{i_1},\ldots,\xi_{i_d}), \qquad \xi_i \in [0,1].$$

*Max Planck Institute for Mathematics in the Sciences, Inselstraße 22–26, D-04109 Leipzig, Germany.
†Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 52056 Aachen, Germany. Email: {lgr,kluge}@igpm.rwth-aachen.de. Financial support from the DFG SPP-1324 under grant GRA2179/2-1 gratefully acknowledged.

For the tensor A, respectively the function f, we require that there exists a low (hierarchical) rank approximation — which is unknown to us. Our goal is to find, i.e. construct, the data-sparse low rank representation without forming the whole tensor. In almost all practical applications where tensors arise it is not known a priori how a low rank representation should be obtained, and thus our approximation scheme can be regarded as a general tool for the conversion into the $\mathcal{H}$-Tucker format. Once the tensor is in the $\mathcal{H}$-Tucker format, one can perform standard arithmetic operations with it in a complexity $\mathcal{O}(dk^4 + k^2 \sum_{\mu=1}^{d} n_\mu)$, where k is an internal rank parameter.

A crucial part for the approximation is to determine the necessary entries of the tensor A that have to be inspected, the so-called pivot elements. We provide an adaptive and incremental construction alongside with an error estimate and stopping criterion.

In [10] a cross approximation algorithm is proposed for the TT format [9, 8] which is a special subclass of the $\mathcal{H}$-Tucker format. Our results can be regarded as an extension to the $\mathcal{H}$-Tucker format, but we also provide an adaptive pivoting strategy, an incremental construction as well as error bounds for the approximation (under reasonable assumptions).

In Section 2 we introduce the hierarchical Tucker format [6, 5] and the corresponding hierarchical rank. In Section 3 we present the hierarchical black box approximation scheme as well as the adaptive pivoting strategy. In Section 4 we derive a priori error bounds and a plausible but heuristic bound. In the last Section 5 we provide a number of numerical examples and counterexamples that underline the efficiency of the method as well as its limitations.

2 The Hierarchical Tucker Format

In the hierarchical Tucker format, the sparsity of the representation of a tensor is determined by the hierarchical rank which is the rank of certain matricizations of the tensor. For the rest of the article we use the notation for index sets

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

2.1 Definition of the $\mathcal{H}$-Tucker Format

Definition 1 (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 (cf. Figure 1)*

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

is defined by its entries

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

Example 2 *The matricizations of the tensor*

$$A_{i_1,i_2,i_3,i_4} := i_1 + 2(i_2 - 1) + 4(i_3 - 1) + 8(i_4 - 1), \qquad i_1, i_2, i_3, i_4 \in \{1, 2\}$$

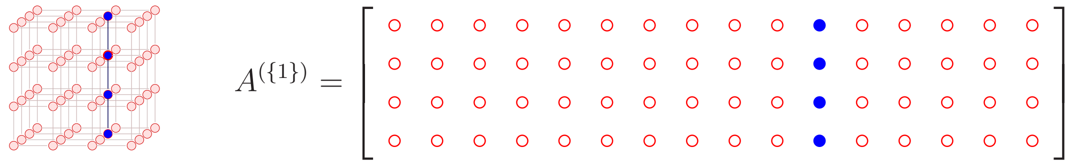

Figure 1: Left: Position of entries of a $4 \times 4 \times 4$ tensor A. Right: Matricization $A^{(\{1\})}$ for the node $t = \{1\}$. The blue dots form a column in the matricization.

are

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

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

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

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.

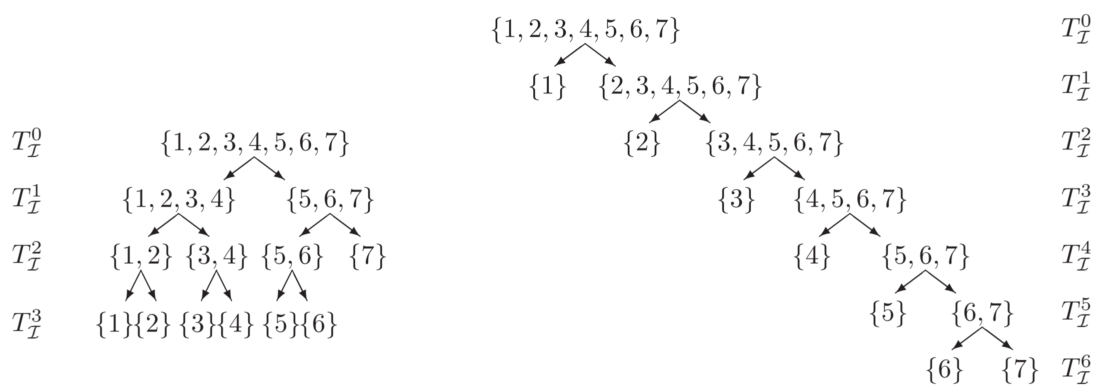

Figure 2: Left: The canonical dimension tree. Right: The degenerate TT-tree.

Definition 3 (Dimension tree) *A dimension tree or mode cluster tree* $T_{\mathcal{I}}$ *for dimension* $d \in \mathbb{N}$ *is a tree with root* $D := \{1, \ldots, d\}$ *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* $S(t) = \{t_1, t_2\}$*:*

$$t = t_1 \dot{\cup} t_2. \tag{1}$$

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 2. We denote the level ℓ of the tree by

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

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 4 (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*

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

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*

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

Remark 5 (Canonical dimension tree and TT-tree) *In the hierarchical format only some of the possible subsets t of all modes appear. A special case is the so-called TT-format [8] with corresponding TT-rank and TT-tree, where all nodes are of the form (cf. Figure 2)*

$$t = \{q\} \quad \textit{or} \quad t = \{q, \dots, d\}, \qquad q = 1, \dots, d.$$

This tree is of maximal depth (cf. [5, Section 5] for a discussion).

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 (cf. Figure 2)

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

2.2 Arithmetics in the $\mathcal{H}$-Tucker Format

The storage complexity for a tensor in the (CP) format (cf. [7] and the references therein) with rank k, mode size n and order d is $\mathcal{O}(dkn)$. In the $\mathcal{H}$-Tucker format the storage complexity will turn out to be roughly one factor k higher.

Lemma 6 (Hierarchical Tucker format, [5]) *Let $T_{\mathcal{I}}$ be a dimension tree and let $A \in \mathcal{H}\text{-Tucker}((k_t)_{t\in T_{\mathcal{I}}})$. 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 in terms of number of entries for B_t, U_t from the previous lemma is

$$\text{Storage}((B_t)_{t\in\mathcal{I}(T_{\mathcal{I}})}, (U_t)_{t\in\mathcal{L}(T_{\mathcal{I}})}) \ \le \ (d-1)k^3 + k\sum_{\mu=1}^{d} n_\mu, \qquad k := \max_{t\in T_{\mathcal{I}}} k_t, \tag{2}$$

i.e. linear in the order d (provided that k is uniformly bounded) [5].

