A Note on Nonclosed Tensor Formats

W. Hackbusch¹

Received: 24 February 2019 / Accepted: 23 August 2019 / Published online: 7 November 2019
© The Author(s) 2019

Abstract

Various tensor formats exist which allow a data-sparse representation of tensors. Some of these formats are not closed. The consequences are (i) possible non-existence of best approximations and (ii) divergence of the representing parameters when a tensor within the format tends to a border tensor outside. The paper tries to describe the nature of this divergence. A particular question is whether the divergence is uniform for all border tensors.

Keywords Tensor representation · Tensor format · Nonclosed tensor format · Numerical instability

Mathematics Subject Classification (2010) 15A69 · 65F99

1 Introduction

Given (finite-dimensional) vector spaces V_j we denote the corresponding tensor space by

$$\mathbf{V} = \bigotimes_{j=1}^d V_j.$$

Since usually the dimension $\prod_{j=1}^d \dim(V_j)$ of $\mathbf{V}$ is rather huge, the numerical treatment of tensor needs special data-sparse representation techniques. The oldest one is the r -term format: given a *representation rank* r we form all tensors which can be written as a sum of r elementary tensors, where $r \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$. This yields the subset

$$\mathcal{R}_r = \left\{ \mathbf{v} \in \mathbf{V} : \mathbf{v} = \sum_{v=1}^r \bigotimes_{j=1}^d v_v^{(j)} \text{ with } v_v^{(j)} \in V_j \right\}$$

Dedicated to the 65th birthday of Volker Mehrmann.

✉ W. Hackbusch
wh@mis.mpg.de

¹ Max-Planck-Institut Mathematik in den Naturwissenschaften, Inselstraße 22, 04103 Leipzig, Germany

of $\mathbf{V}$. Under certain conditions a tensor in $\mathcal{R}_r$ might have an *essentially unique* representation, i.e., different representations only differ by the order of the terms and the scaling of the vectors $\{v_v^{(j)} : 1 \leq j \leq d\}$ (cf. Section 3.2).

Although this approach may be very successful for certain problems, it also has an unpleasant property which we are going to explain.

The *tensor rank* of $\mathbf{v} \in \mathbf{V}$ is defined as the smallest r with $\mathbf{v} \in \mathcal{R}_r$:¹

$$\text{rank}(\mathbf{v}) := \min \{r \in \mathbb{N}_0 : \mathbf{v} \in \mathcal{R}_r\}.$$

This allows to describe $\mathcal{R}_r$ by $\{\mathbf{v} \in \mathbf{V} : \text{rank}(\mathbf{v}) \leq r\}$. In the case of $d = 2$, tensor spaces are isomorphic to matrix spaces. Then the tensor rank coincides with the usual matrix rank. For matrices it is well known that a convergent sequence $\{M_k\}$ with $\text{rank}(M_k) \leq r$ has a limit M with $\text{rank}(M) \leq r$, i.e., the set of matrices of rank $\leq r$ is closed. This is not true for tensors of order $d \geq 3$. As an example consider the tensor-valued function

$$\mathbf{w}(t) := (a + tb) \otimes (a + tb) \otimes (a + tb) \in \otimes^3 V,$$

where $a, b \in V$ are linearly independent vectors (i.e., $V = V_j$ and $\dim(V) \geq 2$). The derivative $\mathbf{v} := \mathbf{w}'(0)$ is the symmetric tensor

$$\mathbf{v} = b \otimes a \otimes a + a \otimes b \otimes a + a \otimes a \otimes b. \quad (1)$$

The derivative can be approximated by Newton's divided difference quotient

$$\tilde{\mathbf{v}}(h) := \frac{1}{h}(\mathbf{w}(h) - \mathbf{w}(0)) \rightarrow \mathbf{v} \quad \text{for } h \rightarrow 0. \quad (2)$$

Remark 1 (a) It can be proved that $\text{rank}(\mathbf{v}) = 3$ for $\mathbf{v}$ in (1).

(b) Since $\mathbf{w}(t)$ is of rank 1, the approximation satisfies $\text{rank}(\tilde{\mathbf{v}}(h)) \leq 2$, i.e., $\tilde{\mathbf{v}}(h) \in \mathcal{R}_2$.

(c) From (a) and (b) we conclude that $\mathcal{R}_2$ is not closed.

Part (c) states that, in general, the r -term representation is not closed. This leads to the notation of the *border rank*

$$\underline{\text{rank}}(\mathbf{v}) := \min \{r : \mathbf{v} \in \overline{\mathcal{R}_r}\}.$$

It is related to the usual rank by $\underline{\text{rank}}(\mathbf{v}) \leq \text{rank}(\mathbf{v}) \leq (\underline{\text{rank}}(\mathbf{v}))^{d-1}$. The first inequality is trivial. For the second one, let $\mathbf{v}_i \in \mathcal{R}_r$ ($r := \underline{\text{rank}}(\mathbf{v})$) be tensors converging to $\mathbf{v}$. Then $\mathbf{v}_i \in \otimes_{j=1}^d U_j^{\min}(\mathbf{v}_i)$ holds for the minimal subspaces $U_j^{\min}(\mathbf{v}_i)$ (cf. Hackbusch [9, Section 6]). $\mathbf{v}_i \in \mathcal{R}_r$ implies $\dim U_j^{\min}(\mathbf{v}_i) \leq r$. [9, Theorem 6.24] proves $\dim U_j^{\min}(\mathbf{v}) \leq r$ for the corresponding subspaces in $\mathbf{v} \in \mathbf{U} := \otimes_{j=1}^d U_j^{\min}(\mathbf{v})$. The maximal rank in $\mathbf{U}$ is bounded by r^{d-1} proving the second inequality (cf. [9, Section 3.2.6.4]).

