THE CONDITION NUMBER OF MANY TENSOR DECOMPOSITIONS IS INVARIANT UNDER TUCKER COMPRESSION

NICK DEWAELE*, PAUL BREIDING†, AND NICK VANNIEUWENHOVEN‡

Abstract. We characterise the sensitivity of several additive tensor decompositions with respect to perturbations of the original tensor. These decompositions include canonical polyadic decompositions, block term decompositions, and sums of tree tensor networks. Our main result shows that the condition number of all these decompositions is invariant under Tucker compression. This result can dramatically speed up the computation of the condition number in practical applications. We give the example of an $265 \times 371 \times 7$ tensor of rank 3 from a food science application whose condition number was computed in 6.9 milliseconds by exploiting our new theorem, representing a speedup of four orders of magnitude over the previous state of the art.

Key words. structured block term decomposition, sum of tree tensor networks, condition number, Tucker compression, invariance

AMS subject classifications. 49Q12, 53B20, 15A69, 65F35

1. Introduction. A tensor of order $D$ is a $D$-array of size $n_1 \times \cdots \times n_D$. In numerous applications (see, e.g., [20, 31] and the references therein), one seeks a decomposition that expresses a tensor $A$ as a sum of $R$ elementary terms:

$\mathbf{A} = \mathbf{A}_1 + \cdots + \mathbf{A}_R$,  

where $\mathbf{A}_r \in \mathcal{M}_r$ and $\mathcal{M}_r$ is a low-dimensional manifold in the space of tensors. Such a decomposition was called a join decomposition in [5].

In this paper, we study the sensitivity properties of a certain subclass of join decompositions related to tensors. We call them structured block term decompositions (SBTD). The formal definition of this class is given in Section 2 below. Informally, an SBTD involves manifolds $\mathcal{M}_r$ that are defined by imposing certain (manifold) structures on the core tensor of a Tucker decomposition with fixed multilinear rank $(l_1, \ldots, l_D)$. Many commonly used decompositions are SBTDs; for instance,

- sums of rank-1 tensors, i.e., canonical polyadic decomposition (CPD) [18],
- sums of Tucker decompositions, i.e., block term decomposition (BTD) [9],
- sums of tensor train decompositions [13, 28], and
- sums of hierarchical Tucker decompositions [15, 17].

The condition number is one way to measure the sensitivity of a tensor decomposition relative to perturbations of the tensor. For general join decompositions this number was analysed in [5]. One main result we establish in this paper is that the condition number of an SBTD is invariant under Tucker compression.

Recall that a Tucker decomposition [33] represents $\mathbf{A}$ in a tensor product subspace by expressing

$\mathbf{A} = (Q_1, \ldots, Q_D) : \mathcal{G} := \sum_{i_1=1}^{m_1} \cdots \sum_{i_D=1}^{m_D} g_{i_1, \ldots, i_D} q_{i_1}^1 \otimes \cdots \otimes q_{i_D}^D$,  

where $Q_d = [q_{i_d}^d] \in \mathbb{R}^{n_d \times m_d}$ with $n_d \geq m_d$ have linearly independent columns for each $d = 1, \ldots, D$.

---

*KU Leuven, Department of Computer Science, Celestijnenlaan 200A, B-3001 Leuven, Belgium. nick.dewaele@kuleuven.be

†MPI MiS, Inselstr. 22, 04103 Leipzig, Germany. paul.breiding@mis.mpg.de. P. B. is funded by the Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 445466444.

‡KU Leuven, Department of Computer Science, Celestijnenlaan 200A, B-3001 Leuven, Belgium. nick.vannieuwenhoven@kuleuven.be. N. V. was supported by a Postdoctoral Fellowship of the Research Foundation—Flanders (FWO) with project 12E8119N.
The core tensor $\mathcal{G}$ is often much smaller than $\mathcal{A}$, and it gives the coordinates of $\mathcal{A}$ with respect to the tensor product basis $Q_1 \otimes \cdots \otimes Q_D$. Note that we will switch freely between two equivalent notations for Tucker decomposition: the first, $(Q_1,\ldots,Q_D) \cdot \mathcal{G}$, is a common notation [12] for multilinear multiplication, while the second, $(Q_1 \otimes \cdots \otimes Q_D) \mathcal{G}$, emphasises that a Tucker decomposition consists of taking a particular linear combination of the tensors in a tensor product basis $Q_1 \otimes \cdots \otimes Q_D$. Herein, $Q_1 \otimes \cdots \otimes Q_D$ denotes the tensor product of matrices, which acts linearly on rank-1 tensors by $(Q_1,\ldots,Q_D) \cdot (v_1 \otimes \cdots \otimes v_D) := (Q_1 v_1 \otimes \cdots \otimes Q_D v_D)$. In coordinates, this matrix is given by $Q_1 \otimes \cdots \otimes Q_D = [q_{1,i} \otimes \cdots \otimes q_{D,i}]_{i=1,\ldots,d}$; see [16].

Originally proposed for CPD, Tucker compression [7] consists of expressing a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ in coordinates in the smallest tensor subspace in which it lives. It does not measure how the points are parametrised, which means fixing the space in which the perturbations $\delta$ are allowed to live. There are four increasingly restrictive ways of looking at the problem:

1. $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ is arbitrary and the SBTD of $\mathcal{A}$ is interpreted as the least-square minimiser $\text{argmin}_{(A_1,\ldots,A_R) \in M_1 \times \cdots \times M_R} \frac{1}{2} \left\| \mathcal{A} - (A_1 + \cdots + A_R) \right\|^2$.
2. $\mathcal{A}$ has an SBTD.
3. $\mathcal{A}$ can be Tucker compressed to a core $\mathcal{G} \in \mathbb{R}^{m_1 \times \cdots \times m_D}$ and $\mathcal{G}$ has an SBTD.
4. $\mathcal{A}$ lives in the same tensor subspace as $\mathcal{A}$, i.e., we have $\mathcal{A} = (Q_1,\ldots,Q_D) \cdot \mathcal{G}$ and $\mathcal{G}$ lives in the same tensor subspace as $\mathcal{A}$, i.e., we have $\mathcal{A} = (Q_1,\ldots,Q_D) \cdot \mathcal{G}$ and $\mathcal{A}$ and $\mathcal{G}$ both have an SBTD.

As explained in [5], under certain mild conditions, it is essential to quantify how sensitive the elementary terms are to the perturbations.

The obtained decomposition can then be extended to a decomposition of the original tensor $\mathcal{A}$ by multilinear multiplication with the basis $(Q_1,\ldots,Q_D)$. Since there are efficient algorithms [10,36] for computing an approximate Tucker decomposition of $\mathcal{A}$, contrary to the mostly optimization-based algorithms for computing CPDs, this compress–decompose–decompress approach can often reduce the overall computation time [7]. Another main result of this paper characterises which manifolds $\mathcal{M}_r$ in the join decomposition (1.1) are compatible with Tucker compression.

The topic of this article is to characterise how a decomposition of the form (1.1) changes if $\mathcal{A}$ is corrupted by noise. In order to obtain a robust interpretation of the elementary terms, it is essential to quantify how sensitive they are to the perturbations. As explained in [5], under certain mild conditions, $\mathcal{A}$ has an isolated decomposition $\mathcal{A} = (A_1,\ldots,A_R)$ and we can find a local inverse function $\Sigma^{-1}_\alpha$ of the addition map $\Sigma : M_1 \times \cdots \times M_R \to \mathbb{R}^{n_1 \times \cdots \times n_D}$, $(A_1,\ldots,A_R) \mapsto A_1 + \cdots + A_R$. The sensitivity of the elementary terms $A_i$ can be measured by the condition number [30]

$$\kappa_{\text{SBTD}}(A_1,\ldots,A_R) := \lim_{\delta \to 0} \sup_{\delta \in \mathcal{D}_\varepsilon|A-\tilde{A}| \leq \delta} \frac{\left| \Sigma^{-1}_\alpha(A) - \Sigma^{-1}_\alpha(\tilde{A}) \right|}{\| A - \tilde{A} \|},$$

where $\mathcal{D}$ is the set of valid perturbations (more on this below), and $\| \cdot \|$ denotes both the Euclidean norm on the ambient space $\mathbb{R}^{n_1 \times \cdots \times n_D}$ and the product Euclidean norm on $\mathbb{R}^{n_1 \times \cdots \times n_D} \times \cdots \times \mathbb{R}^{n_1 \times \cdots \times n_D}$. The condition number measures $\alpha = (A_1,\ldots,A_R)$ in one piece as tensors. It does not measure how the points are parametrised, which would introduce a number of complications.\footnote{See [35] for how to deal with such complications in the context of the CPD.} Furthermore, a priori, the condition number depends on both input $\mathcal{A}$ and output $(A_1,\ldots,A_R)$ because it is defined in terms of a local inverse [8]. However, since $\mathcal{A}$ depends uniquely on $(A_1,\ldots,A_R)$ we can write the condition number as a function of the output only. We have

$$\| \Sigma^{-1}_\alpha(A) - \Sigma^{-1}_\alpha(\tilde{A}) \| \leq \kappa_{\text{SBTD}}(A_1,\ldots,A_R) \| A - \tilde{A} \| + o \left( \| A - \tilde{A} \| \right)$$

as an asymptotically sharp first-order error bound. Equation (1.2) requires specifying the domain $\mathcal{D}$, which means fixing the space in which the perturbations $\delta$ are allowed to live. There are four increasingly restrictive ways of looking at the problem:

1. $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ is arbitrary and the SBTD of $\mathcal{A}$ is interpreted as the least-square minimiser $\text{argmin}_{(A_1,\ldots,A_R) \in M_1 \times \cdots \times M_R} \frac{1}{2} \left| \| \mathcal{A} - (A_1 + \cdots + A_R) \right|^2$.
2. $\mathcal{A}$ has an SBTD.
3. $\mathcal{A}$ can be Tucker compressed to a core $\mathcal{G} \in \mathbb{R}^{m_1 \times \cdots \times m_D}$ and $\mathcal{G}$ has an SBTD.
4. $\mathcal{A}$ lives in the same tensor subspace as $\mathcal{A}$, i.e., we have $\mathcal{A} = (Q_1,\ldots,Q_D) \cdot \mathcal{G}$ and $\mathcal{A}$ and $\mathcal{G}$ both have an SBTD.
A priori, one should expect the problem to become easier in the more restrictive cases in the sense that the condition number decreases. Indeed, the set of allowed perturbations gets strictly smaller. However, we prove the following surprising result.

**Theorem 1.1.** Let $\mathcal{A} = A_1 + \cdots + A_R$ be an SBTD. The condition number $\kappa_{\text{SBTD}}(A_1, \ldots, A_R)$ is the same for all four domains outlined above.

This theorem is implied by Theorem 5.1 and Corollary 5.2 below.

