Three decompositions of symmetric tensors have similar condition numbers

Nick Dewaele
KU Leuven, Department of Computer Science, Leuven, Belgium.

Paul Breiding
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany.

Nick Vannieuwenhoven
KU Leuven, Department of Computer Science, Leuven, Belgium; Leuven.AI; KU Leuven Institute for AI, B-3000 Leuven, Belgium.

Abstract
We relate the condition numbers of computing three decompositions of symmetric tensors: the canonical polyadic decomposition, the Waring decomposition, and a Tucker-compressed Waring decomposition. Based on this relation we can speed up the computation of these condition numbers by orders of magnitude.

Keywords: condition number, canonical polyadic decomposition, Waring decomposition

1. Introduction
Many problems in machine learning and signal processing involve computing a decomposition of a symmetric tensor [1, 2]; an order-$D$ tensor $\mathcal{A} = [a_{i_1 \ldots i_D}]_{i_1 \ldots i_D = 1}^{m \times \times n}$ is symmetric if its entries $a_{i_1 \ldots i_D}$ are invariant under all permutations of the indices $i_1, \ldots, i_D$. We establish a close connection between the numerical sensitivity of the following three increasingly structured decomposition problems associated with a symmetric tensor $\mathcal{A}$:

1. A canonical polyadic decomposition (CPD) of $\mathcal{A}$ expresses $\mathcal{A}$ as a sum of $R$ (not necessarily symmetric) tensors of rank 1, where $R$ is minimal. In other words, $\mathcal{A} = \sum_{r=1}^{R} \mathcal{A}_r$ where $\mathcal{A}_r = \alpha_r \mathbf{a}_r^{(1)} \otimes \cdots \otimes \mathbf{a}_r^{(D)}$, $\alpha_r \in \mathbb{R} \setminus \{0\}$ and each $\mathbf{a}_r^{(i)}$ is a point on the sphere $\mathbb{S}^{n-1} = \{ \mathbf{a} \in \mathbb{R}^n \mid \|\mathbf{a}\|_2 = 1 \}$.

2. A Waring decomposition (WD) is a special case of the CPD where all summands are symmetric. That is, for $r = 1, \ldots, R$, we have that $\mathcal{A}_r = \alpha_r \mathbf{a}_r^{\otimes D}$ where $\alpha_r \in \mathbb{R} \setminus \{0\}$, $\mathbf{a}_r \in \mathbb{S}^{n-1}$, and $\mathbf{a}_r^{\otimes D}$ is the tensor product of $D$ copies of $\mathbf{a}_r$.

3. A $Q$-compressed Waring decomposition ($Q$-WD) is defined as follows. A symmetric tensor $\mathcal{A}$ can be represented in a minimal subspace by a symmetric Tucker decomposition [3, 4], i.e., $\mathcal{A} = (Q, \ldots, Q) \cdot \mathbf{g}$ where $Q \in \mathbb{R}^{n \times m}$ has orthonormal columns and $\mathbf{g} \in \mathbb{R}^{m \times \cdots \times m}$ is symmetric with $m < n$. We write this as $\mathcal{A} = Q^{\otimes D} \mathbf{g}$. If $\mathbf{g}$ has a WD $\mathbf{g} = \sum_{r=1}^{R} \mathbf{g}_r$, then it can be converted to a WD $\mathcal{A} = \sum_{r=1}^{R} Q^{\otimes D} \mathbf{g}_r$.

In all three cases, the summands are points on a smooth manifold $\mathcal{M} \subseteq \mathbb{R}^{n \times \cdots \times n}$, so they are join decompositions [5]. For the CPD, the summands lie on the Segre manifold $\mathcal{S}_{n,D}$, for the WD they lie on the Veronese manifold $\mathcal{V}_{n,D}$ [6], and for the $Q$-WD they lie on the manifold $\mathcal{W}_{Q,D} = Q^{\otimes D}(\mathcal{V}_{m,D})$. In the remainder, we drop the subscripts on the manifolds if they are clear from the context.

We study the sensitivity of the summands in these three decompositions with respect to perturbations of $\mathcal{A}$. Consider a decomposition of $\mathcal{A}$ with summands $\mathbf{a} = (\mathbf{A}_1, \ldots, \mathbf{A}_R) \in \mathcal{M}^{\times R}$, where $\mathcal{M}$ is one of the three manifolds described above and $\mathcal{M}^{\times R}$ is the product of $R$ copies of $\mathcal{M}$. Under mild conditions [5], $\mathbf{a}$ is an
isolated decomposition of $\mathcal{A}$ and the addition map $\mathcal{M}^\times \mathcal{R} \to \mathbb{R}^{n \times \cdots \times n}$, $(\mathcal{A}_1, \ldots, \mathcal{A}_R) \mapsto \mathcal{A}_1 + \cdots + \mathcal{A}_R$ admits a local inverse $\Sigma^{-1}_a$. In this case, the sensitivity of the decomposition with respect to $\mathcal{A}$ can be measured by the condition number [7]:

$$\kappa_{\mathcal{M}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) := \lim_{\delta \to 0} \sup_{\mathcal{B}_\delta(S(M^{\times n})), \|\mathcal{B}_\delta - \mathcal{A}\| \leq \delta} \frac{\|\Sigma^{-1}_a(\mathcal{A}) - \Sigma^{-1}_a(\mathcal{B})\|}{\|\mathcal{B} - \mathcal{A}\|},$$

(1)

where $\|\|$ is the Euclidean or Frobenius norm. If $a$ is not isolated, $\kappa_{\mathcal{M}}(\mathcal{A}_1, \ldots, \mathcal{A}_R)$ := $\infty$.