The nonclosedness implies that a typical approximation problem as

$$\text{find the minimiser of } \inf \{\|\mathbf{v} - \mathbf{w}\| : \mathbf{w} \in \mathcal{R}_r\} \quad \text{for some } \mathbf{v} \in \mathbf{V} \quad (3)$$

might be unsolvable since any convergent sequence $\mathbf{w}_k \in \mathcal{R}_r$ with $\|\mathbf{v} - \mathbf{w}_k\| \rightarrow \inf \|\mathbf{v} - \mathbf{w}\|$ may tend to a tensor $\mathbf{w}$ of larger rank (but with $\underline{\text{rank}}(\mathbf{w}) \leq r$) outside of $\mathcal{R}_r$. Since the border

¹Since we only consider finite-dimensional tensor spaces, all tensors are algebraic tensors, i.e., their rank is finite.

set $\overline{\mathcal{R}_r} \setminus \mathcal{R}_r$ is of measure zero, one might consider this as a marginal problem. However, De Silva–Lim [4] prove that those $\mathbf{v} \in \mathbf{V}$ for which problem (3) is unsolvable have a positive measure. To be precise, this result holds for $\mathbb{R}$ as underlying field. A new result by Qi–Michalek–Lim [13] states that in the complex case this exceptional set is of measure zero.

Many numerical approaches lead to optimisation problems within the set $\mathcal{R}_r$. If the minimiser does not exist, any numerical method is in trouble. As in the case of (2) the coefficients increase to infinity although its sum is bounded. This fact leads to the typical numerical cancellation. The instability of divided difference quotients is well known in numerical mathematics.

The paper is not restricted to the format $\mathcal{R}_r$ but to rather general nonclosed formats. An example of another nonclosed tensor format is the *cyclic matrix product representation*. A special variant is the *site-independent cyclic matrix product representation* for the case of $V_j = V$ (cf. Perez–Garcia et al. [12, Section 3.2.1]). Let $n := \dim(V)$. For tuples $(M[i] : 1 \leq i \leq n)$ of $r \times r$ matrices define the corresponding tensor $\mathbf{v} \in \otimes^d V$ componentwise by

$$\mathbf{v}[i_1, i_2, \dots, i_d] = \text{trace}(M[i_1]M[i_2] \cdot \dots \cdot M[i_d]).$$

Such tensors define the set $\mathcal{C}_{\text{ind}}(d, r, n)$. Already $\mathcal{C}_{\text{ind}}(3, 2, 3)$ and $\mathcal{C}_{\text{ind}}(4, 2, 2)$ are nonclosed (cf. [5]). The general conclusion is that, in general, graph-based tensor format are nonclosed if the graph does not degenerate to a tree (cf. Landsberg [11, Theorem 14.1.2.2]).

In the following we characterise the divergence of the coefficients of a sequence $\mathbf{w}_k \in \mathcal{F}$ converging to a border tensor (i.e., a tensor in $\overline{\mathcal{F}} \setminus \mathcal{F}$). In particular it is interesting to know how strong the divergence is and whether it is uniform for all such tensors. If the divergence could be arbitrarily weak, the numerical instability would be negligible.

We also study the order of divergence in a neighbourhood of a border tensor and whether this quantity behaves continuously.

2 Notations and Definitions

2.1 Tensor Representation

$\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ denotes the field on the following vector spaces. In general, a tensor representation is given by a map ρ from a parameter set into the tensor space. We suppose that

$$P \text{ is a vector space with } \dim(P) < \infty, \quad \mathcal{D} \subset P \text{ a closed subset,} \quad (4a)$$

$$\mathbf{V} \text{ is a tensor space with } \dim(\mathbf{V}) < \infty, \quad \mathcal{F} \subset \mathbf{V} \text{ subset of two-sided cone structure,} \quad (4b)$$

$$\rho : \mathcal{D} \rightarrow \mathcal{F} \quad \text{continuous and surjective,} \quad (4c)$$

$$0 \in \mathcal{D} \quad \text{and } \rho(0) = 0. \quad (4d)$$

In the sequel, (4a–4d) are assumed to be valid.

By definition, the tensor subset $\mathcal{F}$ is the range of ρ . The cone structure ensures that with $\mathbf{v}$ also $\lambda \mathbf{v}$ belongs to $\mathcal{F}$ for all $\lambda \in \mathbb{K}$. In most of the examples, ρ is not injective, and $\mathcal{D} = P$ holds (cf. Section 2.2). The standard representations ρ are multilinear so that (4a–4d) is an easy consequence. In the following, we choose some norms on P and $\mathbf{V}$, both denoted by $\|\cdot\|$. Because of the finite dimensions, the choice of the norm is not essential for the following considerations.

Let $\mathbf{v} = \rho(p)$. The ratio $\|p\|/\|\mathbf{v}\| = \|p\|/\|\rho(p)\|$ may be considered as a *stability measure* for the representation of $\mathbf{v}$ by p (cf. Section 1). Since ρ is not necessarily injective, there might be many p with $\mathbf{v} = \rho(p)$. Therefore, we define

$$\sigma(\mathbf{v}) := \inf\{\|p\| : \mathbf{v} = \rho(p), p \in \mathcal{D}\}.$$

By compactness, the infimum may be replaced by a minimum:

Remark 2 Each $\mathbf{v} \in \mathcal{F}$ has at least one $p_{\mathbf{v}} \in \mathcal{D}$ with $\mathbf{v} = \rho(p_{\mathbf{v}})$ and $\sigma(\mathbf{v}) = \|p_{\mathbf{v}}\|$.

Since, in general, there is no scale invariance (cf. Section 2.2), we shall consider the ratio $\|p\|/\|\mathbf{v}\|$ only for normalised tensors so that $\|p\|/\|\mathbf{v}\| = \|p\|$.

In numerical applications we often work with approximations instead of the exact tensor. Therefore, the quantity $\sigma(\tilde{\mathbf{v}})$ is of interest for $\tilde{\mathbf{v}}$ in a neighbourhood of $\mathbf{v}$. For this purpose we define the ε -neighbourhood of some $\mathbf{v} \in \overline{\mathcal{F}}$ by

$$U_{\mathcal{F},\varepsilon}(\mathbf{v}) := \{\mathbf{w} \in \mathcal{F} : \|\mathbf{v} - \mathbf{w}\| < \varepsilon\}$$

and the stability quantity by

$$\sigma_{\varepsilon}(\mathbf{v}) := \sup\{\sigma(\mathbf{w}) : \mathbf{w} \in U_{\mathcal{F},\varepsilon}(\mathbf{v})\} \quad \text{for } \varepsilon > 0.$$

Note that σ_{ε} is defined for all $\mathbf{v} \in \overline{\mathcal{F}}$ and that $\sigma_{\varepsilon}(\mathbf{v}) = \infty$ may happen.