Theorem 1.1 is in stark contrast to some other problems in which the condition number depends on the domain. For instance, the condition number of the matrix logarithm for perturbations constrained in the symplectic group was studied in [2]. It was shown that the ratio between the unconstrained and constrained condition number can become arbitrarily large.

Our result indicates that computing the SBTDs of $\mathcal{A}$ and $\hat{g}$ are equally difficult from a numerical point of view. Indeed, it is known that the convergence rate of iterative methods to compute the decomposition is related to the condition number [1,4,25]. This suggests that compression, surprisingly, will not improve the local rate of convergence, even though the search space can be much smaller. Compression can, nevertheless, reduce the overall computation time when $\mathcal{A}$ is highly compressible [7].

A major practical advantage of Theorem 1.1 is that the condition number can be computed more efficiently by considering $\mathcal{A}$ as a point in a tensor product subspace: It suffices to compute the condition number of the core $\hat{g}$. An example illustrates the above significant computational advantage. Consider a rank-3 tensor of dimensions $265 \times 371 \times 7$, as in the sugar data set of [7]. Its CPD can be compressed to that of a $3 \times 3 \times 3$-tensor. We implemented two algorithms to compute the condition number in Julia v1.6. [3]: the one from [5] and one based on Theorem 1.1. On a system with an Intel Xeon CPU E5-2697 v3 running on 8 cores and 126GB memory, this took 110 seconds and 6.9 milliseconds, respectively, which gives a speedup of over 15 000. If the CPD is already in compressed form, the time can be reduced further to only 0.089 milliseconds, representing a speedup of more than a million over the state of the art.

1.1. Outline. We introduce the SBTD in Section 2. In Section 3, we derive the geometric foundations of structured Tucker decompositions upon which the theory of the condition number is based. An algorithm to compute the condition number is outlined in Section 4. We also present qualitative properties of well or ill-conditioned SBTDs. Section 5 introduces subspace-constrained SBTDs and proves the main result, Theorem 1.1, which states that the condition number of SBTDs is invariant under Tucker compression. Concluding numerical experiments are found in Section 7.

1.2. Notation. The only norm used in this paper are the Euclidean (or Frobenius) norms for tensors and vectors, all consistently denoted by $\| \cdot \|$. The manifold of real $n \times m$ matrices of rank $m$ is denoted as $\mathbb{R}^{n \times m}_m$, where $n \geq m$. The $n$-dimensional sphere is $\mathbb{S}^n$. The $j$th column of the identity $I_n$ is $e_j^{(n)}$. If the ambient dimension is clear from the context, we also abbreviate $e_j := e_j^{(n)}$. The $d$th unfolding of a tensor $\mathcal{A}$ is $\mathcal{A}_{(d)}$. For any matrix $X$ and any set of matrices $A_n$, $n = 1, \ldots, N$, and any $k = 1, \ldots, N+1$, we denote $X \otimes (A_1 \otimes \cdots \otimes A_N) := A_1 \otimes \cdots \otimes A_{k-1} \otimes X \otimes A_k \otimes \cdots \otimes A_N$.

For a group $G$ acting on a set $\mathcal{M}$, the projection of $x \in \mathcal{M}$ onto its $G$-orbit is $[x]_G$.

2. The structured block term decomposition. In this section, we give a formal definition of the SBTD, the main tensor decomposition that we study in this paper. Just as a BTD is a linear combination of Tucker decompositions, an SBTD is a linear combination of *structured Tucker decompositions*. The structure we consider is imposed on the core tensor of the Tucker decomposition.

**Definition 2.1** (Tucker core structure). A smooth submanifold $\mathcal{M} \subseteq \mathbb{R}^{l_1 \times \cdots \times l_D}$ is a *Tucker core structure* if it is a $(\text{GL}(l_1) \times \cdots \times \text{GL}(l_D))$-homogeneous manifold:

1. $\mathcal{C} \in \mathcal{M}$ has multilinear rank equal to $(l_1, \ldots, l_D)$, and
2. $(A_1, \ldots, A_D) \cdot \mathcal{C} \in \mathcal{M}$ for all $A_d \in \text{GL}(l_d)$ with $d = 1, \ldots, D$. 


Next, we can define the $\mathcal{M}$-structured Tucker decomposition.

**Definition 2.2 (Structured Tucker decomposition).** Let $\mathcal{M} \subseteq \mathbb{R}^{l_1 \times \cdots \times l_D}$ be a Tucker core structure. An $\mathcal{M}$-structured Tucker decomposition of $A \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ is an expression of the form

$$A = (U_1, \ldots, U_D) \cdot C = (U_1 \otimes \cdots \otimes U_D)C$$

with $C \in \mathcal{M}$ and all $U_d \in \mathbb{R}^{n_d \times l_d}$ for $d = 1, \ldots, D$.

The first basic result we establish in Section 3 below ensures that the results from [5] can be applied to study the condition number.

**Proposition 2.3.** The set of all tensors $A$ admitting an $\mathcal{M}$-structured Tucker decomposition forms a smooth embedded submanifold $\mathcal{M}^{n_1, \ldots, n_D} \subseteq \mathbb{R}^{n_1 \times \cdots \times n_D}$, called the $\mathcal{M}$-structured Tucker manifold.

An important subclass of structured Tucker manifolds in applications are defined by tensor networks in which the graph is a tree [27]. This includes tensors with a fixed rank Tucker decomposition [33], fixed-rank tensor train decomposition [28], and fixed rank hierarchical Tucker decomposition [15, 17].

The set of tree tensor networks (i.e., hierarchical Tucker formats) with fixed ranks is known to form a manifold [34]. This manifold is invariant under the natural action of $\mathrm{GL}(l_1) \times \cdots \times \mathrm{GL}(l_D)$. Since multilinear rank is also invariant under this action [21], all concise (i.e., multilinear rank equals the dimension of the ambient space) tree-based tensor networks are valid Tucker core structures. This includes all aforementioned Tucker, tensor trains, and hierarchical Tucker decompositions in $\mathbb{R}^{l_1 \times \cdots \times l_D}$ of multilinear rank $(l_1, \ldots, l_D)$. In particular, $\mathbb{R}^n$ is a valid Tucker core structure.

We will be interested in additive decompositions whose elementary terms lie in structured Tucker manifolds, called structured block term decompositions (SBTDs).

**Definition 2.4 (Structured block term decomposition).** An SBTD of the tensor $A \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ associated with the $\mathcal{M}_r$-structured Tucker manifolds $\mathcal{M}^{n_1, \ldots, n_D}$ is a decomposition of the form $A = A_1 + \cdots + A_R$ with $A_r \in \mathcal{M}^{n_1, \ldots, n_D}$ for $r = 1, \ldots, R$.

Any sum mixing rank-1 tensors, Tucker decompositions, tensor trains decompositions, and hierarchical Tucker decompositions is thus an SBTD.

### 3. The Geometry of the Structured Tucker Manifold

The condition number of join decompositions from [5] requires that the summands in (1.1) live on manifolds. Therefore, we first derive the geometric properties of the manifolds involved in the decomposition. We prove Proposition 2.3, which shows that the $\mathcal{M}^{n_1, \ldots, n_D}$ in Definition 2.2 are indeed manifolds. We also derive an expression for its tangent space. The proofs of these statements are standard computations in differential geometry, similar to those of [34].

The following result establishes the differential structure of our manifolds.

**Proposition 3.1.** Let $\mathcal{M}$ be a Tucker core structure as in Definition 2.1. Define the manifolds

$$\widetilde{\mathcal{M}} := \mathcal{M} \times \mathbb{R}^{n_1 \times l_1} \times \cdots \times \mathbb{R}^{n_D \times l_D} \quad \text{and} \quad \mathcal{G} := \mathrm{GL}(l_1) \times \cdots \times \mathrm{GL}(l_D)$$

and the group action

$$\theta : \mathcal{G} \times \widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}$$

$$(A_1, \ldots, A_D), (C, U_1, \ldots, U_D)) \mapsto ((A_1^{-1}, \ldots, A_D^{-1}) \cdot C, U_1 A_1, \ldots, U_D A_D).$$

Then $\widetilde{\mathcal{M}}/\mathcal{G}$ is a quotient manifold with a unique smooth structure so that the quotient map $\left[\cdot\right]_G : \widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}/\mathcal{G}$ is a smooth submersion.
Proof. By [22, Theorem 21.10], we only need to verify that the action is smooth, free (i.e., it fixes the identity), and proper. The first two properties are straightforward to check. To show that it is proper, consider the sequences \( \{x_n\}_{n \in \mathbb{N}} \rightarrow x \) in \( \tilde{M} \) and \( \{a_n\}_{n \in \mathbb{N}} \) in \( G \) where \( \{\theta(a_n, x_n)\}_{n \in \mathbb{N}} \) converges in \( \tilde{M} \). By [22, Proposition 21.5], \( \theta \) is proper if \( \{a_n\}_{n \in \mathbb{N}} \) converges in \( G \). Consider the map \( f : \tilde{M} \times M \rightarrow G \) taking
\[
(c, U_1, \ldots, U_D), (\hat{c}, \hat{U}_1, \ldots, \hat{U}_D) \mapsto (U^T_1 \hat{U}_1, \ldots, U^T_D \hat{U}_D),
\]
where \( U^T_1 = (U^T_2 U_2)^{-1} U^T_2 \) is the Moore-Penrose inverse. Note \( f(x_n, \theta(a_n, x_n)) = a_n \). Furthermore, \( f \) is continuous by the continuity of the Moore-Penrose inverse. Since \( \{x_n, \theta(a_n, x_n)\}_{n \in \mathbb{N}} \) converges, so does \( \{f(x_n, \theta(a_n, x_n))\}_{n \in \mathbb{N}} = \{a_n\}_{n \in \mathbb{N}} \).

The tangent space to this quotient manifold is derived next.