Suppose $\mathcal{A}$ has a Q-WD $a = (\mathcal{A}_1, \ldots, \mathcal{A}_R)$. It can also be regarded as a WD or CPD by ignoring symmetry or subspace constraints. We investigate the relationship between the condition numbers of these three problems at $a$. Since $\mathcal{W} \subseteq \mathcal{V} \subseteq \mathcal{S}$, it follows from (1) that $\kappa_{\mathcal{V}}(a) \leq \kappa_{\mathcal{A}}(a) \leq \kappa_{\mathcal{S}}(a)$. Similarly to recent findings [8], we show the following results for the WD:

**Theorem 1.1.** Let $\mathcal{G} = \mathcal{G}_1 + \cdots + \mathcal{G}_R \in \mathbb{R}^{n \times \cdots \times m}$ be a WD of an order-D tensor.

1. Take $Q \in \mathbb{R}^{n \times m}$ with orthonormal columns and set $\mathcal{A}_r = Q^\otimes D \mathcal{G}_r$, for $r = 1, \ldots, R$. Then

$$\kappa_{\mathcal{V/Q},D}(\mathcal{A}_1, \ldots, \mathcal{A}_R) \leq \kappa_{\mathcal{V},D}(\mathcal{A}_1, \ldots, \mathcal{A}_R) \leq \sqrt{D} \cdot \kappa_{\mathcal{V},D}(\mathcal{G}_1, \ldots, \mathcal{G}_R) = \sqrt{D} \cdot \kappa_{\mathcal{V/Q},D}(\mathcal{A}_1, \ldots, \mathcal{A}_R).$$

2. Let $U \in \mathbb{R}^{n \times m}$ have orthonormal columns and set $\mathcal{A}_r := U^\otimes D \mathcal{G}_r$ for all $r$. If $\min(\ell, n) > m$, then

$$\kappa_{\mathcal{V},D}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \kappa_{\mathcal{V},D}(\mathcal{B}_1, \ldots, \mathcal{B}_R); \text{ i.e., the condition number is invariant under non-minimal symmetric Tucker compressions.}$$

Numerical evidence indicates a stronger connection, which can be proved in the rank-2 case:

**Conjecture 1.2.** If $\mathcal{A} = \sum_{r=1}^R \mathcal{A}_r$ is a WD of an order-D symmetric tensor $A \in \mathbb{R}^{n \times \cdots \times n}$ with $D \geq 3$, then $\kappa_{\mathcal{V}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \kappa_{\mathcal{S}}(\mathcal{A}_1, \ldots, \mathcal{A}_R)$.

**Proposition 1.3.** Conjecture 1.2 holds for $R \leq 2$.

In conjunction with [8, Theorem 5.1], Conjecture 1.2 would imply that $\kappa_{\mathcal{V}}(a) = \kappa_{\mathcal{A}}(a) = \kappa_{\mathcal{S}}(a)$ for any Q-WD $a$, which is sharper than Theorem 1.1. This entails that the supremum in (1) applied to the Segre manifold (i.e., $M = \mathcal{S}$) can be attained locally with a perturbation $\mathcal{A} \in \mathcal{S}(\mathcal{W}^\times \mathcal{R})$.

A practical consequence of Theorem 1.1 relates to the following procedure from [5] to compute the condition number. Let $\mathcal{M}$ be either $\mathcal{S}$ or $\mathcal{V}$. Let the matrix $T^M_{\mathcal{A}_1, \ldots, \mathcal{A}_R}$ consist of columns of the orthonormal basis of $T_{\mathcal{A}_r, \mathcal{M}}$ for $r = 1, \ldots, R$. Then the condition number is characterized by the Terracini matrix $T^M_{\mathcal{A}_1, \ldots, \mathcal{A}_R}$:

$$T^M_{\mathcal{A}_1, \ldots, \mathcal{A}_R} := \begin{bmatrix} T^M_{\mathcal{A}_1} & \cdots & T^M_{\mathcal{A}_R} \end{bmatrix} \quad \text{and} \quad \kappa_{\mathcal{M}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \sigma_{\min}(T^M_{\mathcal{A}_1, \ldots, \mathcal{A}_R})^{-1},$$

(2)

where $\sigma_{\min}(A)$ is the smallest singular value of $A$. Consider a WD $\mathcal{A} = \sum_{r=1}^R \mathcal{A}_r$ with $\mathcal{A}_r = \alpha_r \mathcal{A}_r^\otimes D$ for some $\alpha_r \in \mathbb{R} \setminus \{0\}$ and $\mathcal{A}_r \in \mathcal{S}^{n-1}$. For this decomposition the Terracini matrices for the CPD and the WD, respectively, are given as follows: for any two matrices $X$ and $A$, let $X \otimes_d A^{\otimes D-d} := A^{\otimes d-1} \otimes X \otimes A^{\otimes D-d}$. Let $U(\mathcal{A}_r)$ be any orthonormal basis of $T_{\mathcal{A}_r, \mathcal{S}^{n-1}} = a_r^\perp$. Then the Terracini matrices are

$$T^S_{\mathcal{A}_1, \ldots, \mathcal{A}_R} = \begin{bmatrix} a_r^\otimes D \cr U_r \otimes_d a_r^\otimes D^{-1} \end{bmatrix}^D_{d=1} \quad \text{and} \quad T^V_{\mathcal{A}_1, \ldots, \mathcal{A}_R} = \begin{bmatrix} a_r^\otimes D \cr \frac{1}{\sqrt{D}} \sum_{d=1}^D U(\mathcal{A}_r) \otimes_d a_r^\otimes D^{-1} \end{bmatrix}^R_{r=1}.$$  

(3)