Since $\sigma_{\varepsilon}(\mathbf{v})$ is weakly decreasing as $\varepsilon \searrow 0$ and bounded from below by zero, the improper limit

$$\sigma_0(\mathbf{v}) := \lim_{\varepsilon \searrow 0} \sigma_{\varepsilon}(\mathbf{v})$$

exists ($\sigma_0(\mathbf{v}) = \infty$ holds if $\sigma_{\varepsilon}(\mathbf{v}) = \infty$ for all $\varepsilon > 0$).

2.2 Standard Example $\mathcal{F} = \mathcal{R}_r$

Let $\mathbf{V} = \bigotimes_{j=1}^d V_j$. In the case of the r -term format $\mathcal{F} = \mathcal{R}_r$ we may choose the vector space $P = (V_1 \times \cdots \times V_d)^r$, the domain $\mathcal{D} = P$, and the mapping

$$\rho\left(\left((v_i^{(j)})_{i=1,\dots,r}\right)_{j=1,\dots,d}\right) = \sum_{i=1}^r \bigotimes_{j=1}^d v_i^{(j)}.$$

A natural choice of the norm on P is $\|(v_i^{(j)})\| = \sqrt{\sum_{i=1}^r \sum_{j=1}^d \|v_i^{(j)}\|^2}$. Since $\rho(\lambda p) = \lambda^d \rho(p)$, the ratio $\|p\|/\|\rho(p)\|$ is not invariant with respect to scaling.

To ensure scalability we may choose $P := \mathbf{V}^r$ instead of which we restrict to the domain of r -tuples of elementary tensors:

$$\mathcal{D} = \{\mathbf{v} \in \mathbf{V} : \text{rank}(\mathbf{v}) \leq 1\}^r = (\mathcal{R}_1)^r.$$

Since $\mathcal{R}_1$ is closed (cf. Hackbusch [9, Lemma 9.11]), $\mathcal{D} \subset P$ satisfies (4a). Now the tensor representation is

$$\rho\left((e_i)_{i=1,\dots,r}\right) = \sum_{i=1}^r e_i \quad (e_i \in \mathcal{R}_1)$$

with the norm $\|(e_i)_{i=1,\dots,r}\| = \sqrt{\sum_{i=1}^r \|e_i\|^2}$. As $\rho(\lambda p) = \lambda \rho(p)$, the ratio $\|p\|/\|\rho(p)\|$ is scale invariant.

3 Nonclosed Formats

As mentioned above, we assume throughout the article that (4a–4d) hold.

3.1 Instability Properties

Now we suppose that the format $\mathcal{F}$ is not closed. Then there is a nonempty set $\mathcal{B}$ such that the closure of $\mathcal{F}$ can be split into

$$\overline{\mathcal{F}} = \mathcal{F} \cup \mathcal{B} \quad (\text{disjoint union}).$$

We call $\mathcal{B}$ the *border set* since in the case of $\mathcal{F} = \mathcal{R}_r$ it consists of tensors with border rank $\leq r$, while the usual rank is $> r$.

Any tensor $\mathbf{v} \in \mathcal{B}$ is the limit of tensors $\mathbf{v}_i$ in $\mathcal{F}$. The next statement is a simple observation, but fundamental for the following.

Lemma 1 *Let $\mathbf{v}_i \in \mathcal{F}$ with $\mathbf{v}_i := \rho(p_i)$ be a convergent sequence with the limit $\mathbf{v} = \lim \mathbf{v}_i$. Then $\sup_i \|p_i\| < \infty$ implies $\mathbf{v} \in \mathcal{F}$. The condition $\sup_i \|p_i\| < \infty$ can be replaced by $\sup_i \sigma(\mathbf{v}_i) < \infty$.*

Proof Since the set $\{p \in \mathcal{D} : \|p\| \leq C\}$ with $C := \sup_i \|p_i\|$ is compact (cf. (4a)), there is a subsequence—again denoted by (p_i) —with $p_i \rightarrow p \in \mathcal{D}$. Continuity of ρ (cf. (4c)) implies that $\mathbf{v} = \lim \rho(p_i) = \rho(p)$. Hence $\mathbf{v}$ belongs to the range $\mathcal{F}$ of ρ , i.e., $\mathbf{v} \in \mathcal{F}$. $\square$

By negation we conclude the following result.

Lemma 2 *Let $\mathbf{v}_i := \rho(p_i) \in \mathcal{F}$ converge to $\mathbf{v} \in \mathcal{B}$. Then $\|p_i\| \rightarrow \infty$.*

Note that $\sigma_\varepsilon(\mathbf{v}) \geq 0$ is well-defined for $\mathbf{v} \in \mathcal{B}$ and $\varepsilon > 0$ since $U_{\mathcal{F},\varepsilon}(\mathbf{v})$ is a nonempty subset of $\mathcal{F}$. A consequence of Lemma 2 is

Conclusion 1 *If $\mathbf{v} \in \mathcal{B}$, then $\sigma_\varepsilon(\mathbf{v}) = \infty$ holds for all $\varepsilon > 0$ and leads to $\sigma_0(\mathbf{v}) = \infty$.*

Proof If $\sigma_\varepsilon(\mathbf{v}) =: C < \infty$ we can choose $\mathbf{v}_i \in U_{\mathcal{F},\varepsilon}(\mathbf{v})$ with $\mathbf{v}_i \rightarrow \mathbf{v}$ and parameters $p_i \in \mathcal{D}$ with $\mathbf{v}_i = \rho(p_i)$ and $\|p_i\| \leq C$ (cf. Remark 2). Lemma 2 yields the contradiction $\mathbf{v} \notin \mathcal{B}$. $\square$

In Section 3.2 we shall comment on the continuity of σ . A general negative result follows.

Remark 3 If $\mathcal{B} \neq \emptyset$ (i.e., if the format is nonclosed), σ is discontinuous at $0 \in \mathbf{V}$.