**Proposition 3.2.** Take the manifold \( \tilde{M}/G \) as in Proposition 3.1 and consider a point \( x = (c, U_1, \ldots, U_D) \in \tilde{M} \). Complete each basis \( [U_d, U^*_d] \) of \( \mathbb{R}^{n_d} \). Then
\[
T_{[x]}(\tilde{M}/G) \cong \left\{ (\hat{c}, U^*_1 B_1, \ldots, U^*_D B_D) \mid \hat{c} \in T_c M, B_d \in \mathbb{R}^{(n_d-l_d) \times l_d} \right\}.
\]

**Proof.** Define the fibre \( F \) of all \( x' \) equivalent to \( x \):
\[
F_x = \left\{ ((A_1, \ldots, A_D) \cdot c, U_1 A_1^{-1}, \ldots, U_D A_D^{-1}) \mid A_d \in \text{GL}(l_d), d = 1, \ldots, D \right\}.
\]
This allows us to define the vertical space as the tangent space to \( F \) at \( x \):
\[
V_x = \left\{ \sum_{d=1}^D \left( \hat{A}_d \otimes d_{d' \neq d} \otimes I_{l_{d'}} \right) c, -U_1 \hat{A}_1, \ldots, -U_D \hat{A}_D \right\} | \hat{A}_d \in \mathbb{R}^{l_d \times l_d} \right\}.
\]
In this expression, we used \( T_{A_d} \text{GL}(l_d) \cong \mathbb{R}^{l_d \times l_d} \) [22] for each \( d = 1, \ldots, D \). Now define the horizontal space as
\[
H_x := \left\{ (\hat{c}, U^*_1 B_1, \ldots, U^*_D B_D) \mid \hat{c} \in T_c M, B_d \in \mathbb{R}^{(n_d-l_d) \times l_d} \right\}.
\]
We will show that \( V_x \oplus H_x = T_x \tilde{M} \). First, we verify that the intersection is trivial. Take \( \xi \in V_x \), parametrised as in (3.1). If also \( \xi \in H_x \), by construction of \( U^*_d \), it must hold that all \( A_d \) in the parametrisation of \( \xi \) are zero and hence \( \xi = 0 \).

Next, we show that the sum is \( T_x \tilde{M} \). We know \( M \) is invariant under the application of \( \text{GL}(l_1) \times \cdots \times \text{GL}(l_D) \). Therefore, for any \( \hat{A}_d \in \mathbb{R}^{l_d \times l_d} \) for \( d = 1, \ldots, D \), there exist curves over \( M \) of the form \( \gamma(t) = (A_1(t), \ldots, A_D(t)) \cdot c \) with \( A_d(0) = I_{l_d} \) and \( \frac{d}{dt} \bigg|_{t=0} A_d(t) = \hat{A}_d \). Hence, all tensors of the form
\[
\gamma'(0) = \sum_{d=1}^D \hat{A}_d \otimes d_{d' \neq d} \otimes I_{l_{d'}} c \quad \text{with} \quad \hat{A}_d \in \mathbb{R}^{l_d \times l_d}, \quad d = 1, \ldots, D,
\]
are tangent to \( M \) at \( x \). Because of this, it is easy to check that
\[
V_x \oplus H_x \cong T_c \tilde{M} \times \mathbb{R}^{n_1 \times l_1} \times \cdots \times \mathbb{R}^{n_D \times l_D} = T_x \tilde{M}.
\]
By the general theory of quotient manifolds, this establishes \( H_x \cong T_{[x]}(\tilde{M}/G) \), where the isomorphism is the unique horizontal lift [1, Section 3.5.8].

We have established that Tucker decompositions with a structured core form a smooth manifold. By definition, a point on an \( \tilde{M} \)-structured Tucker manifold corresponds to a Tucker decomposition that is unique up to basis transform. We now have all the tools we need to show that \( \mathcal{M}^{n_1, \ldots, n_D} \) is a manifold. We do this next.
Proposition 3.3. Let \( \tilde{\mathcal{M}} / \mathcal{G} \) be as in Proposition 3.1 and let \( \mathcal{M}^{n_1 \ldots n_D} \) be the \( \mathcal{M} \)-structured Tucker manifold. Then \( \mathcal{M}^{n_1 \ldots n_D} \) is a smooth embedded submanifold of \( \mathbb{R}^{n_1 \times \ldots \times n_D} \) and the following is a diffeomorphism:

\[
\Phi : \tilde{\mathcal{M}} / \mathcal{G} \to \mathcal{M}^{n_1 \ldots n_D}
\]

\[
[(C, U_1, \ldots, U_D)]_{\mathcal{G}} \mapsto (U_1, \ldots, U_D) \cdot C.
\]

Moreover, the tangent space to \( \mathcal{M}^{n_1 \ldots n_D} \) at \( (U_1, \ldots, U_D) \cdot C \) is generated by all tensors

\[
(3.2)
\]

\[
(U_1, \ldots, U_D) \cdot \hat{C} + \sum_{d=1}^{D} \left( \hat{U}_d \otimes_d \bigotimes_{d' \neq d} U_{d'} \right) C
\]

with \( \hat{C} \in T_{C} \tilde{\mathcal{M}} \) and \( U_d \hat{U}_d = 0_{\alpha d} \cdot l_d \) for all \( d = 1, \ldots, D \).

Proof. By [22, Proposition 5.2], the first claim holds if \( \Phi \) is both a homeomorphism and a smooth immersion. First, we show that it is a homeomorphism. Note that \( \Phi \) is a bijection because \( \mathcal{M}^{n_1 \ldots n_D} \) is precisely the set of all tensors with a Tucker decomposition where the core is in \( \mathcal{M} \). Since \( \Phi \) is induced by a polynomial map, it is also continuous. To show that \( \Phi^{-1} \) is continuous, consider the maps

\[
V_d : \mathcal{M}^{n_1 \ldots n_D} \to \text{Gr}(n_d, l_d)
\]

\[
(U_1, \ldots, U_D) \cdot C \mapsto [U_d]_{\text{GL}(l_d)}
\]

where \( \text{Gr}(n_d, l_d) \cong \mathbb{R}^{n_d \times l_d} / \text{GL}(l_d) \) is the Grassmannian of \( n_d \)-dimensional linear spaces in \( \mathbb{R}^{n_d} \) [1]. That is, \( V_d(x) \) is the column span of its \( d \)-th flattening \( x_{(d)} \).

We will demonstrate continuity of \( V_d \) at any \( x \in \mathcal{M}^{n_1 \ldots n_D} \) by showing that any open neighbourhood \( V \) of \( V_d(x) \) contains the image of a neighbourhood of \( x \) [23, Theorem 18.1]. By the definition of the quotient topology, \( V = [U]_{\text{GL}(l_d)} \) for some open neighbourhood \( U \subseteq \mathbb{R}^{n_d \times l_d} \) of \( U_d \), where \( U_d \) is any representative of \( V_d(x) \). Furthermore, for some ball \( B_\varepsilon(U_d) \) of radius \( \varepsilon \) centered at \( U_d \), we have \( [B_\varepsilon(U_d)]_{\text{GL}(l_d)} \subseteq [U]_{\text{GL}(l_d)} = \mathcal{V} \).

Now we exploit the liberty of choosing the representative \( U_d \). Observe that \( V_d(x) \) is the span of \( l_d \) columns of \( x_{(d)} \). In other words, there exists a column selection operator \( P_d \in \mathbb{R}^{(l_d \times n_d) \times l_d} \), so that \( V_d(x) = [x_{(d)} P_d]_{\text{GL}(l_d)} \). By the upper semicontinuity of matrix rank, there exists \( 0 < \delta < \varepsilon \) so that any perturbation to \( x \) of norm less than \( \delta \) does not change the rank of \( x_{(d)} \) or \( x_{(d)} P_d \). Hence, \( V_d(x) = [x_{(d)} P_d]_{\text{GL}(l_d)} \) for any \( x \) in a ball \( B_\delta(x) \) of radius \( \delta \). Because \( [x_{(d)} P_d - x_{(d)}]_{\text{GL}(l_d)} \) lies in \( \mathcal{V} \), we have \( V_d(B_\delta(x)) \subseteq [B_\varepsilon(x_{(d)}) P_d]_{\text{GL}(l_d)} \), which proves continuity of \( V_d \).

For any \( x \in \mathcal{M}^{n_1 \ldots n_D} \), let \( V_d(x) = [U_d]_{\text{GL}(l_d)} \) for each \( d \). It can be verified that the following is independent of the representatives \( U_d \):

\[
\Psi(x) := \left[ (U^1_1, \ldots, U^D_D) \cdot C, U_1, \ldots, U_D \right]_{\mathcal{G}}.
\]

The right-hand side is the Tucker decomposition of \( x \), which is unique up to the action of \( \mathcal{G} \). Hence, \( \Psi = \Phi^{-1} \). This shows that \( \Phi^{-1} \) is the composition of continuous maps: \( V_d \) for each \( d \), the Moore-Penrose inverse, multilinear multiplication, and the canonical projection map. Hence, \( \Phi^{-1} \) is continuous.

Next, we show that \( \Phi \) is an immersion, i.e., that its derivative maps a basis to a basis, in which case \( T_x \mathcal{M}^{n_1 \ldots n_D} \) is the image of \( d\Phi \). Fix a basis \( \mathcal{B}_0 \) of \( T_x \mathcal{M} \) and, for each \( d = 1, \ldots, D \), a basis \( \mathcal{B}_d \) of all \( U_d \in \mathbb{R}^{n_d \times l_d} \) so that \( U_d \hat{U}_d = 0 \). By Proposition 3.2, the tangent space of \( \tilde{\mathcal{M}} / \mathcal{G} \) can be considered as a product space generated by the canonical product basis derived from \( \mathcal{B}_0, \ldots, \mathcal{B}_D \).

Applying \( d\Phi \) to this basis of \( \tilde{\mathcal{M}} / \mathcal{G} \) gives tangents of the form

\[
T_0 := \left\{ (U_1, \ldots, U_D) \cdot \hat{C} \right\} \quad \text{and} \quad T_d := \left\{ (U_1, \ldots, U_{d-1}, \hat{U}_d, U_{d+1}, \ldots, U_D) \cdot C \right\}
\]
in which $\mathcal{C} \in \mathbb{R}_0$ and $\hat{U}_d \in \mathbb{R}_d$, $d = 1, \ldots, D$. Note that the sets $T_0$ and $T_d$ with $i \neq j$ are pairwise orthogonal due to the constraint on $\hat{U}_d$. Since $(U_1 \otimes \cdots \otimes U_D)$ has full rank, $T_0$ is linearly independent. The tangents in the set $T_d$ with $d \geq 1$ are tensors whose $d$th unfolding is

$$
\hat{U}_d \mathcal{C}_d(u_1 \otimes \cdots \otimes u_{d-1} \otimes u_{d+1} \otimes \cdots \otimes u_D)^T.
$$

Recall from Definition 2.2 that $\mathcal{C}_d$ and all $U_i$ have full row rank. For a set of linearly independent matrices $\hat{U}_d$, all matrices (3.3) are linearly independent. This shows that $\Phi$ is an immersion. By [22, Proposition 5.2], $\mathcal{M}^{n_1 \times \cdots \times n_D}$ is a manifold and $\Phi$ is a diffeomorphism. By [22, Theorem 4.14] $T_\mathcal{X} \mathcal{M}^{n_1 \times \cdots \times n_D}$ is the image of $d\Phi$.

Note that Proposition 2.3 is a corollary of the previous statement.

4. Computing the condition number. Having shown that the structured Tucker decompositions form a manifold, we can investigate their condition number using the tools from [5]. For this, we first derive an orthonormal basis of the structured Tucker manifold, so that the condition number can be computed with efficient algorithms from linear algebra using (4.2) below. We present some examples, as well as useful estimates of the condition number of SBTDs.