A major implication of Theorem 1.1 is that we can speed up the computation of $\kappa_{\mathcal{V}/D}(\mathcal{A}_1, \ldots, \mathcal{A}_R)$. Assuming $n > R$ and $\mathcal{A}_r = \alpha_r \mathcal{A}_r^\otimes D$ with $\alpha_r \neq 0$ and $\mathcal{A}_r \in \mathcal{S}^{n-1}$, the following computes $\kappa_{\mathcal{V}/D}(\mathcal{A}_1, \ldots, \mathcal{A}_R)$:

1. Compute a thin singular value decomposition $[\alpha_r]_{r=1}^R = Q^S \Sigma^V T$ and set $[\mathcal{G}]_{r=1}^R := \Sigma^V T \in \mathbb{R}^{n \times R}$.
2. Construct $b_r = [g_r^T \ 0]^T \in \mathbb{R}^\ell$ where $\ell = m + 1$ and set $\mathcal{B}_r = \alpha_r b_r^\otimes D$ for each $r$.
3. Construct $T^V_{\mathcal{A}_1, \ldots, \mathcal{A}_R}$ as in (3) and compute $\kappa_{\mathcal{V}}(\mathcal{B}_1, \ldots, \mathcal{B}_R)$ by applying (2).

Steps 1-2 give one possible choice of $Q$ and $U = [I \ 0]^T$ and $\mathcal{G}_r = \alpha_r \mathcal{G}_r^\otimes D$ that satisfy Theorem 1.1. A Julia [9] implementation of this method is provided along with the arXiv version of this manuscript.

Since $T^V_{\mathcal{A}_1, \ldots, \mathcal{A}_R} \in \mathbb{R}^{R \times D \times R}$ and $\ell \leq R + 1$, step 3 can be performed in $O(R^{D+4})$ operations, adding to the $O(n R^2)$ cost of step 1. Applying (2) to $T^V_{\mathcal{A}_1, \ldots, \mathcal{A}_R} \in \mathbb{R}^{n \times R \times R}$ would involve $O((n \times R)^2)$ operations. The algorithm can reach significant speedups if $n \gg R$. For instance, we applied the Julia code to a WD with $(n, D, R, \ell) = (100, 3, 10, 11)$ on an Intel Xeon CPU E5-2697 v3 running on 8 cores and 126GB memory. The computation times were 115.6 and 0.0092 seconds for the original and improved algorithms, respectively.
2. Condition number of a Q-WD

In this section, we prove Theorem 1.1 based on the following insight: $\Sigma(\mathcal{V}^n)$ is locally a manifold whose tangent space is decomposed as $\mathbb{T} \oplus \mathbb{T}^\perp$ where $\mathbb{T}$ is the tangent space to $\Sigma(\mathcal{V}^n)$ and $\mathbb{T}^\perp$ is its orthogonal complement. As long as $n > m$, the effect of the worst perturbation to $\mathbb{A}$ inside $\mathbb{T}^\perp$ is independent of $n$ and can be bounded as in the first statement. From this, the second statement follows as well.

**Proof of Theorem 1.1.** The first inequality follows from the inclusion $W_{Q,D} \subseteq \mathbb{V}_{n,D}$. The last follows from the fact that $Q^{\otimes D}$ is an isometry between $\mathbb{V}_{m,D}$ and $W_{Q,D}$. It remains to show the middle inequality. If $m \neq n$, $Q$ is an orthogonal change of basis, so we assume $m > n$.

For each $r$, let $\tilde{g}_r = \alpha_r g_r^{\otimes D}$ with $\alpha_r \in \mathbb{R} \setminus \{0\}$ and $g_r \in S^{m-1}$, let $\alpha_r = Q g_r$ and define $U_r$ so that the matrix $[g_r \ U_r] \in \mathbb{R}^{m \times m}$ is orthogonal. Construct $T_{g_r}$ by applying (3) to $g_r$. Complete $Q$ to an orthonormal basis $[Q_1 \ Q_\perp]$ of $\mathbb{R}^n$. The columns of $U(\alpha_r) := [Q U_r \ Q_\perp]$ form an orthonormal basis of $T_{n,m}$. Substituting this into (3) gives

$$T_{\mathbb{V}_{n,m}} = [T_r \ T_r^\perp]_{r=1}^R$$

with $T_r = [\alpha_r^{\otimes D} \sqrt{D} \left( \sum_{d=1}^D Q U_r \otimes_d a_r^{\otimes D-1} \right)]$ and $T_r^\perp = \frac{1}{\sqrt{D}} \left( \sum_{d=1}^D Q_\perp \otimes_d a_r^{\otimes D-1} \right)$.

Since $\alpha_r = Q g_r$, we have $T_r = Q^{\otimes D} T_{\tilde{g}_r}^{\otimes D}$. Thus, up to a column permutation, $T_{\mathbb{V}_{n,m}}$ is the horizontal concatenation of $Q^{\otimes D} T_{\tilde{g}_r^{\otimes D}}$ and $T_\perp := [T_{\tilde{r}_1} \ \cdots \ T_{\tilde{r}_R}]$. The column spaces of $Q^{\otimes D}$ and $T_\perp$ are orthogonal, so that the singular values of $T_{\mathbb{V}_{n,m}}$ are the union of those of $Q^{\otimes D} T_{\tilde{g}_r^{\otimes D}}$ and $T_\perp$ separately. Since $Q$ has orthonormal columns, $Q^{\otimes D} T_{\tilde{g}_r^{\otimes D}}$ has the same singular values as $T_{g_r}$, so it suffices to show $\sigma_{\min}(T_\perp) \geq \sigma_{\min}(T_{\tilde{g}_r^{\otimes D}}) / \sqrt{D}$.