Proof (4d) implies $\sigma(0) = 0$. Assume that σ is continuous at 0. There is a neighbourhood $U_{\mathcal{F},\varepsilon}(0) =: U_{\mathcal{F},\varepsilon}$ for some $\varepsilon > 0$ with $\sigma(\mathbf{v}) \leq 1$ for all $\mathbf{v} \in U_{\mathcal{F},\varepsilon}$. Hence, $\sigma_{\varepsilon-\|\mathbf{w}\|}(\mathbf{w}) \leq 1$ holds for all $\mathbf{w} \in U_{\mathcal{F},\varepsilon}$. Conclusion 1 implies that $\overline{U_{\mathcal{F},\varepsilon}} \cap \mathcal{B} = \emptyset$. On the other hand, there is some $0 \neq \mathbf{v} \in \mathcal{B}$. The cone property (4b) implies that $\lambda \mathbf{v} \in \mathcal{B}$ for all $\lambda \neq 0$. For sufficiently small $\lambda \neq 0$, $\lambda \mathbf{v} \in \overline{U_{\mathcal{F},\varepsilon}}$ yields the contradiction. $\square$

3.2 Discussion of $\mathcal{F} = \mathcal{R}_r$

Let $\mathbb{K} = \mathbb{C}$ be the underlying field. An interesting question is whether the quantity $\sigma(\mathbf{v})$ is continuous. As known from algebraic geometry *general tensors*² in $\mathcal{R}_r$ admit only finitely many (essentially different) decompositions (cf. Section 1) and these decompositions depend continuously on the tensor, at least for r not too large. Then $\sigma_\varepsilon(\mathbf{v}) < \infty$ holds for sufficiently small $\varepsilon > 0$ and has the limit $\sigma_0(\mathbf{v}) = \sigma(\mathbf{v})$.

A particular positive result holds if the representation $\rho : \mathcal{D}_0 \subset \mathcal{D} \rightarrow \mathbf{V}$ is injective for a certain subset $\mathcal{D}_0$ and the inverse map—the decomposition— $\rho^{-1} : \mathcal{F}_0 := \rho(\mathcal{D}_0) \rightarrow \mathcal{P}$ is continuous. Then $\sigma_\varepsilon(\mathbf{v})$ is bounded for $\mathbf{v} \in \mathcal{F}_0$ and $\sigma(\mathbf{v}) = \sigma_0(\mathbf{v})$. This situation occurs under the conditions studied in Sørensen et al. [14–16] and Domanov–De Lathauwer [6–8].

Above we require that r be not too large. In the case of $d = 3$ and $\mathbf{V} = \mathbb{K}^n \otimes \mathbb{K}^m \otimes \mathbb{K}^p$ the concrete condition is as follows. For $r \leq (n-1)(m-1)$ general tensors have a *unique* decomposition as stated in Domanov–De Lathauwer [7, Corollary 1.7]. However, if $r \leq (n-1)(m-1) + 1$ and³ $\mathbb{K} = \mathbb{C}$, general tensors have *finitely* many decompositions (cf. Chiantini–Ottaviani [2, Proposition 5.4]).

The term ‘general tensor’ admits the existence of exceptional tensors. A particular exceptional situation holds for the tensor in

Example 1 Set $\mathbf{V} = \otimes^3 V$ with $\dim(V) \geq 2$ and choose any linearly independent vectors $a, b \in V$. In the case of the 2-term format $\mathcal{F} = \mathcal{R}_2$, the tensor⁴

$$\mathbf{v}(t) := (a + tb) \otimes a \otimes a + a \otimes b \otimes a + a \otimes a \otimes b$$

belongs to $\mathcal{B}$ for $t > 0$, while $\mathbf{v}(0) \in \mathcal{F}$.

Proof (a) For $t = 0$ we rewrite $\mathbf{v}(0)$ as $a \otimes a \otimes (a + b) + a \otimes b \otimes a \in \mathcal{R}_2 = \mathcal{F}$.

(b) Let $t > 0$. For c linearly independent of a , the tensor $\mathbf{w} := c \otimes a \otimes a + a \otimes c \otimes a + a \otimes a \otimes c$ is well known to be in $\mathcal{B} = \overline{\mathcal{R}_2} \setminus \mathcal{R}_2$ (cf. Remark 1a). Let φ be an isomorphism on V with $\varphi(a) = a$ and $\varphi(c) = \frac{1}{t}c$, while $\psi = t \cdot id$. Then $\Lambda := \psi \otimes \varphi \otimes \varphi$ is an isomorphism on $\mathbf{V}$ with $\mathbf{u} := \Lambda(\mathbf{w}) = tc \otimes a \otimes a + a \otimes c \otimes a + a \otimes a \otimes c$. Hence also $\mathbf{u} \in \mathcal{B}$. Substitution $c = b + \frac{1}{t+2}a$ yields $\mathbf{u} = \mathbf{v}(t)$ and proves $\mathbf{v}(t) \in \mathcal{B}$ for $t > 0$. $\square$

The interesting conclusion from Example 1 is that the set $\mathcal{B}$ is not closed. We define

$$\overline{\mathcal{B}} = \mathcal{B} \cup \partial\mathcal{B} \quad (\text{disjoint union}). \quad (5)$$

Note that the tensor $\mathbf{v}(0)$ defined in Example 1 belongs to $\partial\mathcal{B}$.

3.3 General Case

If $\mathcal{B}$ is nonempty, also $\partial\mathcal{B} \neq \emptyset$ holds since

$$0 \in \partial\mathcal{B} \quad (6)$$

²The term ‘general’ tensor means all tensors for which a polynomial does not vanish. The set of exceptional tensors is of measure zero.

³The case of $\mathbb{K} = \mathbb{R}$ is more involved (cf. Angelini–Bocci–Chiantini [1, Theorem 4.2]).

⁴Private communication by M. Michałek.

is always true (consider $\lambda \mathbf{v}$ with $\mathbf{v} \in \mathcal{B}$ for $\lambda \rightarrow 0$ and note that $0 \in \mathcal{F}$ because of (4d)). Example 1 ensures that $\partial \mathcal{B}$ may also contain *nontrivial* tensors of $\mathcal{F} = \mathcal{R}_2$. We remark that

$$\partial \mathcal{B} \subset \mathcal{F} \quad (7)$$

since $\partial \mathcal{B} \subset \overline{\mathcal{B}} \subset \overline{\mathcal{F}} = \mathcal{F} \cup \mathcal{B}$ and $\partial \mathcal{B} \cap \mathcal{B} = \emptyset$ (cf. (5)).