4.1. A direct algorithm. Let $\mathcal{M}_1^{n_1 \times \cdots \times n_D}, \ldots, \mathcal{M}_R^{n_1 \times \cdots \times n_D}$ be structured Tucker manifolds, and recall the addition map

$$
\Sigma : \mathcal{M}_1^{n_1 \times \cdots \times n_D} \times \cdots \times \mathcal{M}_R^{n_1 \times \cdots \times n_D} \to \mathbb{R}^{n_1 \times \cdots \times n_D}, \, (\mathcal{A}_1, \ldots, \mathcal{A}_R) \mapsto \mathcal{A}_1 + \cdots + \mathcal{A}_R
$$

from the introduction. Computing an SBTD translates to finding a decomposition $(\mathcal{A}_1, \ldots, \mathcal{A}_R)$ so that $\Sigma(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \mathcal{A}$. The condition number $\kappa^{\text{SBTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R)$ from (1.2) is computed as follows [5]. For $r = 1, \ldots, R$, compute orthonormal bases of $T_{\mathcal{A}_r} \mathcal{M}_r^{n_1 \times \cdots \times n_D}$, the tangent space to $\mathcal{M}_r^{n_1 \times \cdots \times n_D}$ at $\mathcal{A}_r$. The basis vectors are the columns of matrices $T_r$. Then, the so-called Terracini matrix is constructed as

$$
T_{\mathcal{A}_1, \ldots, \mathcal{A}_R} := [T_1 \cdots T_R].
$$

The condition number satisfies

$$
\kappa^{\text{SBTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \frac{1}{\sigma_{\min}(T_{\mathcal{A}_1, \ldots, \mathcal{A}_R})},
$$

where $\sigma_{\min}(A) = \sigma_{\min(m,n)}(A)$ denotes the smallest singular value of $A \in \mathbb{R}^{m \times n}$. Thus, the computation of $\kappa^{\text{SBTD}}$ requires orthonormal bases of the tangent spaces to the structured Tucker manifolds. We explain this in the next proposition.

Recall that the compact higher-order singular value decomposition (HOSVD) [10, 33] is an orthogonal Tucker decomposition $X = (U_1, \ldots, U_D) \circ \mathcal{C}$ with $U_d \in \mathbb{R}^{n_d \times l_d}$ a basis of left singular vectors of $X_d$ corresponding to the nonzero singular values. In particular, $U_d^T U_d = I_{l_d}$ and the columns of $U_d$ span the column span of $X_d$. The core tensor $\mathcal{C}$ is the orthogonal projection of $X$ onto the orthonormal basis $U_1 \otimes \cdots \otimes U_D$: $\mathcal{C} = (U_1^T, \ldots, U_D^T) \circ X$. With this terminology in place, we can state the result.

Proposition 4.1. Let $\mathcal{M}^{n_1 \times \cdots \times n_D} \subseteq \mathbb{R}^{n_1 \times \cdots \times n_D}$ be the $\mathcal{M}$-structured Tucker manifold with Tucker core structure $\mathcal{M} \subseteq \mathbb{R}^{l_1 \times \cdots \times l_D}$. Assume that we are given a tensor $X \in \mathcal{M}^{n_1 \times \cdots \times n_D}$ expressed in HOSVD format $X = (U_1, \ldots, U_D) \circ \mathcal{C}$. Complete each $U_d$ to an orthonormal basis $[U_d \ U_d^\perp]$ of $\mathbb{R}^{n_d}$. Let $\sigma_d := \|e_d^T \mathcal{C}_d\|$ and $\mathbf{u}_d := (\sigma_d^d)^{-1} e_d^T \mathcal{C}_d$. If $\mathcal{B}_C$ is an orthonormal basis for $T_{\mathcal{C}} \mathcal{M}$, the following is an orthonormal basis of the tangent space $T_{\mathcal{X}} \mathcal{M}^{n_1 \times \cdots \times n_D}$:

$$
\mathcal{B}_\mathcal{X} := \{(U_1, \ldots, U_D) \circ \mathcal{C} \} \cup \{(U_1, \ldots, U_{d-1}, U_d^\perp e_i (\mathbf{u}_d^T)^T, U_{d+1}, \ldots, U_D) \circ \mathcal{C} \}
$$

in which $\mathbf{C} \in \mathcal{B}, d = 1, \ldots, D$, $i = 1, \ldots, n_d - l_d$ and $j = 1, \ldots, l_d$. 

7
Proof. Equation \((3.2)\) for \(T_x\mathcal{M}^{n_1\cdots n_D}\) suggests a decomposition of the tangent space of the form

\[
T_x\mathcal{M}^{n_1\cdots n_D} = T_0 \oplus T_1 \oplus \cdots \oplus T_D,
\]

where \(T_0\) contains all tangents of the form \((U_1, \ldots, U_D) \cdot \mathcal{C}\) and \(T_d\) with \(d = 1, \ldots, D\) contains the tangents of the form \((U_1, \ldots, U_{d-1}, U_d, U_{d+1}, \ldots, U_D) \cdot \mathcal{C}\). As argued in the proof of Proposition 3.3, this is a decomposition of \(T_x\mathcal{M}^{n_1\cdots n_D}\) into pairwise orthogonal spaces.

First we verify that \((4.3)\) spans \(T_x\mathcal{M}^{n_1\cdots n_D}\). Since we have an orthonormal basis of \(T_x\mathcal{C}\) available, we have

\[
T_0 = \text{span}\left\{(U_1, \ldots, U_D) \cdot \mathcal{C} \mid \mathcal{C} \in \mathcal{B}_{\mathcal{C}}\right\} = (U_1 \otimes \cdots \otimes U_D)(T_x\mathcal{C}).
\]

For the other \(D\) subspaces \(T_d\), we require all \(\hat{U}_d\) such that \(U_d^T \hat{U}_d = 0\), or equivalently \(\hat{U}_d = U_d^T B\) for some \(B \in \mathbb{R}^{(n_d-l_d) \times l_d}\). The \(e_i^{l_d}(\hat{u}_d^j)^T\) with \(i = 1, \ldots, n_d - l_d\) and \(j = 1, \ldots, l_d\) are a basis of \(\mathbb{R}^{(n_d-l_d) \times l_d}\), because they are just a rescaling of the canonical basis \(e_i^{l_d}(\hat{u}_d^j)^T\). Substituting each of these for \(B\), we get a basis of all available \(\hat{U}_d\). This parameterises all of \(T_d\) as

\[
T_d = \text{span}\left\{(U_1, \ldots, U_{d-1}, U_d^\perp e_i(\hat{u}_d^j)^T, U_{d+1}, \ldots, U_D) \cdot \mathcal{C} \mid i = j = 1 \right\}^{i=n_d-l_d,j=l_d}.
\]

Hence, the proposed basis \(\mathcal{B}_x\) generates \(T_0 \oplus T_1 \oplus \cdots \oplus T_D\).

We have yet to verify that the proposed basis is orthonormal. We already know that \(T_0, \ldots, T_D\) are pairwise orthogonal. It thus suffices to show that the bases we constructed for each of these spaces separately are orthonormal. The basis for \(T_0\) is orthonormal because \(\mathcal{B}_x\) is orthonormal and \(U_1 \otimes \cdots \otimes U_D\) is an orthonormal tensor product basis.

For the basis of \(T_d\) with \(d \geq 1\), we use the fact that \(x\) is in HOSVD format. This ensures that an HOSVD of \(\mathcal{C}\) is \((1, \ldots, 1) \cdot \mathcal{C}\). In other words, \(\mathcal{C}_{(d)}\) has singular values \(\sigma_j^d\) as defined above and its corresponding left singular vectors are \(e_j^d\) [10]. Hence, the transpose of its \(j\)th right singular vector is \((v_j^d)^T := (\sigma_j^d)^{-1} e_j^d \mathcal{C}_{(d)} = (u_j^d)^T \mathcal{C}_{(d)}\). With this in mind, we calculate the inner products between the basis vectors of \(T_d\):

\[
\langle (U_1, \ldots, U_{d-1}, U_d^\perp e_i(\hat{u}_d^j)^T, U_{d+1}, \ldots, U_D) \cdot \mathcal{C}, (U_1, \ldots, U_{d-1}, U_d^\perp e_j(\hat{u}_d^l)^T, U_{d+1}, \ldots, U_D) \cdot \mathcal{C} \rangle
\]

\[
= \text{Trace}(U_d^\perp e_i(\hat{u}_d^j)^T \mathcal{C}_{(d)} \mathcal{C}_{(d)^T} U_d^\perp e_j(\hat{u}_d^l)^T)
\]

\[
= \langle U_d^\perp e_i, U_d^\perp e_j \rangle \langle \hat{u}_d^j, \hat{u}_d^l \rangle.
\]

If \(i = j\), the right-hand side is the inner product between two right singular vectors of \(\mathcal{C}_{(d)}\), which is \(\delta_{jj}\), the Kronecker delta. Otherwise, it is zero due to the orthogonality of the columns of \(U_d^\perp\). This ensures that our basis of \(T_d\) is orthogonal, which completes the proof.

Example 1 (BTD). For a BTD with block terms of multilinear rank \((l_1^r, \ldots, l_D^r)\), where \(r = 1, \ldots, R\), we can apply Definition 2.2 in which the Tucker core structure \(\mathcal{M}\) is the submanifold of tensors in \(\mathbb{R}^{l_1^r \times \cdots \times l_D^r}\) with multilinear rank \((l_1^r, \ldots, l_D^r)\). Since \(\mathcal{M}\) is an open subset of \(\mathbb{R}^{l_1^r \times \cdots \times l_D^r}\), the canonical basis of \(\mathbb{R}^{l_1^r \times \cdots \times l_D^r}\) is an orthonormal basis of the tangent space to \(\mathcal{M}\) at any point. The algorithm to compute the condition number \(\kappa_{\text{HOSVD}}\) as follows. For each term \(\mathcal{A}_r\) in the BTD, compute its compact HOSVD \((U_1^r, \ldots, U_D^r) \cdot \mathcal{C}_r\). An orthonormal basis of the tangent space to the Tucker manifold is given by the columns of

\[
T_{\mathcal{X}_r} := \left[ \bigotimes_{d=1}^D U_d^\perp \left[ (U_1^d \otimes \cdots \otimes U_{d-1}^d \otimes U_d^\perp e_i(\hat{u}_d^j)^T \otimes U_{d+1}^d \otimes \cdots \otimes U_D^d) \mathcal{C}_r \right]_{d,i,j,l=1}^{d,m_d-l_d,l_d} \right].
\]
where \( \hat{u}_d^j \) and \( U_d^{-1} \) are as in Proposition 4.1. The condition number of the BTD with terms \( A_1, \ldots, A_R \) can then be computed by applying (4.2):

\[
\kappa^{\text{BTD}}(A_1, \ldots, A_R) = \sigma_{\min}(T_{A_1, \ldots, A_R})^{-1}, \quad T_{A_1, \ldots, A_R} = [T_{A_1}, \ldots, T_{A_R}].
\]

**Example 2 (CPD).** This case was studied in [5]. By applying the Tucker core structure \( M := R \setminus \{0\} \) to Definition 2.2, we get the Segre manifold of rank-1 tensors. If \( A_r = \lambda u_1^r \otimes \cdots \otimes u_D^r \) is a rank-1 tensor with \( \|u_1^r\| = \cdots = \|u_D^r\| = 1 \), Proposition 4.1 gives the following familiar basis:

\[
T_{A_r} := [u_1^r \otimes \cdots \otimes u_D^r, (u_1^r \otimes \cdots \otimes u_D^{r-1} \otimes u_1^{r+1} \otimes \cdots \otimes u_D^{D})_{d=1}^D],
\]

where \( U_d^{-1} \) is an orthonormal basis for the complement of \( u_d^r \) for each \( d \). The condition number \( \kappa^{\text{CPD}} \) can be computed in a similar fashion as for the BTD.