To do this, we compute $(T_\perp)^T T_\perp = [(T_{\tilde{r}_1})^T (T_{\tilde{r}_2})^T]_{r_1,r_2=1}^R$, where the block at $(r_1, r_2)$ is

$$\frac{1}{D} \left( \sum_{d=1}^D Q_\perp \otimes_d a_{r_1}^{\otimes D-1} \right) \left( \sum_{d=1}^D Q \otimes_d a_{r_2}^{\otimes D-1} \right) = (a_{r_1}, a_{r_2})^{D-1} I_{n-m} = (g_{r_1}, g_{r_2})^{D-1} I_{n-m}. \quad (4)$$

Consider two modifications of $T_\perp$ that preserve the singular values: first, let $\tilde{T}_\perp := [I_{n-m} \otimes a_r^{\otimes D-1}]_{r=1}^R$, then by (4), we have $(T_\perp)^T \tilde{T}_\perp = (T_\perp)^T (T_\perp)$, so they have the same singular values. Second, if we define $\check{T}_\perp := [g_r \ U_r] \otimes [g_r^{\otimes D-1}]_{r=1}^R$, then $\tilde{T}_\perp$ and $\check{T}_\perp$ also have the same singular values, since $[g_r \ U_r]$ and $I_{n-m}$ are orthogonal up to multiplicities [8, lemma 5.3]. Hence, we can proceed with $\check{T}_\perp$ instead of $T_\perp$. Similarly, we modify $T_{\tilde{g}_r^{\otimes D}}$ by scaling up all its columns of the form $g_r^{\otimes D}$ by $\sqrt{D}$ gives

$$\check{T}_{\tilde{g}_r^{\otimes D}} := \sqrt{D} g_r^{\otimes D} \left[ \frac{1}{\sqrt{D}} \sum_{d=1}^D U_r \otimes_d a_r^{\otimes D-1} \right]_{r=1}^R = \left[ \frac{1}{\sqrt{D}} \sum_{d=1}^D [g_r \ U_r] \otimes_d a_r^{\otimes D-1} \right]_{r=1}^R,$$

i.e., $\check{T}_{\tilde{g}_r^{\otimes D}} = T_{\tilde{g}_r^{\otimes D}} \Delta$ where $\Delta$ is diagonal and $\sigma_{\min}(\Delta) = 1$. Hence, $\sigma_{\min}(\check{T}_{\tilde{g}_r^{\otimes D}}) \geq \sigma_{\min}(T_{\tilde{g}_r^{\otimes D}})$.

To compare the singular values of $\tilde{T}_\perp$ and $\check{T}_{\tilde{g}_r^{\otimes D}}$, take the singular vector $v = [v_r \in \mathbb{R}^m]_{r=1}^R$ of $\check{T}_\perp$ corresponding to the smallest singular value and compute

$$\check{T}_{\tilde{g}_r^{\otimes D}} v_r = \frac{1}{\sqrt{D}} \left( \sum_{d=1}^D [g_r \ U_r] \otimes_d a_r^{\otimes D-1} \right) v_r = \frac{1}{\sqrt{D}} \left( \sum_{d=1}^D \sum_{r=1}^R [g_r \ U_r] v_r \right) \otimes_d a_r^{\otimes D-1}.$$

Since all the summands in the outer sum have the same norm, the triangle inequality gives

$$\left\| \check{T}_{\tilde{g}_r^{\otimes D}} v_r \right\| \leq \sqrt{D} \left( \sum_{r=1}^R \left\| [g_r \ U_r] v_r \otimes a_r^{\otimes D-1} \right\| \right) = \sqrt{D} \left\| [g_r \ U_r] \otimes a_r^{\otimes D-1} \right\|_r = \sqrt{D} \cdot \sigma_{\min}(\check{T}_\perp).$$

3
As \( \sigma_{\min}(T_{g_1 \ldots g_n}^V) \leq \sigma_{\min}(T_{g_1 \ldots g_n}^{\tilde{V}}) \leq \|T_{g_1 \ldots g_n}^{\tilde{V}} \| \), and \( \sigma_{\min}(T_{g_1 \ldots g_n}^{\perp}) = \sigma_{\min}(T_{g_1 \ldots g_n}) \) this gives the desired bound.

For the second statement, recall that the singular values of \( T_{A_1 \ldots A_n}^{V,D} \) are the union of those of \( T_{g_1 \ldots g_n}^{V,m,D} \) and those of \( \tilde{T}_{A_1 \ldots A_n}^{\perp} \) whenever \( n > m \). Observe that both of these matrices are independent of \( n \) and \( Q \). Hence, applying the above calculation to \( B_1, \ldots, B_n \in V, D \subset \mathbb{R}^{t \times m} \) and orthogonal \( U \in \mathbb{R}^{t \times n} \) under the assumption \( \ell > m \) would reveal the same singular values. \( \square \)

3. Equivalence between the CPD and WD

Conjecture 1.2 is a stronger statement than Theorem 1.1, but it seems too challenging to show in general. We present a proof for the case where \( R = 2 \) and present numerical evidence for the general case.