Conclusion 2 *Let $0 \neq \mathbf{v} \in \partial \mathcal{B}$. Then $\sigma_\varepsilon(\mathbf{v}) = \infty$ for all $\varepsilon > 0$ and $\sigma_0(\mathbf{v}) = \infty$, although $\sigma(\mathbf{v}) < \infty$.*

Proof By definition of $\partial \mathcal{B}$ there is some $\mathbf{w} \in \mathcal{B}$ with $0 < \eta := \|\mathbf{v} - \mathbf{w}\| < \varepsilon/2$ and $\sigma_\eta(\mathbf{w}) = \infty$ (cf. Conclusion 1). Since $U_{\mathcal{F},\eta}(\mathbf{w}) \subset U_{\mathcal{F},\varepsilon}(\mathbf{v})$, $\sigma_\varepsilon(\mathbf{v}) \geq \sigma_\eta(\mathbf{w})$ yields the assertion. $\square$

A consequence is the discontinuity of σ on $\partial \mathcal{B}$. Conclusion 2 ensures the existence of a sequence $\mathbf{v}_i \rightarrow \mathbf{v}$ with $\lim \sigma(\mathbf{v}_i) \rightarrow \infty > \sigma(\mathbf{v})$. This proves:

Conclusion 3 *σ is not continuous at $\mathbf{v} \in \partial \mathcal{B} \setminus \{0\}$.*

4 On the Strength of Divergence

The numerical instability is caused by the fact that $\sigma(\mathbf{v}_i) \rightarrow \infty$ holds for any sequence $\mathcal{F} \ni \mathbf{v}_i \rightarrow \mathbf{v} \in \mathcal{B}$. Whether this is a severe problem or not depends on the order of divergence. The introductory example (2) is the classical one-sided difference quotient. Using the step size h for $\mathbf{v}(h)$, we get $\sigma(\tilde{\mathbf{v}}(h)) = \mathcal{O}(1/h)$ and the accuracy $\varepsilon = \mathcal{O}(h)$. Expressing the quantity σ as a function of the accuracy ε , $\sigma = \mathcal{O}(1/\varepsilon)$ shows divergence of first order. However, this is not the general behaviour. We may choose the central difference quotient. Since still $\sigma(\tilde{\mathbf{v}}(h)) = \mathcal{O}(1/h)$ but $\varepsilon = \mathcal{O}(h^2)$, we now have the weaker divergence $\sigma = \mathcal{O}(1/\sqrt{\varepsilon})$.

This example shows that, given an accuracy $\varepsilon > 0$, we have to look for an approximation $\mathbf{w} \in \mathcal{F}$ with minimal $\sigma(\mathbf{w})$. This leads to the following definition.

Definition 1 Let $\mathbf{v} \in \mathcal{B}$ and $\varepsilon > 0$. The instability of the approximation problem in $\mathcal{F}$ is described by

$$\delta(\mathbf{v}, \varepsilon) := \inf\{\sigma(\mathbf{w}) : \mathbf{w} \in U_{\mathcal{F},\varepsilon}(\mathbf{v})\}.$$

Note that $\delta(\mathbf{v}, \varepsilon)$ is the infimum, whereas $\sigma_\varepsilon(\mathbf{v})$ is the supremum over the same set. Again, $\delta(\mathbf{v}, \varepsilon)$ diverges as $\varepsilon \rightarrow 0$.

Proposition 1 *Weakly monotone divergence $\delta(\mathbf{v}, \varepsilon) \nearrow \infty$ holds for all $\mathbf{v} \in \mathcal{B}$ as $\varepsilon \searrow 0$.*

Proof For an indirect proof assume that $\delta(\mathbf{v}, \varepsilon) \leq K < \infty$ for all $\varepsilon = \frac{1}{n} > 0$. Then, for any $n \in \mathbb{N}$, there are $\mathbf{w}_n \in U_{\mathcal{F},1/n}$ (i.e., $\mathbf{w}_n \in \mathcal{F}$ and $\|\mathbf{v} - \mathbf{w}_n\| \leq 1/n$) with $\sigma(\mathbf{w}_n) \leq K + 1$. Since $\mathbf{w}_n \rightarrow \mathbf{v}$, Lemma 1 proves the contradicting statement $\mathbf{v} \in \mathcal{F}$. $\square$

4.1 Uniform Strength of Divergence

4.1.1 Definitions

The function $\delta(\mathbf{v}, \cdot)$ is the exact description of the kind of divergence at $\mathbf{v} \in \mathcal{B}$. In the case of the model example (1) and $\mathcal{F} = \mathcal{R}_2$ we have seen that the divergence is not stronger than

$\mathcal{O}(1/\sqrt{\varepsilon})$ —so that $\delta(\mathbf{v}, \varepsilon) \lesssim 1/\sqrt{\varepsilon}$ —if we use the central difference quotient. There are difference formulae of higher consistency order, but they involve more than two terms, i.e., such approximations are not in $\mathcal{R}_2$. This leads to the conjecture that $\delta(\mathbf{v}, \varepsilon) \sim 1/\sqrt{\varepsilon}$ holds for $\mathbf{v}$ in (1). The next question is as to whether this behaviour might hold for all $\mathbf{v} \in \mathcal{B}$. The answer will be negative. There are particular tensors which behave differently. On the algebraic side, one might expand the difference quotient with step size h into a power series $\mathbf{v} + \sum_j \mathbf{v}_j h^j$. In general, the central difference leads to $\mathbf{v}_1 = 0$ and some $\mathbf{v}_2$. However, for certain $\mathbf{v}$ it might happen that $\mathbf{v}_2 = \mathbf{v}_3 = \dots = \mathbf{v}_{k-1} = 0$ so that $\delta(\mathbf{v}, \varepsilon) \sim \varepsilon^{-1/k}$. The characterisation of the largest possible k seems to be an unsolved problem. In the following we try to treat this problem by analytic tools.

4.1.2 Uniform Divergence

The strongest formulation of uniform divergence would be an inequality

$$\delta(\mathbf{v}, \varepsilon) \geq \delta_0(\varepsilon) \quad \text{for all } \mathbf{v} \in \mathcal{B} \text{ with } \|\mathbf{v}\| = 1, \quad (8a)$$

where

$$\delta_0(\varepsilon) \nearrow \infty \quad \text{as} \quad \varepsilon \searrow 0. \quad (8b)$$

The best possible δ_0 satisfying (8a) is

$$\begin{aligned} \delta_0(\varepsilon) &:= \inf\{\delta(\mathbf{v}, \varepsilon) : \mathbf{v} \in \mathcal{B}, \|\mathbf{v}\| = 1\} \\ &= \inf\{\sigma(\mathbf{w}) : \mathbf{w} \in \mathcal{F}, \|\mathbf{v} - \mathbf{w}\| \leq \varepsilon, \mathbf{v} \in \mathcal{B}, \|\mathbf{v}\| = 1\}. \end{aligned} \quad (9)$$

By definition, $\delta_0(\varepsilon)$ is weakly increasing. The crucial question is whether $\lim_{\varepsilon \rightarrow 0} \delta_0(\varepsilon) < \infty$ or $= \infty$.