Notation 7 *For a matrix $U \in \mathbb{R}^{I\times J}$ we denote by a single subscript $U_j \in \mathbb{R}^I$ the j-th column of the matrix. If this is ambiguous then the notation $U_{\cdot,j}$ or $U|_{I\times\{j\}}$ is used.*

For each of the interior nodes the corresponding mode frame is implicitly given by the following nestedness property:

$$(U_t)_i = \sum_{j=1}^{k_{t_1}} \sum_{\ell=1}^{k_{t_2}} (B_t)_{i,j,\ell} (U_{t_1})_j \otimes (U_{t_2})_\ell \qquad (3)$$

In particular the tensor A is given by $A = U_D$ for the root D. In the following section we explain how the mode frames U_t for the leaves t and the transfer tensors B_t for the interior nodes can be constructed efficiently.

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 [5].

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,

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

and it can be computed in

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

For the proof we refer to [5, 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 (proportionality factor $\sqrt{2d-3}$).

3 Black Box Approximation and Pivoting

The approximation of tensors in the $\mathcal{H}$-Tucker format is based on two concepts: One has to approximate the matricizations $A^{(t)}$ (Definition 1) by low rank k_t, and one has to ensure that the approximations are nested (3).

3.1 Approximation of Matricizations

In [4] the approximability of matrices by outer products of some of its columns and rows is analysed. The main theorem states that for every matrix $A^{(t)} \in \mathbb{R}^{\mathcal{I}_t\times\mathcal{I}_{t'}}$ and every rank k approximation

$$\|A^{(t)} - R\|_2 \le \varepsilon, \qquad \text{rank}(R) \le k,$$

there exist k row indices $P_t = \{p_1^{(t)}, \cdots, p_k^{(t)}\} \subset \mathcal{I}_t$ and column indices $P_{t'} = \{q_1^{(t')}, \cdots, q_k^{(t')}\} \subset \mathcal{I}_{t'}$ and a matrix S such that

$$\tilde{A}^{(t)} := A^{(t)}|_{\mathcal{I}_t \times P_{t'}} \cdot S^{-1} \cdot A^{(t)}|_{P_t \times \mathcal{I}_{t'}}, \qquad S \in \mathbb{R}^{P_t \times P_{t'}},$$

approximates the whole matrix with an error of the size

$$\|A^{(t)} - \tilde{A}^{(t)}\|_2 \leq \varepsilon \left(1 + 2\sqrt{k}\left(\sqrt{\#\mathcal{I}_t} + \sqrt{\#\mathcal{I}_{t'}}\right)\right).$$

A practical construction based on successive rank one approximations is given in [1]. The

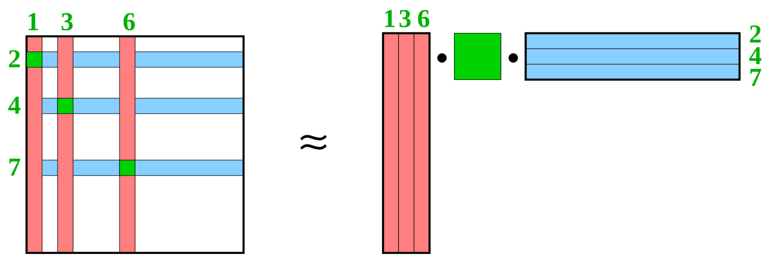

Figure 3: Rows and columns of $A^{(t)}$ are used for a low rank approximation.

idea is to construct rank one approximations of the remainder:

$$\begin{aligned} X^1 &:= A^{(t)}_{\cdot,q_1} \frac{1}{A^{(t)}_{p_1,q_1}} A^{(t)}_{p_1,\cdot}, \\ X^j &:= X^{j-1} + (A^{(t)} - X^{j-1})_{\cdot,q_j} \frac{1}{(A^{(t)} - X^{j-1})_{p_j,q_j}} (A^{(t)} - X^{j-1})_{p_j,\cdot}, \quad j = 2, \ldots, k. \end{aligned} \tag{4}$$

The final approximation is given by $\tilde{A}^{(t)} := X^k$. In the form above, the matrix S is

$$S = A^{(t)}|_{P_t \times P_{t'}}, \qquad P_t := \{p_1, \ldots, p_k\}, \quad P_{t'} := \{q_1, \ldots, q_k\}.$$

In principle, we use exactly this approximation scheme with three necessary modifications:

1. The column and row vectors can only be formed when the index sets $\mathcal{I}_t, \mathcal{I}_{t'}$ are small, i.e., when t, t' are leaves of the tree. Whereas in [10] this is done by fixing an initial guess of all pivot elements and then optimizing them with respect to one mode at a time, we will choose an entirely different approach.

2. The pivot elements are best chosen such that S has maximal volume. Since this is in general practically impossible, one instead chooses the pivot indices p_j, q_j such that the remainder $\left|(A^{(t)} - X^{j-1})_{p_j,q_j}\right|$ is maximized over the whole matrix (full pivoting), or a column/row (partial pivoting), respectively. We will use a maximization over crosses of the tensor as it was introduced in [3, Algorithm 2: Greedy Initial Pivot Search].

3. The choice of possible row and column pivots is restricted in order to ensure the nestedness property.

3.2 Choice of Pivot Elements

In the incremental construction (4), we will choose pivot elements (p_j, q_j) from a restricted set (required in Subsection 3.3)

$$p_j \in P_t \subset \mathcal{I}_t, \qquad q_j \in P_{t'} = \mathcal{I}_{\bar{t}} \times P_{f'} \subset \mathcal{I}_{t'},$$

where $P_{f'} = \{q_1^{(f')}, \dots, q_k^{(f')}\}$ is the set of column pivot indices for the father f of t and $\bar{t}$ ($S(f) = \{t, \bar{t}\}$).

In each step of construction (4) we thus aim at finding

$$(p_j, q_j) := \operatorname*{argmax}_{p_j \in \mathcal{I}_t, q_j \in \mathcal{I}_{\bar{t}} \times P_{f'}} \left| (A^{(t)} - X^{j-1})_{p_j, q_j} \right|.$$

This is done by a simple greedy search in the entries of the remainder (similar to Algorithm 2 in [3]) starting at random entries and then looking for entries by varying only one component at a time. The procedure is given in Algorithm 1.

Algorithm 1 Greedy Pivot Search

Given: an initial index $(i_\mu)_{\mu \in D}$, $D := \{1, \dots, d\}$, a subset $\mathcal{I}_{\bar{t}} \times P_{f'} \subset \mathcal{I}_{t'}$, $t' = \{1, \dots, d\} \setminus t$, a tensor $A \in \mathbb{R}^{\mathcal{I}}$ and an approximation X^{j-1} of A.

for $\ell = 1, 2, \dots, \ell_{max}$ (typically $\ell_{max} := 3$) **do**

  for $\mu \in f = t \cup \bar{t}$ **do**

    Modify the μ-th index by

$$i_\mu := \operatorname*{argmax}_{i_\mu \in \{1, \dots, n_\mu\}} \left| (A - X^{j-1})_{(i_1, \dots, i_d)} \right|$$

  end for

  Modify the remaining indices in $f' = D \setminus f$ by

$$(i_\mu)_{\mu \in f'} := \operatorname*{argmax}_{(i_\mu)_{\mu \in f'} \in P_{f'}} \left| (A - X^{j-1})_{(i_1, \dots, i_d)} \right|$$

end for

Return value: the pivot index $(i_1, \dots, i_d)$.