**Proof of Proposition 1.3.** For \( R = 1 \), both condition numbers are equal to 1 by (2) and (3). For \( R = 2 \), the proof comprises computing the singular values of (2) for the CPD. Let \( A_1 = \lambda_1 u \otimes D \) and \( A_2 = \lambda_2 v \otimes D \) with \( u, v \in S^{n-1} \) and \( \lambda_1, \lambda_2 \neq 0 \). Let \( U, V \in \mathbb{R}^{t \times (n-1)} \) be orthonormal bases of \( T_{u}^{S^{n-1}} = u \perp \) and \( T_{v}^{S^{n-1}} = v \perp \), respectively. Applying (3) and using as before the notation \( Q \) and those of \( \tilde{T} \), Observe that both of these matrices are independent of \( n \) and \( Q \). Together, these terms add up to

\[
\min_{\text{sing values}} \lambda_{\min} \left( T_{A_1 A_2} \right) \leq \| T_{A_1 A_2} \| = \| T_{A_1 A_2} \|_{\text{F}}.
\]

Observe that both of these matrices are independent of \( n \) and \( Q \). Together, these terms add up to

\[
\min_{\text{sing values}} \lambda_{\min} \left( T_{A_1 A_2} \right) \leq \| T_{A_1 A_2} \| = \| T_{A_1 A_2} \|_{\text{F}}.
\]

Observe that both of these matrices are independent of \( n \) and \( Q \). Together, these terms add up to

\[
\min_{\text{sing values}} \lambda_{\min} \left( T_{A_1 A_2} \right) \leq \| T_{A_1 A_2} \| = \| T_{A_1 A_2} \|_{\text{F}}.
\]

Observe that both of these matrices are independent of \( n \) and \( Q \). Together, these terms add up to

\[
\min_{\text{sing values}} \lambda_{\min} \left( T_{A_1 A_2} \right) \leq \| T_{A_1 A_2} \| = \| T_{A_1 A_2} \|_{\text{F}}.
\]

Observe that both of these matrices are independent of \( n \) and \( Q \). Together, these terms add up to

\[
\min_{\text{sing values}} \lambda_{\min} \left( T_{A_1 A_2} \right) \leq \| T_{A_1 A_2} \| = \| T_{A_1 A_2} \|_{\text{F}}.
\]
where each $\beta_{\neq} \alpha$ add up to $-j(j-1)\beta_{\neq} - j'\beta_{\neq}$. This means the terms involving $\beta_{\neq}$ also vanish. Therefore, all inner products $(S_{a,\perp}^j)^T(S_{v,\perp}^j)$ vanish for $j \neq j'$.

Furthermore, the columns of $T_{3,\perp}^j$ are symmetric tensors. The space of symmetric tensors is the linear span of the Veronese manifold $V := \{\alpha x^D \mid \alpha \in \mathbb{R} \setminus \{0\}, x \in S^{n-1}\}$. Since $\sum_{i=1}^{j+1}(q_{D^i})_i = 0$, we have $(x^D)^TS_{a,\perp}^j = \sum_{i=1}^{j+1}(x^T u)^D_{i-1}x^TU(q_{D^i})_i = 0$, so that the columns of $S_{a,\perp}^j$ and $T_{3,\perp}^j$ are pairwise orthogonal. We can therefore conclude that $(5)$ partitions $T^S$ into pairwise orthogonal blocks.

Next, we compute all singular values of $T^S$ by computing the singular values of the blocks in $(5)$ separately. Using the same notation as before, we compute the blocks of $(T_1^T T_1^j)^T$:

$$(S_{a,\perp}^j)^T(S_{v,\perp}^j) = \frac{1}{j(j+1)}(x_1 + \cdots + x_j - jx_{j+1})^T(y_1 + \cdots + y_j - jy_{j+1}) = \frac{1}{j(j+1)}(a + b + c + d)$$

where

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} (x_1 + \cdots + x_j)(y_1 + \cdots + y_j) \\ -(x_1 + \cdots + x_j)jy_{j+1} \\ -jx_{j+1}(y_1 + \cdots + y_j) \\ j^2x_{j+1}y_{j+1} \end{bmatrix} = \begin{bmatrix} j\beta_{\neq} + (j^2 - j)\beta_{\neq} \\ -j^2\beta_{\neq} \\ -j^2\beta_{\neq} \\ j^2\beta_{\neq} \end{bmatrix}.$$}

This gives $(S_{a,\perp}^j)^T(S_{v,\perp}^j) = \beta = \beta_{\neq} = \alpha D^{-1}U^TV - \alpha D^{-2}U^TV$. Hence, the Gramian of $T_1^j$ is

$$G_\perp := (T_1^j)^TT_1^j = \begin{bmatrix} I_{n-1} & \alpha D^{-1}U^TV \\ \alpha D^{-1}U^TV & I_{n-1} \end{bmatrix},$$

which is independent of $j$. The Gramian of $T_{3,\perp}^j$ and $v^\perp$ respectively. By planar geometry, we can choose these bases such that $Ue_j = \frac{-\alpha e_j}{\parallel \alpha e_j \parallel}$, $Ve_j = \frac{-\alpha e_j}{\parallel \alpha e_j \parallel}$ and $Ue_j = Ve_j$ for all $j = 2, \ldots, n-1$. Consequently, $U^TV = \sqrt{1 - \alpha^2}e_1$, $V^TU = \sqrt{1 - \alpha^2}e_1$, and $U^TV = \det(-\alpha, 1, \ldots, 1)$. Plugging these into $G_\perp$, we get

$$G_\perp = \begin{bmatrix} I_{n-1} & \det(-\alpha D - \alpha D^{-2}(1 - \alpha^2), \alpha D^{-1}, \ldots, \alpha D^{-1}) \end{bmatrix} = I_{2(n-1)} + \begin{bmatrix} 0 & A_\perp \\ A_\perp & 0 \end{bmatrix},$$

where $A_\perp := \det(-\alpha D - \alpha D^{-2}(1 - \alpha^2), \alpha D^{-1}, \ldots, \alpha D^{-1})$. Recall that the eigenvalues of $[\begin{smallmatrix} 0 & A \end{smallmatrix}]$ are $\pm \sigma(A)$, where $\sigma$ are the singular values of $A$. Therefore, the eigenvalues of $G_\perp$ are $\lambda(G_\perp) = \{1 \pm \alpha D^{-1}, 1 \pm \alpha D^{-2}\}$. We only need the extreme eigenvalues, which are $1 \pm \alpha D^{-2}$ since $|\alpha| \leq 1$. For $G_S$, we obtain