Proposition 2 *Uniform divergence as in (8a–8b) holds if and only if $\mathcal{B} \cup \{0\}$ is closed.*

Proof As remarked in (6), zero does not belong to $\mathcal{B}$, but to its closure. Therefore, closedness of $\mathcal{B} \cup \{0\}$ means that $\overline{\mathcal{B}}$ contains no *nontrivial* tensor. In particular $\mathcal{B}_1 := \mathcal{B} \cap \{\mathbf{v} : \|\mathbf{v}\| = 1\}$ would be closed.

(a) Let $\mathcal{B} \cup \{0\}$ be closed. For an indirect proof assume $\lim_{\varepsilon \rightarrow 0} \delta_0(\varepsilon) =: K < \infty$. Then for any $\varepsilon = 1/n$, $n \in \mathbb{N}$, there are $\mathbf{v}_n \in \mathcal{B}_1$ and $\mathbf{w}_n \in \mathcal{F}$ with $\sigma(\mathbf{w}_n) \leq K + 1$ and $\|\mathbf{v}_n - \mathbf{w}_n\| \leq 1/n$. By compactness we may take subsequences—again denoted by $\mathbf{v}_n, \mathbf{w}_n$ —so that $\mathbf{v}_n \rightarrow \mathbf{v}$ and $\mathbf{w}_n \rightarrow \mathbf{w}$. Since $\mathcal{B}_1$ is closed, we obtain $\mathbf{v} \in \mathcal{B}_1 \subset \mathcal{B}$. As $\sigma(\mathbf{w}_n)$ is uniformly bounded, the limit belongs to $\mathcal{F}$ (cf. Lemma 1), i.e., $\mathbf{w} \in \mathcal{F}$. Now $\|\mathbf{v}_n - \mathbf{w}_n\| \leq 1/n$ yields the contradiction $\mathbf{v} = \mathbf{w}$ ($\mathcal{F}$ and $\mathcal{B}$ are disjoint!).

(b) If $\mathcal{B} \cup \{0\}$ is not closed, there is some $0 \neq \mathbf{w} \in \partial\mathcal{B} := \overline{\mathcal{B}} \setminus \mathcal{B}$. Thanks to the cone property (4b), we may assume without loss of generality that $\|\mathbf{w}\| = 1$. Note that $\partial\mathcal{B} \subset \mathcal{F}$ (cf. (7)). Hence $\mathbf{w}$ has a finite value $\omega := \sigma(\mathbf{w})$. For any $\varepsilon > 0$ we find some $\mathbf{v} \in \mathcal{B}_1$ with $\|\mathbf{v} - \mathbf{w}\| \leq \varepsilon$. Now (9) implies that $\delta_0(\varepsilon) \leq \omega$ for all $\varepsilon > 0$, i.e., the property (8b) is not valid. $\square$

In the interesting case of $\mathcal{F} = \mathcal{R}_r$ we know that $\mathcal{B} \cup \{0\}$ is not closed (cf. Example 1). Hence uniform divergence (8a–8b) does not hold for $\mathcal{F} = \mathcal{R}_r$. Nevertheless it is possible to refine the definition of divergence.

4.1.3 Weaker Form of Uniform Divergence

In the case of $\mathcal{F} = \mathcal{R}_r$, the exceptional set $\partial\mathcal{B} = \overline{\mathcal{B}} \setminus \mathcal{B}$ is a rather small subset of $\mathcal{F}$. In the following we formulate an inequality involving the distance from $\partial\mathcal{B}$.

Theorem 4 *There is a function δ_1 with*

$$\delta_1(\varepsilon) \nearrow \infty \quad \text{as} \quad \varepsilon \searrow 0 \quad (10a)$$

such that

$$\delta(\mathbf{v}, \varepsilon) \geq \text{dist}(\mathbf{v}, \partial\mathcal{B}) \delta_1(\varepsilon) \quad \text{for all } \mathbf{v} \in \mathcal{B} \text{ with } \|\mathbf{v}\| = 1. \quad (10b)$$

Proof (a) If $\text{dist}(\mathbf{v}, \partial\mathcal{B}) = 0$, the estimate $\delta(\mathbf{v}, \varepsilon) \geq 0$ is trivial.

(b) In the following we consider those $\mathbf{v}$ with $\mathbf{v} \in \mathcal{B}$, $\|\mathbf{v}\| = 1$, and $\text{dist}(\mathbf{v}, \partial\mathcal{B}) > 0$. In this case the best possible $\delta_1(\varepsilon)$ is

$$\begin{aligned} \delta_1(\varepsilon) &:= \inf\{\delta(\mathbf{v}, \varepsilon)/\text{dist}(\mathbf{v}, \partial\mathcal{B}) : \mathbf{v} \in \mathcal{B}, \|\mathbf{v}\| = 1, \text{dist}(\mathbf{v}, \partial\mathcal{B}) > 0\} \\ &= \inf\{\sigma(\mathbf{w})/\text{dist}(\mathbf{v}, \partial\mathcal{B}) : \mathbf{w} \in \mathcal{F}, \|\mathbf{v} - \mathbf{w}\| \leq \varepsilon, \mathbf{v} \in \mathcal{B}, \|\mathbf{v}\| = 1, \text{dist}(\mathbf{v}, \partial\mathcal{B}) > 0\}. \end{aligned}$$