3.3 Ensuring Nestedness

Let t be a node with sons $S(t) = \{t_1, t_2\}$. We denote the three corresponding approximations of the matricizations by

$$N_t := A_t S_t^{-1} A_t', \qquad A_t := A^{(t)}|_{\mathcal{I}_t \times P_{t'}},\ S_t := A^{(t)}|_{P_t \times P_{t'}},\ A_t' := A^{(t)}|_{P_t \times \mathcal{I}_{t'}}, \tag{5}$$

$$N_{t_1} := A_{t_1} S_{t_1}^{-1} A_{t_1}', \qquad A_{t_1} := A^{(t_1)}|_{\mathcal{I}_{t_1} \times P_{t_1'}},\ S_{t_1} := A^{(t_1)}|_{P_{t_1} \times P_{t_1'}},\ A_{t_1}' := A^{(t_1)}|_{P_{t_1} \times \mathcal{I}_{t_1'}}, \tag{6}$$

$$N_{t_2} := A_{t_2} S_{t_2}^{-1} A_{t_2}', \qquad A_{t_2} := A^{(t_2)}|_{\mathcal{I}_{t_2} \times P_{t_2'}},\ S_{t_2} := A^{(t_2)}|_{P_{t_2} \times P_{t_2'}},\ A_{t_2}' := A^{(t_2)}|_{P_{t_2} \times \mathcal{I}_{t_2'}}. \tag{7}$$

In the leaves t we set

$$(U_t)_i := A^{(t)}|_{\cdot,q_i}, \quad q_i \in P_{t'}.$$

For all interior nodes we define

$$(B_t)_{i,j,\ell} := \sum_{p\in P_{t_1}} \sum_{q\in P_{t_2}} (S_{t_1}^{-1})_{q_j,p} A^{(t)}_{(p,q),q_i} (S_{t_2}^{-1})_{q_\ell,q}, \qquad q_i \in P_{t'}. \tag{8}$$

The construction breaks down if a non-zero pivot element could not be determined (the matrix S becomes singular). This can either be the case when the tensor is of rank $j-1$ and the remainder zero, or it could be that there are non-zero elements that we could not find. For a discussion and counterexamples see, e.g. [2].

The construction (4) with the greedy pivot search and setup of the transfer tensors (8) is adaptive in the sense that the size of the pivot element gives an estimate for the $\|\cdot\|_\infty$ norm of the remainder. Also, one can update an approximation that is not accurate enough by continuation of the pivot search — independently for each node t. This is a difference to the procedure in [9], where the rank distribution has to be determined or guessed in advance. In the numerical tests we will see that the (heuristic) error estimation gives good results. In the following section we will estimate the approximation error under stronger assumptions.

Lemma 8 (Complexity) *The complexity N_{setup} for the setup of the mode frames U_t for the leaves $t \in \mathcal{L}(T_\mathcal{I})$ and the transfer tensors B_t for the interior nodes $t \in \mathcal{I}(T_\mathcal{I})$ is*

$$N_{setup} = \mathcal{O}\left(dk^4 + \log(d)k^2 \sum_{\mu=1}^{d} n_\mu\right), \qquad k := \max_{t\in T_\mathcal{I}} k_t.$$

The number of entries $N_{entries}$ required from A is of the size

$$N_{entries} = \mathcal{O}\left(dk^3 + \log(d)k^2 \sum_{\mu=1}^{d} n_\mu\right).$$

Proof: (N_{setup}): Once the pivot elements are determined, it is clear that the setup of all U_t for leaves t is of complexity $\sum_{\mu=1}^{d} n_\mu k_\mu$ (copy of entries from A to U_t). For the transfer tensors the $k_t k_{t_1} k_{t_2}$ entries of $A^{(t)}_{(p,q),q_i}$ are transformed by the inverses $S_{t_1}^{-1}, S_{t_2}^{-1}$ which in sum provides the first part of the claimed estimate. The setup of S_t^{-1} is of complexity at most $\mathcal{O}(k_t^3)$, and this is required k_t times ($j = 1,\ldots,k_t$). In total, this is of complexity $\mathcal{O}(k_t^4)$, which gives again the first term in the estimate.

The evaluation of X^{j-1} in a single point (i,ℓ) is of complexity $k_t^2 \sum_{\mu\in t} n_\mu$ because the low rank form allows the evaluation of the form

$$\tilde{A}^{(t)}_{i,\ell} = (A_t)_{i,\cdot} S_t^{-1} (A'_t)_{\cdot,\ell}.$$

When one of the two indices is fixed and we compute entries, e.g., for several indices i, then $S_t^{-1}(A'_t)_{\cdot,\ell}$ can be precomputed in $\mathcal{O}(k_t^2)$. Afterwards, each evaluation is of complexity

$k_t \sum_{\mu \in t} n_\mu$. For all $j = 1, \ldots, k$ and all nodes of the tree this sums up to the second term in the claimed estimate:

$$\begin{aligned}\sum_{t \in T_\mathcal{I}} \left(k_t^2 \sum_{\mu \in t} n_\mu + k_t^3 \right) &\leq dk^3 + k^2 \sum_{t \in T_\mathcal{I}} \sum_{\mu \in t} n_\mu \quad \leq dk^3 + k^2 \sum_{\ell=0}^{\mathrm{depth}(T_\mathcal{I})} \sum_{t \in T_\mathcal{I}^\ell} \sum_{\mu \in t} n_\mu \\ &\leq dk^3 + k^2 \sum_{\ell=0}^{\mathrm{depth}(T_\mathcal{I})} \sum_{\mu=1}^{d} n_\mu \quad \leq dk^3 + k^2(\log(d) + 1) \sum_{\mu=1}^{d} n_\mu\end{aligned}$$

$(N_{entries})$: The entries required for the setup of U_t, B_t are $\mathcal{O}(\sum_{\mu=1}^d n_\mu k_\mu + dk^3)$. For the pivot search we have in each node t a number of evaluations $\mathcal{O}(\sum_{\mu \in t} n_\mu k_t^2)$ (k_t iterations), and summing this up over all nodes gives $\mathcal{O}(\log(d) k^2 \sum_{\mu=1}^d n_\mu)$. ∎

4 A Priori Error Estimation

If we would approximate each matricization $A^{(t)}$ independently by a skeleton decomposition with error ε, then form the orthogonal factor Q_t,

$$Q_t R_t = A^{(t)}|_{\mathcal{I}_t \times P_{t'}},$$

and use $Q_t Q_t^T$ as orthogonal projector

$$(\pi_t A)^{(t)} := Q_t Q_t^T A^{(t)},$$

then [5, Theorem 3.11, Remark 3.12] shows that the projection

$$A_Q := \left(\prod_{t \in T_\mathcal{I}} \pi_t \right) A$$

fulfills

$$\|A - A_Q\| \leq \sqrt{2d-3}\, \varepsilon, \qquad A_Q \in \mathcal{H}\text{-Tucker}((k_t)_{t \in T_\mathcal{I}}).$$

The exact projection with the orthogonal factors, however, is practically impossible. Instead, we try to approximate

$$A^{(t)}|_{\cdot,i} \approx \sum_{j=1}^{k_{t_1}} \sum_{\ell=1}^{k_{t_2}} A^{(t_1)}|_{\cdot,q_j} \otimes A^{(t_2)}|_{\cdot,q_\ell} \cdot (B_t)_{i,j,\ell},$$

where B_t was defined in (8). We will now estimate the error of this approximation step.