### 4.2. Examples of well and ill-conditioned SBTDs

In this subsection, we present some qualitative properties that determine the condition number of the SBTD. As a general rule, the tensor subspace in which the summands live already gives some information about the condition number. For instance, one instance where the condition number is perfect is when the subspaces \( U_d^r \) in the Tucker decompositions are pairwise orthogonal. This can be considered as the SBTD equivalent of an orthogonally decomposable (odeco) tensor [19]. In such cases, tangent spaces are pairwise orthogonal, so that the following result holds.

**Proposition 4.2.** Suppose \( A_1 + \cdots + A_R \) is an SBTD with \( A_r = (U_1^r, \ldots, U_D^r) \cdot C_r \)

in HOSVD form for \( r = 1, \ldots, R \). Assume that \((U_d^r)^T U_d^r = 0\) for each \( d \) and each \( r_1 \neq r_2 \). Then \( \kappa^{\text{SBTD}}(A_1, \ldots, A_r) = 1 \).

**Proof.** The columns of the Terracini matrix can be grouped into orthonormal bases of \( \text{span}\{\otimes_{d=1}^D U_d^r\} \) and \( \text{span}\{U_d^{-1} \otimes_d \otimes_d U_d^r\} \) for all \( r = 1, \ldots, R \) and for all \( d = 1, \ldots, D \). By assumption, these bases are orthonormal. Since their bases are orthonormal, all columns of the Terracini matrix are orthonormal. \( \Box \)

The fact that this result does not depend on the cores \( C_r \) may be surprising if the problem is not considered geometrically. \( C_r \) may be arbitrarily close to having a multilinear rank lower than the specified \((l_1^r, \ldots, l_D^r)\) without it affecting the condition number. Despite this, summands which are close to being low multilinear rank are a notorious obstacle in practical algorithms to compute the BTD, the other being correlations between the terms [24]. Note that the latter is essentially what the condition number measures. For tensors of lower multilinear rank than \((l_1^r, \ldots, l_D^r)\), there are more ways to parametrise it than is accounted for by the usual symmetries. For instance, for a BTD with multilinear ranks \((l_1, l_2, 1)\), the Jacobian of the residual \( A - \sum_{r=1}^R A_r \) with respect to the parameters becomes singular at such points [32]. The ALS algorithm for a general block term decomposition requires solving a system which also becomes singular at the boundary [11].

However, if the summands are considered as one geometric object, summands close to tensors of lower multilinear rank are not an issue, which explains why it is still reasonable to expect the condition number to be 1 even near the boundary. This suggests that Riemannian optimisation algorithms to compute the BTD could have a significant advantage in these cases, as the convergence rate tends to be related to the condition number. This is analogous to the case of the CPD, where [5] showed experimentally and theoretically that classic flat optimization methods perform worse if the CPD contains summands of small norm—the analogous situation to a lower multilinear rank in \( M \)-structured Tucker decompositions—while Riemannian optimisation methods that treat the summands as one geometric object did not suffer as much.

In general, the condition number of any SBTD can be upper bounded by the condition number of the corresponding BTD, and for the latter we can get a useful lower bound for the condition number. We show this in the next proposition.
Proposition 4.3. Given any SBTD $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_R$, we can also regard it as a BTD of $\mathcal{A}$. The condition numbers satisfy

$$\kappa^{\text{BTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) \geq \kappa^{\text{SBTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R).$$

If the terms $\mathcal{A}_r = (U_r^1, \ldots, U_r^p) \cdot \mathcal{C}_r$ are in HOSVD form for $r = 1, \ldots, R$, then

$$\kappa^{\text{BTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) \geq \sigma_{\min}\left(\left[U_1^1 \otimes \cdots \otimes U_D^1\right]^{R}_{r=1}\right)^{-1}.$$

Proof. Assume $\kappa^{\text{BTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) < \infty$, since otherwise the statement is trivially true. For each $r = 1, \ldots, R$, the $r$th structured Tucker manifold $\mathcal{M}_1^{r_1} \cdots n_D$ of the SBTD is a subset of the manifold $\mathcal{N}_1^{r_1} \cdots n_D$ of tensors of fixed multilinear rank. By assumption, the addition map $\Sigma(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \mathcal{A}_1 + \cdots + \mathcal{A}_R$ of the BTD has a local inverse function $\Sigma^{-1}_{\mathcal{A}_1, \ldots, \mathcal{A}_R}$, defined on a neighbourhood $\mathcal{I} \subseteq \Sigma(\mathcal{A}_1, \ldots, n_D^{r_1} \cdots n_D)$ of $\mathcal{A}$. On the other hand, for any $\tilde{\mathcal{A}} \in \mathcal{I}' := \mathcal{I} \cap \Sigma(\mathcal{A}_1, \ldots, n_D^{r_1} \cdots n_D)$, the (locally unique) SBTD of $\tilde{\mathcal{A}}$ is $\Sigma^{-1}_{\mathcal{A}_1, \ldots, \mathcal{A}_R}(\tilde{\mathcal{A}})$. The condition numbers of the BTD and SBTD are (1.2) applied to $\Sigma^{-1}_{\mathcal{A}_1, \ldots, \mathcal{A}_R}$ and the restriction of $\Sigma^{-1}_{\mathcal{A}_1, \ldots, \mathcal{A}_R}$ onto $\mathcal{I}'$, respectively. Since $\mathcal{I}' \subseteq \mathcal{I}$, the first statement follows.

For the second assertion, observe that the columns of $[U_1^1 \otimes \cdots \otimes U_D^1]^{R}_{r=1}$ are a subset of the columns of the Terracini matrix $T$ of the BTD. By [14, Theorem 8.1.7],

$$\sigma_{\min}(U_1^1 \otimes \cdots \otimes U_D^1)^R_{r=1} \geq \sigma_{\min}(T).$$

The result follows from (4.2).

The second item in the above proposition shows that,

$$\text{if } \ker(U_1^1 \otimes \cdots \otimes U_D^1)^R_{r=1} \neq \{0\}, \text{ then } \kappa^{\text{BTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \infty.$$
First, we introduce subspace-constrained SBTDs as the formal model of decompositions resulting from Bro and Andersson’s [7] compress-decompose-expand approach. Subspace-constrained CPDs were also considered in the recent paper [29]. Suppose that we want to compute an SBTD of the tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ associated with the $\mathcal{M}_r$-structured Tucker manifolds $\mathcal{M}_r^{n_1,\ldots,n_D}$. Then compress-decompose-expand proceeds as follows.

**Compress.**

$\mathcal{A}$ lives in a minimal tensor product subspace of $\mathbb{R}^{n_1 \times \cdots \times n_D}$ (possibly trivial). Its minimal Tucker decomposition is $\mathcal{A} = (Q_1, \ldots, Q_D) \cdot \mathcal{G}$ with core tensor $\mathcal{G} \in \mathbb{R}^{m_1 \times \cdots \times m_D}$ and matrices $Q_d \in \mathbb{R}^{n_d \times m_d}$ for $d = 1, \ldots, D$. The decomposition is minimal if $\mathcal{G}$ has multilinear rank equal to $(m_1, \ldots, m_D)$. It can be computed with a (sequentially) truncated higher-order singular value decomposition [10,36].

**Decompose.**

Compute an SBTD of $\mathcal{G}$ be

$$\mathcal{G} = \mathcal{G}_1 + \cdots + \mathcal{G}_R \quad \text{with} \quad \mathcal{G}_r \in \mathcal{M}_r^{m_1,\ldots,m_D}.$$  

That is, the $\mathcal{G}_r$ are $\mathcal{M}_r$-structured Tucker tensors.

**Expand.**

We expand $\mathcal{A}_r = (Q_1 U_{1r}, \ldots, Q_D U_{Dr}) \cdot \mathcal{G}_r$ and find a decomposition $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_R$ of $\mathcal{A}$. In this decomposition the summands are also $\mathcal{M}_r$-structured Tucker tensors; if we have the $\mathcal{M}_r$-structured Tucker decomposition $\mathcal{G}_r = (U_{1r}, \ldots, U_{Dr}) \cdot \mathcal{G}_r$ such that $U_{dr} \in \mathbb{R}^{n_d \times l_d}$ and $\mathcal{G}_r$ is a point of $\mathcal{M}_r \subset \mathbb{R}^{l_1 \times \cdots \times l_D}$ satisfying the assumptions of Definition 2.2, then for all $r$ and $d$ the matrices $Q_d U_{dr}$ are of full rank. Hence $\mathcal{A}_r = (Q_1 U_{1r}, \ldots, Q_D U_{Dr}) \cdot \mathcal{G}_r$ is a point of the $\mathcal{M}_r$-structured Tucker manifold $\mathcal{M}_r^{n_1,\ldots,n_D}$. Summarising, the SBTD (5.1) of the compressed tensor $\mathcal{G}$ can be expanded to an SBTD of $\mathcal{A}$:

$$\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_R \quad \text{with} \quad \mathcal{A}_r \in \mathcal{M}_r^{n_1,\ldots,n_D}, \; r = 1, \ldots, R.$$  

We call the resulting SBTD of $\mathcal{A}$ a subspace-constrained SBTD, because it is an SBTD all of whose summands are contained in the same tensor subspace $Q_1 \otimes \cdots \otimes Q_d$ that $\mathcal{A}$ lives in.

Given that a subspace-constrained SBTD can be computed by the foregoing compress-decompose-expand approach, it is natural to wonder about the relationship between the condition numbers of $\mathcal{A}$ and $\mathcal{G}$. Since $\mathcal{G}$ lives in a much more constrained space, it seems natural to assume that its condition number could be much lower, similar to the ideas in [2]. In subsection 5.1 below, we prove the following main result about the condition numbers of computing the SBTD (5.2) of the original tensor and the SBTD of the compressed Tucker core (5.1). A priori, the condition number of the decomposition problem (5.1) is bounded above by the condition number of problem (5.2). The next results shows that they are, in fact, always equal.