δ_1 is weakly increasing as $\varepsilon \rightarrow 0$. For an indirect proof of (10a) we assume that $\delta_1(\varepsilon) \leq K < \infty$. As in the proof of Proposition 2 there are convergent subsequences $\mathbf{w}_n \in \mathcal{F}$, $\mathbf{v}_n \in \mathcal{B}$ with

$$\begin{aligned} \mathbf{v} &= \lim_{n \rightarrow \infty} \mathbf{v}_n, \quad \mathbf{w} = \lim_{n \rightarrow \infty} \mathbf{w}_n, \\ \|\mathbf{v}_n - \mathbf{w}_n\| &\leq \frac{1}{n}, \quad \mathbf{w}_n \in \mathcal{F}, \quad \mathbf{v}_n \in \mathcal{B}, \quad \|\mathbf{v}_n\| = 1, \quad \text{dist}(\mathbf{v}_n, \partial\mathcal{B}) > 0, \quad \text{and} \\ \sigma(\mathbf{w}_n) &\leq (K + 1)\text{dist}(\mathbf{v}_n, \partial\mathcal{B}). \end{aligned}$$

We conclude from (6) that $\text{dist}(\mathbf{v}_n, \partial\mathcal{B}) \leq \text{dist}(\mathbf{v}_n, 0) = \|\mathbf{v}_n\| = 1$. Therefore $\sigma(\mathbf{w}_n) \leq K + 1$ implies

$$\mathbf{w} \in \mathcal{F} \quad (11)$$

(cf. Lemma 1).

Next we check the limit $\text{dist}(\mathbf{v}, \partial\mathcal{B}) = \lim_{n \rightarrow \infty} \text{dist}(\mathbf{v}_n, \partial\mathcal{B}) > 0$. Assume that $\text{dist}(\mathbf{v}_n, \partial\mathcal{B}) \rightarrow 0$. Then also $\sigma(\mathbf{w}_n) \leq (K + 1)\text{dist}(\mathbf{v}_n, \partial\mathcal{B}) \rightarrow 0$. By Remark 2 there are parameters $p_n \in \mathcal{D}$ with $\mathbf{w}_n = \rho(p_n)$ and $\sigma(\mathbf{w}_n) = \|p_n\|$. Now $\sigma(\mathbf{w}_n) = \|p_n\| \rightarrow 0$ proves $p_n \rightarrow 0$, while (4c–4d) show that $\mathbf{w} = \lim \mathbf{w}_n = \lim \rho(p_n) = \rho(0) = 0$. However, since the norm is continuous, $\|\mathbf{w}\| = 0$ is a contradiction to $\|\mathbf{w}\| = 1$. The latter equality follows from $\|\mathbf{v}_n - \mathbf{w}_n\| \leq \frac{1}{n}$ and $\|\mathbf{v}_n\| = 1$. Hence $\lim_{n \rightarrow \infty} \text{dist}(\mathbf{v}_n, \partial\mathcal{B}) = \text{dist}(\mathbf{v}, \partial\mathcal{B}) > 0$ holds and implies that $\mathbf{v} \notin \partial\mathcal{B}$. Since $\mathbf{v}_n \in \mathcal{B}$, the limit $\mathbf{v}$ is in $\overline{\mathcal{B}} = \mathcal{B} \cup \partial\mathcal{B}$ (cf. (5)) and $\mathbf{v} \notin \partial\mathcal{B}$ proves

$$\mathbf{v} \in \mathcal{B}. \quad (12)$$

From $\|\mathbf{v}_n - \mathbf{w}_n\| \leq \frac{1}{n}$ we conclude $\mathbf{v} = \mathbf{w}$ which is a contradiction since both tensors are in disjoint sets (cf. (11), (12)). $\square$

The interpretation of Theorem 4 depends on the topological structure of $\partial\mathcal{B}$ as seen next.

Remark 4 If $\partial\mathcal{B}$ is closed, the distance $\text{dist}(\mathbf{v}, \partial\mathcal{B})$ is positive for all $\mathbf{v} \in \mathcal{B}$. This yields a nontrivial estimate (10b) for all $\mathbf{v} \in \mathcal{B}$.

Proof $\mathbf{v} \in \mathcal{B}$ and $\partial\mathcal{B} \subset \mathcal{F}$ implies $\mathbf{v} \notin \partial\mathcal{B}$. Note that $\text{dist}(\mathbf{v}, \partial\mathcal{B}) = 0$ for a closed set $\partial\mathcal{B}$ is equivalent to $\mathbf{v} \in \partial\mathcal{B}$. $\square$

Finally we consider the case of a nonclosed set $\partial\mathcal{B}$. We split the closure $\overline{\partial\mathcal{B}}$ into disjoint sets

$$\overline{\partial\mathcal{B}} = \partial\mathcal{B} \cup \mathcal{C}.$$

Remark 5 (a) $\mathcal{C}$ is a subset of $\mathcal{B}$. (b) $\text{dist}(\mathbf{v}, \partial\mathcal{B}) = 0$ holds for $\mathbf{v} \in \mathcal{B}$ if and only if $\mathbf{v} \in \mathcal{C}$.

Proof (a) $\partial\mathcal{B} \subset \overline{\mathcal{B}}$ implies $\overline{\partial\mathcal{B}} \subset \overline{\mathcal{B}} = \mathcal{B} \cup \partial\mathcal{B}$ and $\mathcal{C} \subset \mathcal{B} \cup \partial\mathcal{B}$. Since $\mathcal{C} \cap \partial\mathcal{B} = \emptyset$, $\mathcal{C} \subset \mathcal{B}$ is proved.

(b) Note that $\text{dist}(\mathbf{v}, \partial\mathcal{B}) = 0$ is equivalent to $\text{dist}(\mathbf{v}, \overline{\partial\mathcal{B}}) = 0$. In the latter case there is some $\mathbf{w} \in \overline{\partial\mathcal{B}}$ with $\|\mathbf{v} - \mathbf{w}\| = \text{dist}(\mathbf{v}, \overline{\partial\mathcal{B}}) = 0$, i.e., $\mathbf{v} = \mathbf{w}$. Comparing $\mathbf{v} \in \mathcal{B}$ and $\mathbf{w} \in \overline{\partial\mathcal{B}} = \partial\mathcal{B} \cup \mathcal{C}$ and noting that $\partial\mathcal{B} \subset \mathcal{F}$, it follows that $\mathbf{v} = \mathbf{w} \in \mathcal{C}$. $\square$

In case of a nonclosed $\partial\mathcal{B}$, the estimate (10b) degenerates to $\delta(\mathbf{v}, \varepsilon) \geq 0$ if and only if $\mathbf{v} \in \mathcal{C}$.

Remark 6 For $\mathcal{F} = \mathcal{R}_r$ it is not hard to prove that the divergence behaviour only depends on the border rank and the order d of the tensor, but not on $\dim(V_j)$.