From [2] it is well known that the pivot search can fail for matrices. Here, we will assume that the pivot search is successful for all matricizations.

Assumption 9 *We assume that for every matricization $A^{(t)}$ the pivot search (Algorithm 1) is almost optimal, i.e., for some $\delta \geq 0$*

$$|(A^{(t)} - X^{j-1})_{p_j,q_j}| \geq \frac{1}{1+\delta} \|A^{(t)} - X^{j-1}\|_\infty, \qquad j = 1, \ldots, k$$

Further, we assume that for each matricization the error of the non-nested approximation is bounded by ε:

$$\|A^{(t)} - X^{k_t}\|_\infty \leq \varepsilon$$

We use the same notation as in the previous section. We define the defect matrices

$$F_t := (A^{(t)} - N_t)|_{\mathcal{I}_t \times P_{t'}}, \quad F_{t_1} := (A^{(t_1)} - N_{t_1})|_{\mathcal{I}_{t_1} \times (\mathcal{I}_{t_2} \times P_{t'})}, \quad F_{t_2} := (A^{(t_2)} - N_{t_2})|_{\mathcal{I}_{t_2} \times (\mathcal{I}_{t_1} \times P_{t'})}.$$

Lemma 10 *Let N_{t_1}, N_{t_2} from (6),(7) be approximations of $A^{(t_1)}, A^{(t_2)}$ with defect F_{t_1}, F_{t_2}. Then for $i_1 \in \mathcal{I}_{t_1}, i_2 \in \mathcal{I}_{t_2}$ and $q \in P_{t'}$ holds*

$$A^{(t)}_{(i_1,i_2),q} = \sum_{j=1}^{k_{t_1}} \sum_{\ell=1}^{k_{t_2}} A^{(t_1)}_{i_1,q^1_j} \cdot A^{(t_2)}_{i_2,q^2_\ell} \cdot (B_t)_{q,j,\ell} \quad + (F_t)_{(i_1,i_2),q} \qquad \left(q^1_j \in P_{t'_1}, q^2_\ell \in P_{t'_2}\right)$$

for the defect tensor

$$(F_t)_{(i_1,i_2),q} := \sum_{\alpha=1}^{k_{t_1}} (A_{t_1} S^{-1}_{t_1})_{i_1,\alpha} (F_{t_2})_{i_2,(\alpha,q)} + (F_{t_1})_{i_1,(i_2,q)}$$

Proof: First, we approximate $A^{(t)}_{(i_1,i_2),q}$ by $(N_{t_1})_{i_1,(i_2,q)}$ with error $(F_{t_1})_{i_1,(i_2,q)}$, then we approximate $(A'_{t_1})_{\alpha,(i_2,q)}$ by $(N_{t_2})_{i_2,(\alpha,q)}$ with error $(F_{t_2})_{i_2,(\alpha,q)}$. Both together yield the stated error. ■

In the defect tensor F_t each entry is possibly amplified by $A_{t_1} S^{-1}_{t_1}$. This factor is estimated next.

Lemma 11 *Let A_{t_1}, S_{t_1} denote the factors from (6). Then for $i_1 \in \mathcal{I}_{t_1}$ and $\alpha \in P_{t_1}$ holds*

$$|(A_{t_1} S^{-1}_{t_1})_{i_1,\alpha}| \leq \delta_k := k(1+\delta)^2(2+\delta)^{k-2}.$$

Proof: The pivot elements are denoted by p_j, q_j. We introduce the notation

$$G^j_\cdot := -(A^{(t_1)} - X^{j-1})^{-1}_{p_j,q_j} (A^{(t_1)} - X^{j-1})_{p_j,\cdot}$$

and observe that due to Assumption 9: $\|G^j\|_\infty \leq 1 + \delta$.

The first approximation step is

$$X^1 = A^{(t_1)}_{\cdot,q_1} (A^{(t_1)}_{p_1,q_1})^{-1} A^{(t_1)}_{p_1,\cdot} = \left[A^{(t_1)}_{\cdot,q_1}\right] [1] \left[-G^1\right]^T.$$

One can in a straight-forward way derive that the approximation after k steps is of the form

$$X^k = \left[A^{(t_1)}_{\cdot,q_1} \,|\, A^{(t_1)}_{\cdot,q_2} \,|\, \dots \,|\, A^{(t_1)}_{\cdot,q_k}\right] \ \left[M_0 \,|\, M_1 \,|\, M_2 \,|\, \cdots \,|\, M_{k-1}\right] \ \left[-G^1 \,|\, \cdots \,|\, -G^k\right]^T$$

with vectors

$$M'_1(l) := \left[\ G^1_{q_l}\ \right] \in \mathbb{R}^1, \dots, \quad M'_{k-1}(l) := \left[\begin{array}{c} M'_{k-2}(l) + M'_{k-2}(k-1) G^{k-1}_{q_l} \\ G^{k-1}_{q_l} \end{array}\right] \in \mathbb{R}^{k-1}, l \leq k$$

and their extensions to vectors of length k by appending a one and zeros:

$$M_0 := \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^k, \quad M_j := \begin{bmatrix} M'_j(j+1) \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^k, \quad M_{k-1} := \begin{bmatrix} M'_{k-1}(k) \\ 1 \end{bmatrix} \in \mathbb{R}^k.$$

We use the notation $Y^k := M_{k-1}(-G^k)^T$ and show per induction

$$\left\|Y^k\right\|_\infty \leq (1+\delta)^2(2+\delta)^{k-2}. \tag{9}$$

From Assumption 9 we derive the maximum norm bound

$$\left\|Y^k\right\|_\infty = \|M_{k-1}\|_\infty \left\|G^k\right\|_\infty \leq (1+\delta)\,\|M_{k-1}\|_\infty .$$

For $k=3$ we can directly estimate

$$\begin{aligned} \left\|Y^3\right\|_\infty &\leq (1+\delta)\,\|M_2\|_\infty = (1+\delta)\left\| \begin{bmatrix} G^1_{q_3} + G^1_{q_2}G^2_{q_3} \\ G^2_{q_3} \\ 1 \end{bmatrix} \right\|_\infty \\ &\leq (1+\delta)(\left|G^1_{q_3}\right| + \left|G^1_{q_2}G^2_{q_3}\right|) \leq (1+\delta)\big((1+\delta)+(1+\delta)^2\big) = (1+\delta)^2(2+\delta). \end{aligned}$$