$$G_S = \begin{bmatrix} 1 & \alpha D \\ \times & \times \\ \times & \times \sqrt{\alpha D^{-1} \sqrt{1 - \alpha^2}} \\ \times & \times & \times & 1 \\ \times & \times & \times & \times \end{bmatrix},$$

where $A_S = \det(-\alpha D + (D - 1)\alpha D^{-2}(1 - \alpha^2), \alpha D^{-1}, \ldots, \alpha D^{-1})$. Define the two matrices

$$Z = \begin{bmatrix} \alpha D \\ \sqrt{\alpha D^{-1} \sqrt{1 - \alpha^2}} \end{bmatrix}, \quad Z' := \begin{bmatrix} \alpha D \\ \sqrt{\alpha D^{-1} \sqrt{1 - \alpha^2}} \end{bmatrix}.$$
The eigenvalues of $G_S$ are $1 \pm \sigma(Z)$. Due to the sparse structure of $Z$, its singular values are $a^{D-1}$ and the singular values of $Z'$. Since $Z'$ is symmetric, its eigenvalues and singular values coincide. We factor out $a^{D-2}$ and compute the eigenvalues in terms of the trace $\tau$ and determinant $\Delta$. This gives $\tau = (D - 1)(1 - \alpha^2)$, $\Delta = -\alpha^2$, and $\lambda_1(Z') = \frac{a^{D-2}}{2} (\tau + \sqrt{\tau^2 - 4\Delta})$ and $\lambda_2(Z') = \frac{a^{D-2}}{2} (\tau - \sqrt{\tau^2 - 4\Delta})$. Finally, we compare the eigenvalues of $G_S$ to the extreme eigenvalues of $G_{1\perp}$. Since $\alpha^2 \leq 1$ and $D \geq 3$, 

\[
4\tau \geq 4(1 + \Delta) \Rightarrow \tau^2 - 4\Delta \geq \tau^2 - 4\tau + 4 \Rightarrow \sqrt{\tau^2 - 4\Delta} \geq 2 - \tau \Rightarrow \frac{1}{2}(\tau + \sqrt{\tau^2 - \Delta}) \geq 1.
\]

Hence, $G_S$ has at least one eigenvalue less than or equal to the smallest of $G_{1\perp}$, namely $1 + \lambda_1(Z') \leq 1 + a^{D-2}$ if $a^{D-2}$ is negative, and $1 - \lambda_1(Z') \leq 1 - a^{D-2}$ otherwise. This shows that the smallest singular value of $\tilde{T}^S$ in (5) is a singular value of $T_{\alpha,\tau,n}^{\perp}$, as required. 

3.1. Numerical experiments

We tested Conjecture 1.2 for third order tensors. For $n = 3 \ldots 18$, we generated 500 random symmetric rank $R$ decompositions $\sum_{r=1}^{R} a_r^{n,3}$ where $a_r \sim \mathcal{N}(0, I_n)$ using Julia v1.6 [9]. For each decomposition, we computed the two condition numbers. By dimensionality arguments, the condition number can only be finite if $Rn < \binom{n+2}{2}$, where the right-hand side is the dimension of the space of symmetric $n \times n \times n$ tensors [6]. We tested all values of $R$ below this upper bound.

Figure 1 shows the ratio between the condition number of the CPD and the WD. A priori, it can never be less than 1. In practice, numerical computations would sometimes find a ratio of $1 - 10^{-11}$ or less. This suggests that ratios exceeding 1 by less than $10^{-11}$ can be explained by numerical roundoff. All measurements lie below this threshold.

References

[1] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM Review 51 (3) (2009) 455–500.
[2] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, M. Telgarsky, Tensor decompositions for learning latent variable models, Journal of Machine Learning Research 15 (2014) 2773–2832.
[3] L. R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (3) (1966) 279–311.
[4] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, SIAM Journal on Matrix Analysis and Applications 21 (4) (2000) 1253–1278.
[5] P. Breiding, N. Vannieuwenhoven, The condition number of join decompositions, SIAM Journal on Matrix Analysis and Applications 39 (1) (2018) 287–309.
[6] J. M. Landsberg, Tensors: Geometry and Applications, Representation theory 381 (402) (2012) 3.
[7] J. R. Rice, A theory of condition, SIAM Journal on Numerical Analysis 3 (2) (1966) 287–310.
[8] N. Dewaele, P. Breiding, N. Vannieuwenhoven, The condition number of many tensor decompositions is invariant under Tucker compression (2021) 1–18arXiv:2106.13034.
[9] J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, Julia: A Fresh Approach to Numerical Computing, SIAM Review 59 (1) (2017) 65–98.
[10] F. R. Helmert, Die Genaugkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genaugkeit, Astronomische Nachrichten 88 (1876) 113.
Appendix A. The condition number of the partially symmetric decomposition

In this section, we present a generalisation of Theorem 1.1 to the partially symmetric case. We say that a tensor \( \mathcal{A} \) is partially symmetric if it is invariant under the permutation of some (but not all) of its indices. When this symmetry constraint is imposed on the summands in its CPD, the CPD is called a partially symmetric rank decomposition (PSRD). Write the size and degree of the tensors as \( \mathbf{n} = (n_1, \ldots, n_K) \) and \( \mathbf{d} = (d_1, \ldots, d_K) \), respectively. Then partially symmetric tensors of rank 1 form the image of the map

\[
\Phi : \mathbb{R} \setminus \{0\} \times S^{n_1-1} \times \cdots \times S^{n_K-1} \to \mathbb{R}^{n_1 \times \cdots \times n_1 \times \cdots \times n_K}
\]

\[
(a_1, \ldots, a_K) \mapsto \alpha a_1^\otimes d_1 \otimes \cdots \otimes a_K^\otimes d_K.
\]