**Theorem 5.1.** Let $\mathcal{M}_r \subset \mathbb{R}^{l_1 \times \cdots \times l_D}$ be Tucker core structures. Assume that the tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ has an orthogonal Tucker decomposition $\mathcal{A} = (Q_1, \ldots, Q_D) \cdot \mathcal{G}$ with $\mathcal{G} \in \mathbb{R}^{m_1 \times \cdots \times m_D}$ and all $Q_d \in \mathbb{R}^{n_d \times m_d}$ having orthonormal columns. Let the subspace-constrained SBTD be $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_R$ with $\mathcal{A}_r \in \mathcal{M}_r^{n_1,\ldots,n_D}$ and the SBTD of the Tucker core be $\mathcal{G} = \mathcal{G}_1 + \cdots + \mathcal{G}_R$ with $\mathcal{G}_r \in \mathcal{M}_r^{m_1,\ldots,m_D}$, and assume that that they are related by $\mathcal{A}_r = (Q_1, \ldots, Q_D) \cdot \mathcal{G}_r$ for each $r = 1, \ldots, R$. Then,

$$\kappa_{\text{SBTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \kappa_{\text{SBTD}}(\mathcal{G}_1, \ldots, \mathcal{G}_R).$$  

Before presenting the proof in subsection 5.1, let us investigate some consequences of Theorem 5.1. In the subspace-constrained SBTD there are two levels of multilinear
multiplication. The first level is in Definition 2.2 of the structured Tucker decompositions. This level is always written with matrices $U_1^1, \ldots, U_2^1$ that depend on the index of the summand. The second level is the multilinear multiplication defining the subspace constraint on the tensor $A$. This level is denoted with matrices $Q_1, \ldots, Q_D$ and it is the same for all summands. This is summarised in the following diagram:

$$A_r \in M_r^{n_1 \cdots n_D} \leftarrow (Q_1 \otimes \cdots \otimes Q_D), \quad G_r \in M_r^{m_1 \cdots m_D} \leftarrow (U_1^r \otimes \cdots \otimes U_D), \quad C_r \in M_r,$$

It is imperative to note, however, that $(Q_1 \otimes \cdots \otimes Q_D)M_r^{m_1 \cdots m_D} \subseteq M_r^{n_1 \cdots n_D}$.

When evaluating the sensitivity of a subspace-constrained SBTD (5.2) via the condition number (1.2), there are at least four natural sets of perturbations $I$ to consider. Let $A$ denote the perturbed tensor. It could have resulted from one of the following increasingly restrictive perturbations of the subspace-constrained SBTD $A$:

1. $A$ was perturbed with no constraints and $A$ was approximated by the closest SBTD $A \approx \hat{A}_1 + \cdots + \hat{A}_R$ with $\hat{A}_r \in M_r^{n_1 \cdots n_D}$;
2. $A$ was perturbed so $A$ has an SBTD $A = \hat{A}_1 + \cdots + \hat{A}_R$ with $\hat{A}_r \in M_r^{n_1 \cdots n_D}$;
3. $A$ was perturbed so $A$ has a subspace-constrained SBTD $\hat{A} = (Q_1, \ldots, Q_D) \cdot \hat{G}$ with core $\hat{G} = \hat{G}_1 + \cdots + \hat{G}_R$ and $\hat{G}_r \in M_r^{m_1 \cdots m_D}$; or
4. $A$ was perturbed inside the fixed subspace $Q_1 \otimes \cdots \otimes Q_D$ so $A$ has a subspace-constrained SBTD $\hat{A} = (Q_1, \ldots, Q_D) \cdot \hat{G}$ with core $\hat{G} = \hat{G}_1 + \cdots + \hat{G}_R$ and terms in the decomposition $\hat{G}_r \in M_r^{m_1 \cdots m_D}$.

Since there are 4 domains of perturbations we can consider in (1.2), there are also 4 associated, a priori distinct, condition numbers. Let us denote the condition number corresponding to the $i$th type of perturbation by $\kappa_i$. Then we have

$$(5.3) \quad \kappa_1 \geq \kappa_2 = \kappa_{\text{SBTD}}(A_1, \ldots, A_R) \geq \kappa_3 \geq \kappa_4 = \kappa_{\text{SBTD}}(G_1, \ldots, G_R).$$

However, we already proved in [6, Corollary 5.5] that $\kappa_1 = \kappa_2$, i.e., arbitrary perturbations in combination with a least-squares approximation are no worse than structured perturbations. Combining this with Theorem 5.1 immediately implies the following more formal restatement of Theorem 1.1.

**Corollary 5.2.** Suppose that we have SBTDs $\hat{G} = \hat{G}_1 + \cdots + \hat{G}_R \in \mathbb{R}^{m_1 \times \cdots \times m_D}$ and $\hat{A} = \hat{A}_1 + \cdots + \hat{A}_R \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ related by $\hat{A}_r = (Q_1 \otimes \cdots \otimes Q_D) \cdot \hat{G}_r$ for each $r = 1, \ldots, R$. If all $Q_d$ have orthonormal columns, then (5.3) is an equality.

**5.1. Proof of the main result.** Proposition 4.1 allows us to prove our main result, Theorem 5.1. Before we do this, we need the following lemma.

**Lemma 5.3.** For any set of matrices $A_k \in \mathbb{R}^{m \times n_k}$ and any set of orthogonal matrices $Q_k \in \mathbb{R}^{p \times p}$ where $k = 1, \ldots, K$, the matrices

$$X := [A_1 \cdots A_K] \quad \text{and} \quad Y := [A_1 \otimes Q_1 \cdots A_K \otimes Q_K]$$

have the same singular values up to multiplicities.

**Proof.** Define the block diagonal matrix $D := \text{diag}(I_{n_1} \otimes Q_1, \ldots, I_{n_K} \otimes Q_K)$. Then $Y = [A_1 \otimes I_p \cdots A_K \otimes I_p] \cdot D$. Since $D$ is orthogonal, $Y$ has the same singular values as $[A_1 \otimes I_p \cdots A_K \otimes I_p]$. Up to a permutation of rows and columns, this is $X \otimes I_p$. The proof is completed by applying the singular value property of Kronecker products [14, section 12.3.1].

**Remark 5.4.** If each $A_k$ in the above lemma is itself a Kronecker product of at least $d$ matrices and we replace $Y$ by $[Q_1 \otimes d A_1 \cdots Q_K \otimes d A_K]$, the statement still holds, because changing the order of the factors only changes their Kronecker product by a permutation of rows and columns [14, Equation 12.3.1].

Now we can prove that the condition number of the SBTD is invariant under Tucker compression.
Proof of Theorem 5.1. For the SBTDs $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_R$ and $\mathcal{G} = \mathcal{G}_1 + \cdots + \mathcal{G}_R$ we denote their associated Terracini matrices by $T_{\mathcal{A}_1,\ldots,\mathcal{A}_n}$ and $T_{\mathcal{G}_1,\ldots,\mathcal{G}_n}$, respectively (see (4.1)). Our strategy is assembling the Terracini matrices in an appropriate way and so that we can compare their singular values.

For each $r = 1, \ldots, R$, we apply Proposition 4.1 to obtain a basis $\mathcal{B}_{\mathcal{G}_r}$ for $T_{\mathcal{G}_r} \mathcal{M}_r^{m_1,\ldots,m_p}$. That is, we write $\mathcal{G}_r$ in HOSVD form $(U_1^r, \ldots, U_D^r) \cdot \mathcal{C}_r$ and compute matrices $U_d^r \perp \perp$ so that $[U_d^r \ U_d^r \perp]$ is orthogonal for each $d$. The vectors $\hat{u}_d^j$ are scaled versions of $e_j^{(U_d^r)\perp}$ as in Proposition 4.1. That is, they are defined such that the transpose of $j$th right singular vector of the $d$th flattening of the $r$th core $(\mathcal{C}_r)_d$ is $(\hat{u}_d^j)^T (\mathcal{C}_r)_d$. This gives the basis

$$
\mathcal{B}_{\mathcal{G}_r} := \{(U_1^r, \ldots, U_D^r) \cdot \hat{\mathcal{C}}_r \} \cup \{(U_1^{r-1}, \ldots, U_d^r \perp, U_d^r \perp e_i (\hat{u}_d^j)^T, U_{d+1}^r, \ldots, U_D^r) \cdot \mathcal{C}_r \}
$$

with $\hat{\mathcal{C}}_r \in \mathcal{C}_r$, $d = 1, \ldots, D$, $i = 1, \ldots, m_d - l_d$ and $j = 1, \ldots, l_d$.

For $T_{\mathcal{G}_r} \mathcal{M}_r^{m_1,\ldots,m_p}$, we can use a basis of the same form, constructed as follows:

We form $Q_d^+$ so that the columns of $[Q_d \ Q_d^+]$ are an orthonormal basis of $\mathbb{R}^{n_d}$. Then define

$$(Q_d U_d^r)^\perp := [Q_d U_d^r \perp, Q_d^+] \in \mathbb{R}^{n_d \times (n_d - l_d)}.$$

The columns of this matrix are a basis for the orthogonal complement of the column space of $Q_d U_d^r$. A basis $\mathcal{B}_d$ of $T_{\mathcal{A}_1} \mathcal{M}_r^{m_1,\ldots,m_p}$ is obtained by applying (4.3) where $Q_d U_d^r$ fulfills the role of $U_d$ and $(Q_d U_d^r)^\perp$ fulfills that of $U_d^r$.

By rearranging the order of the basis vectors and factoring out all $Q_d$ and $Q_d^+$, we get a partition of this basis:

$$(Q_1 \otimes \cdots \otimes Q_D)(\mathcal{B}_{\mathcal{G}_r}) \cup \mathcal{B}_1^r \cup \cdots \cup \mathcal{B}_D^r,$$

where

$$
\mathcal{B}_d^r := \left\{ \left( Q_d^\perp \otimes_{d' \neq d} Q_{d'} \right) \left( e_i^{(n_d - m_d)} (\hat{u}_d^j)^T \otimes_{d' \neq d} U_{d'}^r \right) \mathcal{C}_r \right\}_{i,j=1}^{n_d-m_d,l_d}.
$$

By construction of $Q_d$, the subspaces that for a fixed $r$ are spanned by the $D + 1$ bases in (5.4) are pairwise orthogonal. Therefore, collecting $\mathcal{B}_d$ for all $r$ gives a Terracini matrix of $\mathcal{A}$, which splits into $D + 1$ pairwise orthogonal blocks. Up to a permutation of the columns,

$$(5.5) \ T_{\mathcal{A}_1,\ldots,\mathcal{A}_n} = \left[ (Q_1 \otimes \cdots \otimes Q_D)T_{\mathcal{G}_1,\ldots,\mathcal{G}_n} T_1^\perp \cdots T_D^\perp \right],$$

where the columns of each $T_d^\perp$ are $\mathcal{B}_d^1 \cup \cdots \cup \mathcal{B}_d^D$. Explicitly,

$$
T_d^\perp = \left( Q_d^\perp \otimes_{d' \neq d} Q_{d'} \right) \left[ e_i^{(n_d - m_d)} \otimes_{d' \neq d} (\hat{u}_d^j)^T \otimes_{d' \neq d} U_{d'}^r \right] \mathcal{C}_r \right\}_{i,j=1}^{n_d-m_d,l_d}.
$$

Because the blocks are pairwise orthogonal, the singular values of $T_{\mathcal{A}_1,\ldots,\mathcal{A}_n}$ are the union of those of $T_{\mathcal{G}_1,\ldots,\mathcal{G}_n}$ and those of each $T_d^\perp$ separately.