By induction the assertion (9) follows from

$$\begin{aligned} \left\|Y^k\right\|_\infty &\leq (1+\delta)\,\|M_{k-1}\|_\infty = (1+\delta)\left\| \begin{bmatrix} M'_{k-2}(k) + M'_{k-2}(k-1)G^{k-1}_{q_k} \\ G^{k-1}_{q_k} \\ 1 \end{bmatrix} \right\|_\infty \\ &\leq (1+\delta)\left(\left\| \begin{bmatrix} M'_{k-2}(k) \\ 0 \\ 0 \end{bmatrix} \right\|_\infty + \left\| \begin{bmatrix} M'_{k-2}(k-1)G^{k-1}_{q_k} \\ G^{k-1}_{q_k} \\ 1 \end{bmatrix} \right\|_\infty \right) \\ &\leq (1+\delta)\big((1+\delta)(2+\delta)^{k-3} + (1+\delta)^2(2+\delta)^{k-3}\big) = (1+\delta)^2(2+\delta)^{k-2} \end{aligned}$$

The final assertion follows from (9):

$$\begin{aligned} \left\|A_{t_1}S^{-1}_{t_1}\right\|_\infty &= \left\| \Big[M_0 \,\big|M_1 \,\big|M_2 \,\big|\cdots \,\big|M_{k-1}\Big] \; \Big[-G^1 \,\big|\cdots\big|-G^k\Big]^T \right\|_\infty \\ &\leq k\left\|Y^k\right\|_\infty \leq k(1+\delta)^2(2+\delta)^{k-2} \end{aligned}$$

■

In the following we will assume for simplicity $k_t = k$ for all nodes t. We use the results of the above lemmata to get a bound of the defect tensor F_t in the mode cluster t.

Lemma 12 *Let the matricizations $A^{(t)}, A^{(t_1)}$ and $A^{(t_2)}$ with $S(t) = \{t_1, t_2\}$ fulfill Assumption 9. We use the notation from Lemma 11. Then the error bound of the approximation of $A^{(t)}$ is given by*

$$\|F_t\|_\infty \leq \varepsilon\,(k\,\delta_k + 1) =: \varepsilon\,\overline{c}(k)$$

Proof: According to Assumption 9 and Lemma 11 we obtain

$$\begin{aligned}\|F_t\|_\infty &= \max_{i_1,i_2,q} \left|(F_t)_{(i_1,i_2),q}\right| \leq \max_{i_1,i_2,q} \sum_{\alpha=1}^{k} \left|(A_{t_1} S_{t_1}^{-1})_{i_1,\alpha}\right| \left|(F_{t_2})_{i_2,(\alpha,q)}\right| + \varepsilon \\ &\leq k\varepsilon \max_{i_1,\alpha} \left|(A_{t_1} S_{t_1}^{-1})_{i_1,\alpha}\right| + \varepsilon \leq \varepsilon(k\delta_k + 1).\end{aligned}$$

■

Let t_1, t_2 be mode clusters on level $p-1$ with sons

$$S(t_1) = \{s_1, s_2\}, \qquad S(t_2) = \{s_3, s_4\}.$$

Then the above error bound of the approximation is described by the constant $\overline{c}(k)$. In order to define the error bound for all other levels $l < p$, we denote this constant by

$$c_{p-1}(k) := \overline{c}(k).$$

Lemma 13 *Let $t \in T_\mathcal{I} \setminus \mathcal{L}(T_\mathcal{I})$ be a mode cluster on level $l-1$ with the sons t_1 and $t_2 \in T_\mathcal{I} \setminus \mathcal{L}(T_\mathcal{I})$ on level l. By using the notation*

$$\overline{A}^{(t)}_{(i_1,i_2),q} := \sum_{j=1}^{k} \sum_{\ell=1}^{k} (\tilde{N}_{t_1})_{i_1,q_j^1} \cdot (\tilde{N}_{t_2})_{i_2,q_\ell^2} \cdot (B_t)_{q,j,\ell}$$

with $\tilde{N}_{t_1}, \tilde{N}_{t_2}$ approximations of N_{t_1}, N_{t_2} with $\|\cdot\|_\infty$ error $\varepsilon c_l(k)$, the error of the approximation in level $l-1$ is bounded by

$$\left\|(A^{(t)} - \overline{A}^{(t)})|_{\mathcal{I}_t \times P_{t'}}\right\|_\infty \leq \varepsilon c_{l-1}(k), \qquad c_{l-1}(k) := \overline{c}(k) + k^3\, \delta_k\, c_l(k)^2.$$

Proof: For the short notation

$$\widetilde{A}^{(t)}_{(i_1,i_2),q} := \sum_{j=1}^{k} \sum_{\ell=1}^{k} A^{(t_1)}_{i_1,q_j^1} \cdot A^{(t_2)}_{i_2,q_\ell^2} \cdot (B_t)_{q,j,\ell}$$

we obtain

$$\left|(\widetilde{A}^{(t)} - \overline{A}^{(t)})_{(i_1,i_2)}\right| \leq \left|\sum_{j=1}^{k} \sum_{\ell=1}^{k} (B_t)_{q,j,\ell}\, (A^{(t_1)} - \tilde{N}_{t_1})_{i_1,q_j^1} \cdot (A^{(t_2)} - \tilde{N}_{t_2})_{i_2,q_\ell^2}\right|.$$

From the definition (8) of the transfer tensor, and due to the pivoting (with stopping tolerance ε) we obtain the bound

$$|(B_t)_{q,j,\ell}| \leq k\, \delta_k/\varepsilon.$$

Both together yield

$$\begin{aligned}\left\|(A^{(t)} - \overline{A}^{(t)})|_{\mathcal{I}_t \times P_{t'}}\right\|_\infty &\leq \left\|(A^{(t)} - \widetilde{A}^{(t)})|_{\mathcal{I}_t \times P_{t'}}\right\|_\infty + \left\|(\widetilde{A}^{(t)} - \overline{A}^{(t)})|_{\mathcal{I}_t \times P_{t'}}\right\|_\infty \\ &\leq \varepsilon\, (\overline{c}(k) + k^3\, \delta_k\, c_l(k)^2).\end{aligned}$$

■

The error bound of the tensor is the bound of the approximation in the root D with sons $S(D) := \{t, t'\}$. The entry (m, n) of the cross approximation in the root D is of the form

$$\sum_{j=1}^{k}\sum_{\ell=1}^{k} A^{(t)}_{m,j}\, A^{(t')}_{n,\ell}\, (B_D)_{1,j,\ell} = \sum_{j=1}^{k}\sum_{\ell=1}^{k} A^{(t)}_{m,j}\, (B_D)_{1,j,\ell}\, A^{(t)}_{\ell,n}$$
$$= \left(A^{(t)}\,|_{\mathcal{I}_t\times P_{t'}}\ (B_D)_{1,\cdot,\cdot}\, A^{(t)}\,|_{P_t\times\mathcal{I}_{t'}}\right)_{m,n}$$