The image of \( \Phi \) is known as the Segre-Veronese manifold \( SV_{n,d} \) [6]. Analogous to the Q-WD, a \((Q_1, \ldots, Q_K)\)-PSRD is a PSRD of the form \( \mathcal{A} = \sum_{r=1}^{R} (Q_1 \otimes \cdots \otimes Q_K) \mathcal{G}_r \) where each \( Q_r \) has orthonormal columns and \( \mathcal{G}_r \in SV_{m,d} \) where \( m < n \) elementwise. We write \( W = (Q_1 \otimes \cdots \otimes Q_K) \) \((SV_{m,d})\).

To determine the condition number, we apply (2) to the PSRD. The derivative of \( \Phi \) at any point \( \mathcal{A}_r = \Phi(a, a_1, \ldots, a_K) \) is

\[
d\Phi(\hat{a}, \hat{a}_1, \ldots, \hat{a}_K) = \frac{\partial}{\partial a_1} \otimes \cdots \otimes \frac{\partial}{\partial a_K}
\]

\[
= \frac{\alpha}{\sqrt{d_k}} \left( \sum_{d=1}^{d_k} \hat{a}_k^\otimes d-a_k^\otimes d-1 \right) \otimes_k \left( \mathbf{a}_k^\otimes d_{k'} \right). \quad \sum_{k'} \neq k
\]

If \( U(a_k) \) spans an orthonormal basis of \( T_{\mathcal{A}_r} S^{n_k-1} \), the tangent space to \( SV_{m,d} \) is the column space of

\[
T^{SV_{n,d}}_{\mathcal{A}_r} := \left[ \bigotimes_{k=1}^{K} a_k^\otimes d_k \left( \frac{1}{\sqrt{d_k}} \left( \sum_{d=1}^{d_k} U(a_k) \otimes_d a_k^\otimes d-1 \right) \otimes_k \left( \mathbf{a}_k^\otimes d_{k'} \right) \right) \right]^{K}_{k=1}.
\]

Observe that all \( K + 1 \) blocks of this matrix have orthonormal columns and are pairwise orthogonal by construction of \( U_k \). Therefore, the condition number of any PSRD can be computed using (2) where the blocks in the Terracini matrix are as in (A.1). Now we can present a generalisation of Theorem 1.1.

**Theorem Appendix A.1.** Let \( \mathcal{G} = \mathcal{G}_1 + \cdots + \mathcal{G}_R \) be a PSRD with summands in \( SV_{m,d} \). For \( k = 1, \ldots, K \), take \( Q_k \in \mathbb{R}^{n_k \times m_k} \) with orthonormal columns and set \( \mathcal{A}_r := (Q_1^\otimes d_1 \otimes \cdots \otimes Q_K^\otimes d_K) \mathcal{G}_r \). Then

\[
\kappa_{SV_{n,d}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) \leq \max_{1 \leq r \leq R} \kappa_{SV_{m,d}}(\mathcal{G}_1, \ldots, \mathcal{G}_R).
\]

Similarly, for \( k = 1, \ldots, K \), let \( U_k \in \mathbb{R}^{d_k \times m_k} \) have orthonormal columns and set \( \mathcal{B}_r := (U_1^\otimes d_k \otimes \cdots \otimes U_k^\otimes d_k) \mathcal{A}_r \) for \( r = 1, \ldots, R \). If \( \min(k, \kappa_k) > m_k \) for all \( k \), then

\[
\kappa_{SV_{n,d}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \kappa_{SV_{m,d}}(\mathcal{B}_1, \ldots, \mathcal{B}_R).
\]

**Remark Appendix A.2.** The case \( K = 1 \) is exactly Theorem 1.1. The case \( d_1 = \cdots = d_K = 1 \) is a statement about the CPD. In this case the theorem reads \( \kappa_{SV_{n,k}}(\mathcal{A}_1, \ldots, \mathcal{A}_R) = \kappa_{SV_{m,k}}(\mathcal{G}_1, \ldots, \mathcal{G}_R) \), which is a special case of [8, Theorem 5.1].

**Proof.** For each \( r \), let \( \mathcal{G}_r = \alpha_r (g_1^r)^\otimes d_1 \otimes \cdots \otimes (g_K^r)^\otimes d_K \) with \( \alpha_r \neq 0 \) and \( \mathbf{g}_r \in SV_{m_k} \) for all \( k \). Let \( a_k^r = Q_k g_k^r \) and define \( U_k^r \) so that \( \mathbf{g}_r = U_k^r \mathbf{U}_k \in \mathbb{R}^{m_k \times m_k} \) is orthogonal. Construct \( T^{SV_{n,d}}_{\mathcal{A}_r} \) by applying (A.1) to \( \mathcal{G}_r \). Complete each \( Q_k \) to an orthonormal basis \( \mathbf{Q}_k \) \((R_k)\) of \( \mathbb{R}^{n_k} \). If \( n_k = m_k \), \( Q_k^\perp \) is an \( n_k \times 0 \) matrix. The columns of \( U(a_k^r) := [Q_k U_k^r Q_k^\perp] \) form an orthonormal basis of \( T_{a_k^r} S^{n_k-1} \). For each \( r \), these can be substituted into (A.1) applied to \( \mathcal{A}_1, \ldots, \mathcal{A}_R \), respectively. Similarly to the symmetric case, this gives