The factor $(Q_1^\perp \otimes_{d' \neq d} Q_{d'} \right)$ is orthogonal and thus can be omitted for the purpose of computing singular values. By the definition of the Kronecker product, the columns of the remaining factor can be permuted to

$$
\tilde{T}_d^\perp := \left[ e_{n_d-m_d} \otimes_{d' \neq d} (\hat{u}_d^j)^T \otimes_{d' \neq d} U_{d'}^r \right] \mathcal{C}_r \right\}_{i,j=1}^{n_d-m_d,l_d},
$$

hence, its singular values are just those of $T_d^\perp$. We will show that this is effectively a submatrix of $T_{\mathcal{G}_1,\ldots,\mathcal{G}_n}$ so that the desired result follows from the interlacing property of singular values [14, Theorem 8.1.7].
For any \( r \) and \( d \), take all tangent vectors at \( \mathcal{G}_r \) in the set \( \mathcal{V}_r^d \cup \mathcal{W}_r^d \) with 
\[
\mathcal{V}_r^d := \left\{ \left( \begin{pmatrix} U_d \tilde{e}_i^d \end{pmatrix} (\tilde{W}_r^d)^T \otimes_d U_d \right) \mathcal{C}_r \right\}_{i,j=1}^{m_d \times \ell_d^r} \quad \text{and} \quad \mathcal{W}_r^d := \left\{ \left( \begin{pmatrix} U_d \tilde{e}_i^d \end{pmatrix} \otimes_d U_d \right) \mathcal{C}_r \right\}_{i,j=1}^{m_d \times \ell_d^r},
\]

In the proofs of Propositions 3.1 and 3.3, we showed that all vectors in the same form as \( \mathcal{V}_r^d \) are tangent to \( \mathcal{M}_{\mathcal{G}_r}^{(m_1, \ldots, m_D)} \). \( \mathcal{W}_r^d \) is just a subset of \( \mathcal{R}_d \), the basis we used for \( \mathcal{G}_r \). By construction of \( U_d^r \), the spaces \( \mathcal{V}_r^d \) and \( \mathcal{W}_r^d \) are orthogonal. The inner products between the elements of \( \mathcal{V}_r^d \) (respectively, \( \mathcal{W}_r^d \)) are of the same form as (4.4). Hence, they are also zero. By collecting \( \mathcal{V}_r^d \cup \mathcal{W}_r^d \) for all \( r \), we get a subset of the columns of \( T_{\mathcal{G}_1, \ldots, \mathcal{G}_n} \):
\[
\tilde{T}_d^\text{part} := \left[ \begin{pmatrix} U_d^r & U_d^r \end{pmatrix} \otimes_d \left( \begin{pmatrix} \tilde{W}_r^d \end{pmatrix}^T \otimes_d U_d^r \right) \mathcal{C}_r \right]_{r,j=1}^{R_d \times \ell_d^r},
\]

which has the same singular values as \( \tilde{T}^\bot_{d} \) by Lemma 5.3. Hence, the singular values of \( T_d^\text{part} \) are interlaced between those of \( T_{\mathcal{G}_1, \ldots, \mathcal{G}_n} \). By reminding ourselves that (5.5) is a decomposition into pairwise orthogonal blocks, we can see that \( T_{\mathcal{A}_1, \ldots, \mathcal{A}_R} \) and \( T_{\mathcal{G}_1, \ldots, \mathcal{G}_n} \) must have the same extreme singular values.

6. Computing the condition number. One computational advantage of Theorem 5.1 is that for any SBTD that is computed using the compress-decompose-expand strategy from Section 5, the condition number can be computed at a low extra cost right after the decompose phase. That is, it is not necessary to compute the expanded decomposition in order to know its condition number. Furthermore, if a subspace-constrained SBTD \( \mathcal{A}_1 + \cdots + \mathcal{A}_R \) is given, it can be compressed prior to computing its condition number. This gives Algorithm 6.1.

**Algorithm 6.1** Computation of \( \kappa^{\text{SBTD}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) \) with \( \mathcal{A}_r = (U_1^r, \ldots, U_D^r) \cdot \mathcal{C}_r \).

```plaintext
for d = 1, \ldots, D do
    Compute a QR decomposition \( Q_d R_d = [U_1^d, \ldots, U_D^d] \).
end for

for r = 1, \ldots, R do
    Compute \( \mathcal{G}_r \leftarrow (Q_1^r, \ldots, Q_D^r) \cdot \mathcal{A}_r \).
end for

Compute \( \kappa^{\text{SBTD}}(\mathcal{G}_1, \ldots, \mathcal{G}_R) \) using the algorithm from subsection 4.1.
```

This algorithm was applied to the numerical example mentioned in the introduction. Its computational complexity compares to that of the naive approach as follows.

**Proposition 6.1.** Let \( \mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_R \in \mathbb{R}^{n \times \cdots \times n} \) be a subspace-constrained SBTD with core structures \( \mathcal{M}_r \), each \( \mathcal{M}_r \) is an open submanifold of \( \mathbb{R}^{I \times \cdots \times I} \). Assume that the summands \( \mathcal{A}_r \) are given in HOSVD form. Assume that computing the QR and singular value decomposition of an \( m \times n \times n \)-matrix with \( m \geq n \) both take \( O(mn^2) \) arithmetic operations. The number of arithmetic operations involved in applying (4.2) directly and applying Algorithm 6.1 is
\[
O(n^D l^{2D} + n^D R^2 D^2 l^2 (n-l)^2) \quad \text{and} \quad O(D_n R^2 l^2 + R^D + l^D D^2 R^D + D^D l^4 D^4 D^2),
\]
respectively.
whose dimensions are $O(n_2)$ with orthonormal columns. The dimensions of the tensors are $O(t)$ values requires $O(nD p^2)$ time. This gives a time of $O(R n D p + R l^2 (n - l) n D)$ to construct the full Terracini matrix, whose dimensions are $n_2 D \times p$ where $p = R (l D + D (n - l))$. Computing its singular values requires $O(n D p^2) = O(n D R_1 D + n D R_2 D^2 (n - l)^2)$ operations [14].

Next, we consider Algorithm 6.1. The matrices $[U d^1, \ldots, U d^p]$ have $m := R l$ columns and $n$ rows, which gives a complexity of $O(n R_2 D^3)$ for each QR decomposition [14]. Converting each $G_r$ to HOSVD form takes $O(D m D^3 + 1)$ time [36].

Constructing the Terracini matrix is negligible compared to computing its singular values, as before. In this case, the Terracini matrix has dimensions $m D \times q$ where $q = R (l D + D (m - l)) = O(R D + R D^2 D^2)$. The computation of the singular values requires $O(m D q^2) = O(R D l D q^2) = O(R D + 4 l D + 4 D^2 + 4 D)$ operations.

If $n$ is significantly larger than $R l$ in this proposition, the complexity is approximated by $O(n D + R_2 D^2 D^2)$ and $O(R l n D)$, respectively, which shows the superiority of Algorithm 6.1. On the other hand, if $n \ll R l$, the algorithm does not compress the decomposition and merely adds overhead.

7. Numerical experiments. We present a few numerical experiments illustrating the main result, Theorem 5.1, with a sequence of ill-conditioned block term decompositions. All numerical computations were performed on an Intel Xeon CPU E5-2697 v3 running on 16 cores and 126GB memory. The tensor decompositions were computed in MATLAB R2018b with Tensorlab 3.0 [37] and the other computations were performed in Julia v1.6. [3].

De Silva and Lim [12] give an exact parametrisation of a general curve of rank-2 tensors $X_N$ that converges to a rank-3 tensor as $N \to \infty$. In such cases, the condition number diverges to infinity [5]. Given the vectors $x_d$ and $y_d$ for $d = 1, 2, 3$, the sequence $\{X_N\}_{N=1}^\infty$ is given by

$$N \bigotimes_{d=1}^3 \left( x_d + \frac{1}{N} y_d \right) - N \bigotimes_{d=1}^3 x_d = y_1 \otimes x_2 \otimes x_3 + x_1 \otimes y_2 \otimes x_3 + x_1 \otimes x_2 \otimes y_3 + O\left( \frac{1}{N} \right) .$$

This example can easily be generalised to block term decompositions. Take any third-order core tensor $C$ of full multilinear rank and any two sets of full-rank matrices $\{A_d\}_{d=1}^3$ and $\{B_d\}_{d=1}^3$. Then set

$$g_N := \left( N \bigotimes_{d=1}^3 B_d + \frac{1}{N} A_d \right) - N \bigotimes_{d=1}^3 B_d .$$

Both blocks have the same multilinear rank assuming $B_d$ and $B_d + \frac{1}{N} A_d$ have full rank. Similarly to $X_N$, we can see that $g_N$ equals a three-term BTD independent of $N$, plus $o(N^{-1})$ terms. Its condition number diverges as $N \to \infty$ by a special case of [5, Theorem 1.4].

We generated tensors of this model where the all the core tensor $C \in \mathbb{R}^{2 \times 2 \times 1}$ and the matrices $A_1, A_2, A_3 \in \mathbb{R}^{4 \times 2}, A_3 \in \mathbb{R}^{2 \times 1}$ all have standard normally distributed entries and $B_d$ is the $Q$-factor of the QR decomposition of a matrix with standard normal entries. For several values of $N$, we generated 2000 tensors of model (7.1). For each of these we generated an “inflated” version $A_N = (Q_1, Q_2, Q_3) \cdot g_N$ for some $Q_1, Q_2, Q_3$ with orthonormal columns. The dimensions of the tensors are $g_N \in \mathbb{R}^{4 \times 4 \times 4 \times 2}$ and $A_N \in \mathbb{R}^{60 \times 40 \times 40}$. We used Tensorlab’s Gauss–Newton method 111.nlas [37] to compute a two-term $(2, 2, 1)$-BTD of both the (sequences of) tensors $A_N$ and $g_N$ independently. Since $A_N$ has a subspace-constrained BTD with core tensor $g_N$, by Theorem 5.1 their condition numbers are the same. Some built-in optimisations were disabled, namely automatic
Fig. 7.1. Condition number of the BTD of $\hat{G}_N \in \mathbb{R}^{4 \times 4 \times 2}$ and that of $A_N \in \mathbb{R}^{60 \times 40 \times 40}$ from the experiments in Section 7.