4.1.4 Example $\otimes^3 \mathbb{R}^2$

Obviously, it is interesting to know more about the topological structure of $\mathcal{B}$ for various nonclosed tensor formats. Finally we consider the tensor space

$$\mathbf{V} = \otimes^3 \mathbb{R}^2$$

which is the smallest nontrivial example. The maximal rank in $\mathbf{V}$ is 3 (cf. Kruskal [10]). Hence $\mathcal{R}_3$ coincides with $\mathbf{V}$ and is obviously closed. As seen by the tensor (1), $\mathcal{R}_2$ is not closed. In fact, (1) describes all border tensors up to tensor space isomorphisms:

$$\mathcal{B} = \left\{ \left(\bigotimes_{j=1}^3 \phi^{(j)} \right) (\mathbf{v}) : \phi^{(j)} \in L(\mathbb{R}^2, \mathbb{R}^2) \text{ isomorphism} \right\},$$

where $\mathbf{v}$ is defined in (1) with $\{a, b\}$ being a fixed basis of $\mathbb{R}^2$. Let $\phi^{(2)} = \phi^{(3)}$ be the identity and define $\phi^{(1)}$ by $\phi^{(1)}(a) = a$, $\phi^{(1)}(b) = a + tb$. For $t \neq 0$, $\phi^{(1)}$ is an isomorphism, whereas for $t = 0$ it is not invertible. Note that with these mappings $\left(\bigotimes_{j=1}^3 \phi^{(j)} \right) (\mathbf{v})$ coincides with the tensor in Example 1. For $t = 0$ we obtain the tensor

$$\mathbf{w}_1 = a \otimes a \otimes a + a \otimes b \otimes a + a \otimes a \otimes b \in \partial\mathcal{B}.$$

The same construction with respect to the directions $j = 2$ and $j = 3$ yields

$$\mathbf{w}_2 = b \otimes a \otimes a + a \otimes a \otimes a + a \otimes a \otimes b \in \partial\mathcal{B},$$

$$\mathbf{w}_3 = b \otimes a \otimes a + a \otimes b \otimes a + a \otimes a \otimes a \in \partial\mathcal{B}.$$

We obtain all tensors in $\partial\mathcal{B}$ by $\left(\bigotimes_{j=1}^3 \phi^{(j)} \right) (\mathbf{v})$ when at least one $\phi^{(j)} \in L(\mathbb{R}^2, \mathbb{R}^2)$ is not invertible. Such a tensor can be written as $\left(\bigotimes_{j=1}^3 \psi^{(j)} \right) (\mathbf{w}_i)$ with $\mathbf{w}_i$, $i \in \{1, 2, 3\}$, and general linear maps $\psi^{(j)} \in L(\mathbb{R}^2, \mathbb{R}^2)$, i.e.,

$$\partial\mathcal{B} = \bigcup_{1 \leq i \leq 3} \left\{ \left(\bigotimes_{j=1}^3 \psi^{(j)} \right) (\mathbf{w}_i) : \psi^{(j)} \in L(\mathbb{R}^2, \mathbb{R}^2) \right\}.$$

Since $L(\mathbb{R}^2, \mathbb{R}^2)$ is closed we obtain the desired result:

Proposition 3 $\partial\mathcal{B}$ is closed.

Acknowledgments Open access funding provided by Max Planck Society. I thank Mateusz Michałek (Leipzig) for many instructive discussions. From him I got better insight into the nature of the set $\partial\mathcal{B}$.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Angelini, E., Bocci, C., Chiantini, L.: Real identifiability vs. complex identifiability. *Linear Multilinear Algebra* **66**, 1257–1267 (2018)
2. Chiantini, L., Ottaviani, G.: On generic identifiability of 3-tensors of small rank. *SIAM J. Matrix Anal. Appl.* **33**, 1018–1037 (2012)
3. Coppi, R., Bolasco, S. (eds.): *Multway Data Analysis*. North-Holland, Amsterdam (1989)
4. De Silva, V., Lim, L.H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.* **30**, 1084–1127 (2008)
5. Czapliński, A., Michałek, M., Seynnaeve, T.: Uniform matrix product states from an algebraic geometer's point of view. [arXiv:1904.07563](https://arxiv.org/abs/1904.07563) (2019)
6. Domanov, I., De Lathauwer, L.: Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition. *SIAM J. Matrix Anal. Appl.* **35**, 636–660 (2014)
7. Domanov, I., De Lathauwer, L.: Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL. *SIAM J. Matrix Anal. Appl.* **36**, 1567–1589 (2015)
8. Domanov, I., De Lathauwer, L.: Canonical polyadic decomposition of third-order tensors: Relaxed uniqueness conditions and algebraic algorithm. *Linear Algebra Appl.* **513**, 342–375 (2017)
9. Hackbusch, W.: *Tensor Spaces and Numerical Tensor Calculus*. Springer, Berlin (2012). 2nd edn. appears in 2020
10. Kruskal, J.B.: Rank, decomposition, and uniqueness for 3-way and N -way arrays. In: Coppi, Bolasco (eds.) *Multway Data Analysis*, pp. 7–18. North-Holland, Amsterdam (1989)
11. Landsberg, J.M.: *Tensors: Geometry and Applications*. AMS, Providence (2012)
12. Perez-García, D., Verstraete, F., Wolf, M.M., Cirac, J.I.: Matrix product state representations. *Quantum Inf. Comput.* **7**, 401–430 (2007)
13. Qi, Y., Michałek, M., Lim, L.H.: Complex best r -term approximations almost always exist in finite dimensions. *Appl. Comput. Harmon. Anal.* Available online (2019)
14. Sørensen, M., De Lathauwer, L.: Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(l_{r,n}, L_{r,n}, 1)$ terms—part I: Uniqueness. *SIAM J. Matrix Anal. Appl.* **36**, 496–522 (2015)
15. Sørensen, M., De Lathauwer, L., Comon, P., Icart, S., Deneire, L.: Canonical polyadic decomposition with a columnwise orthonormal factor matrix. *SIAM J. Matrix Anal. Appl.* **33**, 1190–1213 (2012)
16. Sørensen, M., Domanov, I., De Lathauwer, L.: Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(l_{r,n}, L_{r,n}, 1)$ terms—part II: Algorithms. *SIAM J. Matrix Anal. Appl.* **36**, 1015–1045 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.