Lemma 14 *We use the notation of Lemma 13 and additionally*

$$B_D := \left(A^{(t)}\,|_{P_t\times P_{t'}}\right)^{-1}.$$

The error of the tensor approximation is given by

$$\left\|A - \overline{A}\right\|_\infty \leq \varepsilon\,(1 + 2\,c_1(k)\,\delta_k + c_1(k)^2). \tag{10}$$

Proof: By the triangle inequality we get

$$\begin{aligned}
\left\|A - \overline{A}\right\|_\infty &= \left\|A^{(t)} - \overline{A}^{(t)}\,|_{\mathcal{I}_t\times P_{t'}}\ \left(A^{(t)}\,|_{P_t\times P_{t'}}\right)^{-1}\,\overline{A}^{(t)}\,|_{P_t\times\mathcal{I}_{t'}}\right\|_\infty\\
&\leq \left\|A^{(t)} - N_t\right\|_\infty + \left\|N_t - \overline{A}^{(t)}\,|_{\mathcal{I}_t\times P_{t'}}\ \left(A^{(t)}\,|_{P_t\times P_{t'}}\right)^{-1}\,A^{(t)}\,|_{P_t\times\mathcal{I}_{t'}}\right\|_\infty\\
&\quad + \left\|\overline{A}^{(t)}\,|_{\mathcal{I}_t\times P_{t'}}\ \left(A^{(t)}\,|_{P_t\times P_{t'}}\right)^{-1}\left[A^{(t)}\,|_{P_t\times\mathcal{I}_{t'}} - \overline{A}^{(t)}\,|_{P_t\times\mathcal{I}_{t'}}\right]\right\|_\infty\\
&\leq \varepsilon + \varepsilon\,c_1(k)\,\delta_k + \varepsilon\,c_1(k)\,\left\|\overline{A}^{(t)}\,|_{\mathcal{I}_t\times P_{t'}}\ \left(A^{(t)}\,|_{P_t\times P_{t'}}\right)^{-1}\right\|_\infty\\
&\leq \varepsilon + \varepsilon\,c_1(k)\,\delta_k + \varepsilon\,c_1(k)\,\delta_k + \varepsilon\,c_1(k)^2 = \varepsilon\,(1 + 2\,c_1(k)\,\delta_k + c_1(k)^2).
\end{aligned}$$

■

Remark 15 *The bound from equation (10) can be bounded by*

$$\varepsilon\,(1 + 2\,c_1(k)\,\delta_k + c_1(k)^2) \leq \varepsilon\, k^{2d-3}(\bar{c}(k))^{2d-2}.$$

This bound is only a worst case estimate. In practice the true error seems to be at most a small constant times ε, which is underlined by the numerical experiments of the following section.

5 Numerical Experiments

As a first example, we consider a tensor $A \in \mathbb{R}^{\mathcal{I}}$ given by the entries

$$A_{(i_1,\ldots,i_d)} := \left(\sum_{\mu=1}^{d} i_\mu^2\right)^{-1/2}. \tag{11}$$

It is well known that A can be approximated by a tensor $A_E \in \mathbb{R}^{\mathcal{I}}$ given as a sum of exponentials,

$$A_E := \sum_{j=1}^{k} \omega_j \bigotimes_{\mu=1}^{d} a_{j,\mu}, \qquad (a_{j,\mu})_{i_\mu} = \exp(-i_\mu^2 \alpha_j / d),$$

such that

$$\|A - A_E\|_\infty \leq 7.315 \times 10^{-10}.$$

The weights ω_j and exponents α_j were obtained from W. Hackbusch and are available on the webpage ($k = 35$, $R = 1000000$)

`http://www.mis.mpg.de/scicomp/EXP_SUM.`

We now consider A as a tensor given in a black box fashion and seek a tensor $A_B \in \mathcal{H}$-Tucker such that $A_B \approx A$. In particular, we are interested in how the error norms $\|A - A_B\|_2$ and $\|A - A_B\|_\infty$ will behave if we prescribe an accuracy of $\varepsilon > 0$ in the black box algorithm. Since we cannot measure both norms directly, we will look at the error norm $\|A_E - A_B\|_2$ instead of $\|A - A_B\|_2$ which can easily be computed within the $\mathcal{H}$-Tucker framework. For the $\|\cdot\|_\infty$ norm, we will randomly choose a set of indices $\mathcal{J} \subset \mathcal{I}$ with $\#\mathcal{J} = 10^5$ and look at the error

$$\|A - A_B\|_{\mathcal{J},\infty}, \qquad \|X\|_{\mathcal{J},\infty} := \max_{(i_1,\ldots,i_d)\in\mathcal{J}} \left| X_{(i_1,\ldots,i_d)} \right|.$$

In a first experiment, we fix $n := n_1 = \ldots = n_d := 32$ and look at the errors for dimensions $d = 8, 16, 32, 64$ (cf. Figure 4). The effective rank k_{eff} which corresponds to a storage complexity of $\mathcal{O}((d-1)k_{\text{eff}}^3 + k_{\text{eff}} \sum_{\mu=1}^{d} n_\mu)$ and the timings needed for the black box approximation are reported in Table 1. In a second experiment, we fix $d := 16$ and look at the errors, effective ranks, and timings for different mode sizes $n = 32, 64, 128, 256$. The results are summarised in Figure 5 and Table 2. Note that the accuracy for the Euclidean norm cannot be smaller than the accuracy of the exponential sum.

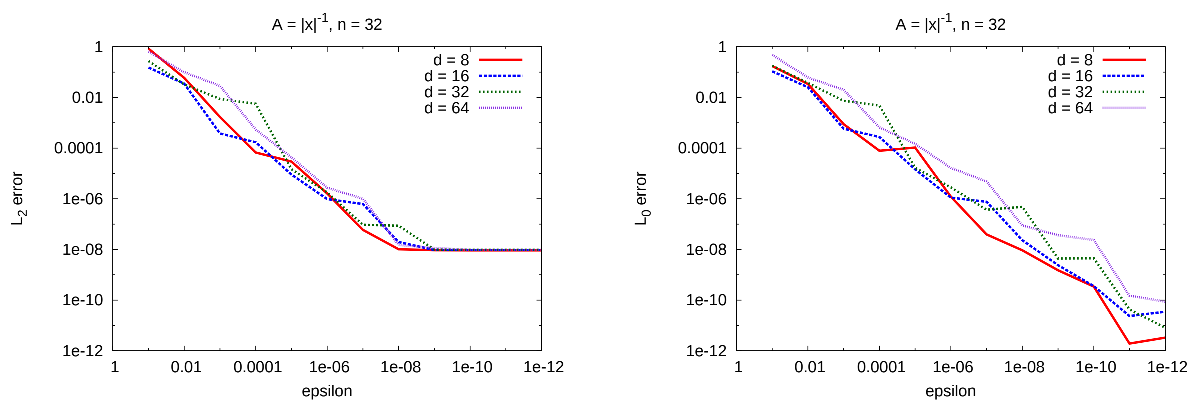

Figure 4: A as in (11). Left: $\|A_E - A_B\|_2 / \|A_E\|_2$. Right: $\|A - A_B\|_{\mathcal{J},\infty} / \|A\|_\infty$ for fixed $n = 32$