Fig. 7.2. Ratio between the estimated forward error based on (1.3) and the true forward error for $\hat{G}_N$ in the experiments in Section 7. Only cases with a residual $\| \hat{G} - G \| / \epsilon \leq 10^{-8}$ were considered.

Fig. 7.3. Number of iterations of btd with applied to $\hat{G}_N \in \mathbb{R}^{4 \times 4 \times 2}$ and $A_N \in \mathbb{R}^{60 \times 40 \times 40}$ from the experiments in Section 7.
Tucker compression and the use of the iterative solver to solve the linear system to compute the quasi–Newton update direction. This is to ensure the same algorithm is used for both tensors. Since 

111.nls stops when the backward error reaches a certain threshold, this generates exact decompositions of nearby tensors, which allows us to compare the forward and backward error.

A violin plot of the condition number of both BTDs is shown in Figure 7.1. The condition number does indeed increase with the parameter $N$. Moreover, the distribution of the condition number of the BTD of $\hat{G}_N$ is the same as that of $A_N$. We did find that the ratio between the computed condition numbers $\hat{\kappa}$ deviated slightly from one in the more ill-conditioned cases. The most extreme case was $\hat{\kappa}(A_1, \ldots, A_R) \approx (1- 2 \cdot 10^{-5})\kappa(\hat{G}_1, \ldots, \hat{G}_R)$ where $\kappa > 10^{12}$. We attribute this to numerical roundoff. These results thus provide a numerical verification of Theorem 5.1.

A major application of the condition number is to estimate the forward error. For a true decomposition $G = \sum_{r=1}^{N} G_r$ and a computed decomposition $\hat{G} = \sum_{r=1}^{N} \hat{G}_r$, the forward error is measured as

$$e = \min_{x \in \mathcal{S}_R} \sqrt{\sum_{r=1}^{N} \| G_r - \hat{G}_{\pi(r)} \|^2},$$

where $\mathcal{S}_R$ is the symmetric group of $R$ elements. By (1.3) we can estimate that $e \leq \kappa^{\text{BTD}} \| G - \hat{G} \|$ as long as the residual $\| G - \hat{G} \|$ is not too large. Figure 7.2 shows that this bound tends to hold when the residual is at most $10^{-8}$.

Finally, condition numbers tend to be related to convergence rate. Because the condition number is equal for $G_N$ and $A_N$, one could expect the convergence behaviour to be the same. This is reflected in the number of iterations, see Figure 7.3. It also shows that convergence gets slower as the condition number increases. The cost per iteration is expected to be a function of only the dimensions of the tensor and the block terms, as only direct linear algebra routines are used to compute the iteration steps [32]. By using the compressed tensor $G_N$ instead of $A_N$, the geometric mean of the speedup per iteration was 9.5. In [7], speedup factors of up to 40 were observed for the ALS algorithm applied to tensors used in chemometrics.

Note that this is the speedup of computing the decomposition. For the sugar data set of [7], the computation of the condition number, as mentioned in the introduction, was sped up by a factor of 15 000 by first Tucker compressing the tensor from size $265 \times 371 \times 7$ to $3 \times 3 \times 3$.

REFERENCES

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, 2008.

[2] B. Arslan, V. Noferini, and F. Tisseur, The structured condition number of a differentiable map between matrix manifolds, with applications, SIAM Journal on Matrix Analysis and Applications, 40 (2019), pp. 774–799.

[3] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A Fresh Approach to Numerical Computing, SIAM Review, 59 (2017), pp. 65–98.

[4] P. Breiding and N. Vannieuwenhoven, Convergence analysis of Riemannian Gauss–Newton methods and its connection with the geometric condition number, Applied Mathematics Letters, 78 (2018), pp. 42–50.

[5] ———, The Condition Number of Join Decompositions, SIAM Journal on Matrix Analysis and Applications, 39 (2018), pp. 287–309.

[6] ———, The Condition Number of Riemannian Approximation Problems, SIAM Journal on Optimization, 31 (2021), pp. 1049–1077.

[7] R. Bro and C. A. Andersson, Improving the speed of multiway algorithms part II: Compression, Chemometrics and Intelligent Laboratory Systems, 42 (1998), pp. 105–113.

[8] P. Burgisser and F. Cucker, Condition, vol. 349 of Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[9] L. De Lathauwer, Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and Uniqueness, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1033–1066.
[10] L. De Lathauwer, B. De Moor, and J. Vandewalle, A multilinear singular value decomposition, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1253–1278.

[11] L. De Lathauwer and D. Nion, Decompositions of a Higher-Order Tensor in Block Terms—Part III: Alternating Least Squares Algorithms, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1067–1083.

[12] V. de Silva and L.-H. Lim, Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1084–1127.

[13] V. Ehrlacher, M. Fuente-Ruiz, and D. Lombardi, SoTT: greedy approximation of a tensor as a sum of Tensor Train. jun 2021.

[14] G. H. Golub and C. F. Van Loan, Matrix computations, vol. 3, JHU press, 2013.

[15] L. Grasedyck, Hierarchical singular value decomposition of tensors, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2029–2054.

[16] W. Greub, Linear Algebra, Springer-Verlag, 2 ed., 1978.

[17] W. Hackbusch and S. Kühn, A new scheme for the tensor representation, Journal of Fourier Analysis and Applications, 15 (2009), pp. 706–722.

[18] F. L. Hitchcock, The Expression of a Tensor or a Polyadic as a Sum of Products, Journal of Mathematics and Physics, 6 (1927), pp. 164–189.

[19] T. G. Kolda, Orthogonal tensor decompositions, SIAM Journal on Matrix Analysis and Applications, 23 (2001), pp. 243–255.

[20] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM Review, 51 (2009), pp. 455–500.

[21] J. M. Landsberg, Tensors: Geometry and Applications Graduate Studies in Mathematics, vol. 128, 2012.

[22] J. M. Lee, Introduction to Smooth Manifolds, Springer New York, 2013.

[23] J. Munkres, Topology, Pearson Education, second ed., 2014.

[24] C. V. de Silva and L.-H. Lim, Tensor-Train decomposition of higher-order tensors, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10891 LNCS (2018), pp. 139–148.

[25] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Annals of Physics, 349 (2014), pp. 117–158.

[26] I. V. Oseledets, Tensor-Train Decomposition, SIAM Journal on Scientific Computing, 33 (2011), pp. 2295–2317.

[27] A.-H. Phan, P. Tichavský, K. Sobolev, K. Sozykin, D. Ermilov, and A. Cichocki, Canonical polyadic tensor decomposition with low-rank factor matrices, in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 4690–4694.

[28] J. R. Rice, A Theory of Condition, SIAM Journal on Numerical Analysis, 3 (1966), pp. 287–310.

[29] R. A. Ricker, Matrix computations, vol. 3, JHU press, 2013.

[30] L. De Lathauwer, B. De Moor, and J. Vandewalle, A multilinear singular value decomposition, SIAM Journal on Matrix Analysis and Applications, 21 (2000), pp. 1253–1278.

[31] L. De Lathauwer and D. Nion, Decompositions of a Higher-Order Tensor in Block Terms—Part III: Alternating Least Squares Algorithms, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1067–1083.

[32] V. de Silva and L.-H. Lim, Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1084–1127.

[33] V. Ehrlacher, M. Fuente-Ruiz, and D. Lombardi, SoTT: greedy approximation of a tensor as a sum of Tensor Train. jun 2021.

[34] G. H. Golub and C. F. Van Loan, Matrix computations, vol. 3, JHU press, 2013.

[35] L. Grasedyck, Hierarchical singular value decomposition of tensors, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2029–2054.

[36] W. Greub, Linear Algebra, Springer-Verlag, 2 ed., 1978.

[37] W. Hackbusch and S. Kühn, A new scheme for the tensor representation, Journal of Fourier Analysis and Applications, 15 (2009), pp. 706–722.

[38] F. L. Hitchcock, The Expression of a Tensor or a Polyadic as a Sum of Products, Journal of Mathematics and Physics, 6 (1927), pp. 164–189.

[39] T. G. Kolda, Orthogonal tensor decompositions, SIAM Journal on Matrix Analysis and Applications, 23 (2001), pp. 243–255.

[40] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM Review, 51 (2009), pp. 455–500.

[41] J. M. Landsberg, Tensors: Geometry and Applications Graduate Studies in Mathematics, vol. 128, 2012.

[42] J. M. Lee, Introduction to Smooth Manifolds, Springer New York, 2013.

[43] J. Munkres, Topology, Pearson Education, second ed., 2014.

[44] C. V. de Silva and L.-H. Lim, Tensor-Train decomposition of higher-order tensors, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10891 LNCS (2018), pp. 139–148.

[45] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Annals of Physics, 349 (2014), pp. 117–158.

[46] I. V. Oseledets, Tensor-Train Decomposition, SIAM Journal on Scientific Computing, 33 (2011), pp. 2295–2317.

[47] A.-H. Phan, P. Tichavský, K. Sobolev, K. Sozykin, D. Ermilov, and A. Cichocki, Canonical polyadic tensor decomposition with low-rank factor matrices, in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 4690–4694.

[48] J. R. Rice, A Theory of Condition, SIAM Journal on Numerical Analysis, 3 (1966), pp. 287–310.

[49] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, Tensor Decomposition for Signal Processing and Machine Learning, IEEE Transactions on Signal Processing, 65 (2017), pp. 3551–3582.

[50] L. Sorber, M. Van Barel, and L. De Lathauwer, Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(Lr, Lr, 1) terms, and a new generalization, SIAM Journal on Optimization, 23 (2013), pp. 695–720.

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

[52] A. Uschmajew and B. Vandereycken, The geometry of algorithms using hierarchical tensors, Linear Algebra and Its Applications, 439 (2013), pp. 133–166.

[53] N. Vannieuwenhoven, Condition numbers for the tensor rank decomposition, Linear Algebra and its Applications, 535 (2017), pp. 35–86.

[54] N. Vannieuwenhoven, R. Vandervelp, and K. Meerbergen, A new truncation strategy for the higher-order singular value decomposition, SIAM Journal on Scientific Computing, 34 (2012), pp. 1027–1052.

[55] N. Verhulst, O. Debals, and L. De Lathauwer, Tensorlab 3.0 - Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization, Conference Record - Asilomar Conference on Signals, Systems and Computers, 32 (2017), pp. 1733–1738.