\[
T^{SV_{n,d}}_{\mathcal{A}_r} = [T_r T_r^\perp \cdots T_r^{K_r^\perp}] \quad \text{where} \quad T_r = \left( \bigotimes_{k=1}^{K} Q_k^\otimes d_k \right) T^{SV_{m,d}}_{\mathcal{G}_r}
\]
\[ T_r^\perp = \frac{1}{\sqrt{d_k}} \left( \sum_{d=1}^{d_k} Q_k^d \otimes_d (a_k^r)^{d_k-1} \right) \otimes_k \left( \otimes_{k' \neq k} \left( (a_k^r)^{d_k} \right) \right) \]

for each \( r \) and \( k \). Define \( T = [T]_{r=1}^R \) and \( T_k^\perp = [T_k^\perp]_{r=1}^R \). Observe that these \( K + 1 \) matrices are pairwise orthogonal since \( Q_k^T Q_k = 0 \) and \( a_k^r \in \text{span} Q_k \). Furthermore, note that \( T_{\text{SV},n,a} = [T \ T^\perp \ ... \ T_K^\perp] \) up to a column permutation. Finally, \( T = \left( \bigotimes_{k=1}^K Q_k \right) T_{\text{SV},m,a} \) has the same \( T_{\text{SV},m,a} \) by orthogonality. The combination of these three observations implies that the singular values of \( T_{\text{SV},n,a} \) are the union of the singular values of \( T, T^\perp, \ldots, T_K^\perp \) separately. Consequently, it suffices to show that for each \( k \) we have \( \sigma_{\min}(T_k^\perp) \geq \sigma_{\min}(T_{\text{SV},m,a})/\sqrt{d_k} \). To do this, we compute the Gramian \( (T_k^\perp)^* T_k^\perp \). Define the following auxiliary matrices:

\[ S^k_r = \frac{1}{\sqrt{d_k}} \left( \sum_{d=1}^{d_k} Q_k^d \otimes_d (a_k^r)^{d_k-1} \right) \]
\[ A^k_r = \otimes_{k' \neq k} \left( (a_k^r)^{d_k} \right) \]
\[ G^k_r = \otimes_{k' \neq k} \left( (g_k^r)^{d_k} \right) \]

This allows us to write \( T_k^\perp = S^k_r \otimes_k A^k_r \). For general \( r_1, r_2 \), the inner products between the columns of \( S^k_{r_1} \) and \( S^k_{r_2} \) are

\[ (S^k_{r_1})^* S^k_{r_2} = (a_k^{r_1}, a_k^{r_2})^{d_k-1} I_{n_k-m_k} = (g_k^{r_1}, g_k^{r_2})^{d_k-1} I_{n_k-m_k}. \]

Hence, if we replace the factors \( S^k_r \) in \( T_k^\perp \) by \( I_{n_k-m_k} \otimes (g_k^r)^{d_k-1} \), the Gramian remains unchanged. Similarly, \( (A^k_{r_1})^* A^k_{r_2} = (G^k_{r_1})^* G^k_{r_2} \) for all \( r_1 \) and \( r_2 \), so that we can replace each \( A^k_r \) in \( T_k^\perp \) by \( G^k_r \). Define

\[ \tilde{T}^\perp := [I_{n_k-m_k} \otimes (g_k^r)^{d_k-1} \otimes_k G^k_r]_{r=1}^R \quad \text{and} \quad \tilde{T}^\perp := \left[ [g_k^r \ U_k^r] \otimes (g_k^r)^{d_k-1} \otimes_k G^k_r \right]_{r=1}^R. \]

\( \tilde{T}^\perp \) is \( T^\perp \) with the aforementioned replacements applied. Since \([g_k^r \ U_k^r] \) is orthogonal, the singular values of \( \tilde{T}^\perp \) and \( \tilde{T}^\perp \) are the same up to multiplicities [8, Lemma 5.3]. Hence, for the purpose of comparing singular values, we can proceed with \( \tilde{T}^\perp \) instead of \( T^\perp \).

Next, we also modify \( T_{\text{SV},m,a} \). First, take the following subset of its columns:

\[ T := \left[ \bigotimes_{k=1}^K (g_k^r)^{d_k} \right] \frac{1}{\sqrt{d_k}} \left( \sum_{d=1}^{d_k} U_k^r \otimes_d (g_k^r)^{d_k-1} \right) \otimes_k G^k_r \quad \text{where columns are scaled up by } \sqrt{d_k}. \]

The first column of the \( r \)th block is \( \bigotimes_{k=1}^K (g_k^r)^{d_k} = (g_k^r)^{d_k} \otimes_k G^k_r \). Define \( \tilde{T} \) as a modification of \( T \) where these \( R \) columns are scaled up by \( \sqrt{d_k} \). Rearranging the columns gives

\[ \tilde{T} = \left[ \frac{1}{\sqrt{d_k}} \left( \sum_{d=1}^{d_k} [g_k^r \ U_k^r] \otimes_d (g_k^r)^{d_k-1} \right) \otimes_k G^k_r \right]_{r=1}^R. \]

Since \( T \) is a submatrix of \( T_{\text{SV},m,a} \), we have \( \sigma_{\min}(T_{\text{SV},m,a}) \leq \sigma_{\min}(T^k) \). Because of how we defined \( \tilde{T} \), we also have \( \sigma_{\min}(T^k) \leq \sigma_{\min}(\tilde{T}) \). From here on, we can compare the singular values of \( T^k \) and \( \tilde{T}^\perp \) the same way as their counterparts in the proof of Theorem 1.1. This completes the proof. \( \square \)