As a second example, we consider a tensor $A \in \mathbb{R}^{\mathcal{I}}$ given by the entries

$$A_{(i_1,\ldots,i_d)} := \exp\left\{ - \left(\sum_{\mu=1}^{d} (i_\mu / n_\mu)^2 \right)^{1/2} \right\}. \tag{12}$$

ε	d 8	16	32	64
1e-01	3.0	3.0	3.0	3.0
1e-02	4.6	4.8	4.5	4.9
1e-03	6.1	7.0	6.7	8.5
1e-04	8.1	8.6	9.8	10.7
1e-05	10.0	11.1	11.5	12.8
1e-06	11.0	13.1	13.4	15.5
1e-07	13.3	14.9	15.9	17.6
1e-08	15.2	16.2	18.4	20.7
1e-09	17.0	18.8	20.7	23.5
1e-10	18.2	20.4	23.2	25.0
1e-11	19.6	22.9	26.3	27.8
1e-12	21.1	23.8	27.3	30.3

ε	d 8	16	32	64
1e-01	0.03s	0.04s	0.08s	0.15s
1e-02	0.05s	0.11s	0.17s	0.38s
1e-03	0.10s	0.19s	0.33s	1.03s
1e-04	0.17s	0.25s	0.58s	1.38s
1e-05	0.24s	0.43s	0.80s	1.89s
1e-06	0.29s	0.63s	1.08s	3.11s
1e-07	0.42s	0.76s	1.70s	4.11s
1e-08	0.56s	0.90s	2.31s	5.77s
1e-09	0.69s	1.23s	3.00s	7.19s
1e-10	0.63s	1.46s	3.61s	8.31s
1e-11	0.77s	1.81s	4.87s	10.46s
1e-12	0.84s	1.96s	5.17s	12.59s

Table 1: A as in (11). Left: effective rank k_{eff}. Right: time for the black box algorithm for fixed $n = 32$

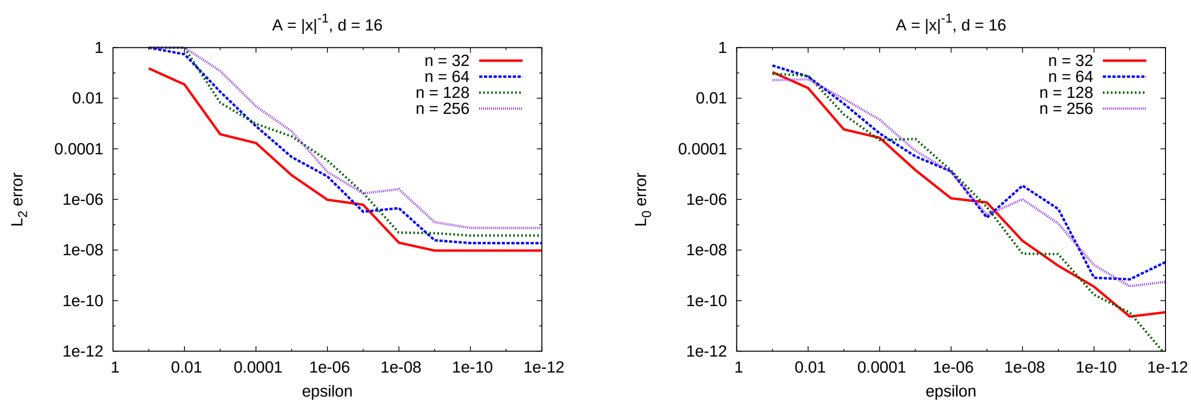

Figure 5: A as in (11). Left: $\|A_E - A_B\|_2 \,/\, \|A_E\|_2$. Right: $\|A - A_B\|_{\mathcal{J},\infty} \,/\, \|A\|_\infty$ for fixed $d = 16$

Again, one can find a tensor $A_E \in \mathbb{R}^{\mathcal{I}}$ given as a sum of exponentials ($k = 40$, $R = 300$) such that

$$\|A - A_E\|_\infty \leq 8.507 \times 10^{-9}.$$

The results for the approximation of A by the black box algorithm are summarised in Figures 6 and 7.

In a third example, we tested the black box algorithm for an $\mathcal{H}$-Tucker tensor A for which the entries of the transfer tensors and of the leaf frames were chosen randomly. We then artificially forced the node-wise singular values to roughly decay like α^j, $\alpha \in (0, 1)$, $j = 1, \ldots, k$. In a first experiment, we fix $d := 8$ and $n := 32$ and set all ranks to $k = 20$. We now look at the Euclidean error for $\alpha = 0.25$ corresponding to a fast decay of the singular values (cf. Table 3). The accuracy of the smallest singular value which is in the range of $\approx 10^{-12}$ can be met by the black box algorithm. In a second experiment, we set all ranks to $k = 20$ and look at the Euclidean error for $\alpha = 0.75$ corresponding to a slow decay of the singular values. Again, the accuracy of the smallest singular value which is in the range of

ε	n = 32	64	128	256
1e-01	3.0	3.0	3.0	3.0
1e-02	4.6	4.6	4.7	4.6
1e-03	6.8	6.1	5.6	5.7
1e-04	8.2	8.5	7.4	7.2
1e-05	9.4	10.1	9.4	8.7
1e-06	11.3	11.0	10.5	10.0
1e-07	13.2	12.9	12.1	11.4
1e-08	14.3	15.1	14.0	13.0
1e-09	16.2	16.3	15.5	14.5
1e-10	17.8	17.8	16.8	16.3
1e-11	19.4	19.6	19.3	17.5
1e-12	20.7	20.4	20.5	19.2

ε	n = 32	64	128	256
1e-01	0.01s	0.02s	0.07s	0.22s
1e-02	0.01s	0.04s	0.14s	0.43s
1e-03	0.03s	0.08s	0.20s	0.69s
1e-04	0.04s	0.13s	0.34s	0.97s
1e-05	0.05s	0.18s	0.55s	1.44s
1e-06	0.07s	0.21s	0.63s	2.06s
1e-07	0.10s	0.29s	0.85s	2.42s
1e-08	0.12s	0.41s	1.11s	3.17s
1e-09	0.15s	0.47s	1.38s	3.87s
1e-10	0.19s	0.59s	1.59s	5.27s
1e-11	0.24s	0.77s	2.24s	5.83s
1e-12	0.28s	0.80s	2.63s	7.56s

Table 2: A as in (11). Left: effective rank k_{eff}. Right: time for the black box algorithm for fixed $d = 16$

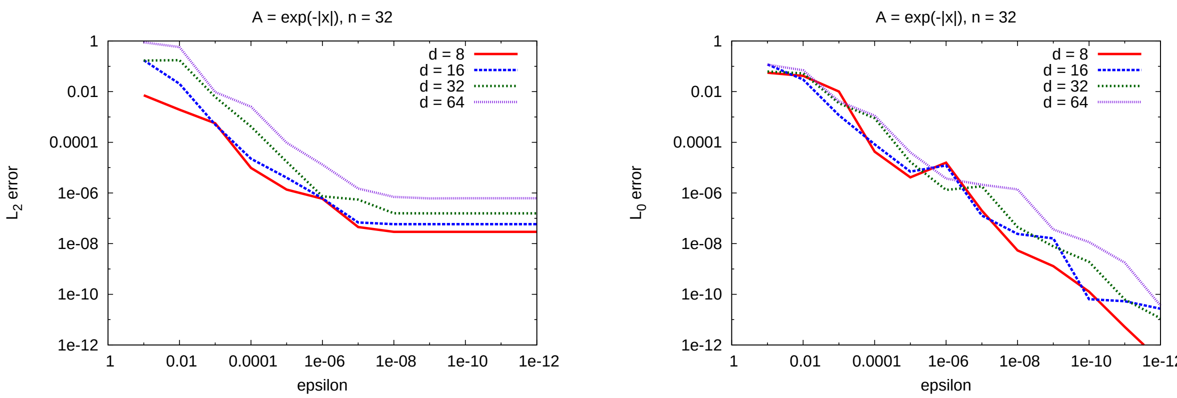

Figure 6: A as in (12). Left: $\|A_E - A_B\|_2 \,/\, \|A_E\|_2$. Right: $\|A - A_B\|_{\mathcal{J},\infty} \,/\, \|A\|_\infty$ for fixed $n = 32$

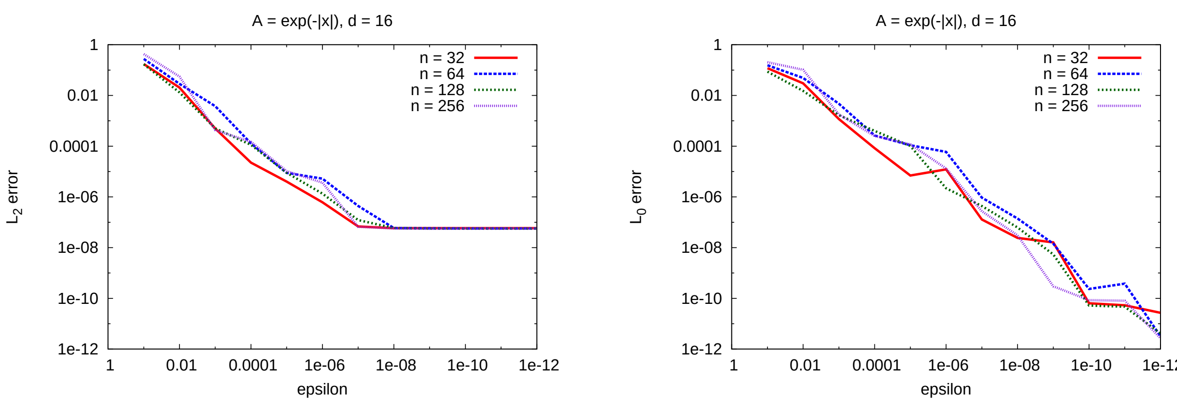

Figure 7: A as in (12). Left:$\|A_E - A_B\|_2 \,/\, \|A_E\|_2$. Right: $\|A - A_B\|_{\mathcal{J},\infty} \,/\, \|A\|_\infty$ for fixed $d = 16$

$\approx 10^{-3}$ can be met by the black box algorithm (cf. Table 3).

ε	$\|A - A_B\|_2 \,/\, \|A\|_2$	k_{eff}
1e-01	4.71e-04	2.0
1e-02	3.98e-04	2.3
1e-03	3.83e-04	2.9
1e-04	2.02e-05	4.1
1e-05	3.75e-06	5.2
1e-06	2.64e-07	6.5
1e-07	5.23e-08	8.0
1e-08	2.17e-09	9.3
1e-09	1.51e-10	10.8
1e-10	2.22e-11	12.1
1e-11	1.74e-12	13.7
1e-12	2.64e-13	14.5

ε	$\|A - A_B\|_2 \,/\, \|A\|_2$	k_{eff}
5.00e-01	6.29e-02	2.0
2.50e-01	1.10e-01	2.4
1.25e-01	6.07e-02	3.6
6.25e-02	3.98e-02	4.9
3.12e-02	1.91e-02	8.6
1.56e-02	1.58e-02	9.1
7.81e-03	4.97e-03	12.8
3.91e-03	3.75e-03	13.7
1.95e-03	1.28e-03	16.9
9.77e-04	4.33e-04	18.3
4.88e-04	3.37e-05	19.4
2.44e-04	5.32e-06	19.7

Table 3: A chosen randomly. Error and effective ranks for fixed $d := 8$, $n := 32$, $k = 20$. Left: $\alpha = 0.25$. Right: $\alpha = 0.75$

Conclusions

For tensors that can be represented exactly in the $\mathcal{H}$-Tucker format with representation ranks $(k_t)_{t \in T_{\mathcal{I}}}$, we can reconstruct these by inspection of only $\mathcal{O}(dk^3 + d\log(d)nk^2)$ entries ($k := \max_{t \in T_{\mathcal{I}}} k_t$, $n := \max_{\mu \in D} n_\mu$) in complexity $\mathcal{O}(dk^4 + d\log(d)nk^2)$. A similar result is obtained by Oseledets and Tyrtyshnikov [10] for the TT format. The difference is not only that our construction applies for the $\mathcal{H}$-Tucker format, but also that the pivot elements as well as their number are chosen adaptively and incremental to achieve a prescribed accuracy ε in the $\|\cdot\|_\infty$-norm. One can therefore estimate the error of the remainder during the construction, one can determine the necessary ranks and one can update an already computed approximation if a higher accuracy is required.

Under rather strong assumptions we have derived an error bound for the approximation in the case that the tensor has a higher (possibly full) representation rank. In the numerical examples we observe that the error in the $\|\cdot\|_\infty$-norm is typically close to the prescribed stopping tolerance ε, i.e., in practice there is almost no error amplification. However, the construction is heuristic and thus there can always be exceptions. It is therefore notable that even for random tensors the numerical results show a stable and almost optimal approximation.

References

[1] M. Bebendorf. Approximation of boundary element matrices. *Numer. Math.*, 86(4):565–589, 2000.

[2] S. Börm and L. Grasedyck. Hybrid cross approximation of integral operators. *Numerische Mathematik*, 101:221–249, 2005.

[3] M. Espig, L. Grasedyck, and W. Hackbusch. Black box low tensor rank approximation using fibre-crosses. *Constr. Approx.*, 30:557–597, 2009.

[4] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin. A theory of pseudoskeleton approximations. *Lin. Alg. Appl.*, 261:1–22, 1997.

[5] Lars Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.*, 31:2029–2054, 2010.

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

[7] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

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

[9] I.V. Oseledets and E.E. Tyrtyshnikov. Breaking the curse of dimensionality, or how to use svd in many dimensions. *SIAM J. Sci. Comp.*, 31(5):3744–3759, 2009.

[10] I.V. Oseledets and E.E. Tyrtyshnikov. TT-cross approximation for multidimensional arrays. *Linear Algebra and its Applications*, 432:70–88, 2010.