IMA Journal of Numerical Analysis (2019) **00**, 1–29
doi:10.1093/imanum/drz026

On the average condition number of tensor rank decompositions

PAUL BREIDING*
Max Planck Institute for Mathematics in the Sciences, Inselstraße 22–26, 04103 Leipzig, Germany
*Corresponding author: breiding@mis.mpg.de

AND

NICK VANNIEUWENHOVEN
KU Leuven, Department of Computer Science, Celestijnenlaan 200A, 3001 Heverlee, Belgium
nick.vannieuwenhoven@kuleuven.be

[Received on 7 June 2018; revised on 3 December 2018]

We compute the expected value of powers of the geometric condition number of random tensor rank decompositions. It is shown in particular that the expected value of the condition number of $n_1 \times n_2 \times 2$ tensors with a random rank-r decomposition, given by factor matrices with independent and identically distributed standard normal entries, is infinite. This entails that it is expected and probable that such a rank-r decomposition is sensitive to perturbations of the tensor. Moreover, it provides concrete further evidence that tensor decomposition can be a challenging problem, also from the numerical point of view. On the other hand, we provide strong theoretical and empirical evidence that tensors of size $n_1 \times n_2 \times n_3$ with all $n_1, n_2, n_3 \geqslant 3$ have a finite average condition number. This suggests that there exists a gap in the expected sensitivity of tensors between those of format $n_1 \times n_2 \times 2$ and other order-3 tensors. To establish these results we show that a natural weighted distance from a tensor rank decomposition to the locus of ill-posed decompositions with an infinite geometric condition number is bounded from below by the inverse of this condition number. That is, we prove one inequality towards a so-called condition number theorem for the tensor rank decomposition.

Keywords: tensor rank decomposition; CPD; condition number; ill-posed problems; inverse distance to ill-posedness; average complexity; condition number theorem.

1. Introduction

An order-d real tensor $\mathcal{A}$ of size $n_1 \times \cdots \times n_d$ is an element of the tensor product of d vector spaces $\mathbb{R}^{n_k}$ for $k = 1, \ldots, d$ (Greub, 1978; Landsberg, 2012). After choosing bases for $\mathbb{R}^{n_k}$, this tensor can be represented as a d-array,

$$\mathcal{A} = \left[a_{i_1,i_2,\ldots,i_d}\right]_{i_1,i_2,\ldots,i_d=1}^{n_1,n_2,\ldots,n_d} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d},$$

which we denote using the same symbol $\mathcal{A}$. We informally refer to this d-array as a tensor. Due to the curse of dimensionality, representing a tensor by this coordinate d-array is neither feasible nor insightful. Fortunately, the tensor often admits additional structure that can be exploited. One particular tensor decomposition is the *tensor rank decomposition*, or *canonical polyadic decomposition* (CPD). It was proposed by Hitchcock (1927) and expresses a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ as a minimum-length linear combination of rank-1 tensors:

$$\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 + \cdots + \mathcal{A}_r, \quad \text{where} \quad \mathcal{A}_i = \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d, \tag{CPD}$$

and where $\otimes$ is the *tensor product*:

$$\mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d = \left[a_{i_1}^{(1)} a_{i_2}^{(2)} \cdots a_{i_d}^{(d)} \right]_{i_1, i_2, \ldots, i_d = 1}^{n_1, n_2, \ldots, n_d} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}, \quad \text{where } \mathbf{a}^k = [a_i^{(k)}]_{i=1}^{n_k}. \tag{1.1}$$

The smallest r for which the expression (CPD) is possible is called the *rank* of $\mathcal{A}$.

One of the main reasons why the CPD has found applications in domains as diverse as psychometrics (Kroonenberg, 2008), chemical sciences (Smilde *et al.*, 2004), theoretical computer science (Bürgisser *et al.*, 1997), signal processing (Comon, 1994; Comon & Jutten, 2010), statistics (McCullagh, 1987; Allman *et al.*, 2009) and machine learning (Anandkumar *et al.*, 2014) is because of its strong uniqueness properties, which allow it to reveal domain-specific information. The expression (CPD) is called unique, or *r-identifiable*, if the set of rank-1 tensors $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$ whose sum is $\mathcal{A}$, is unique, given the tensor $\mathcal{A}$. One of the foundational r-identifiability results for third-order tensors is due to Kruskal (1977).[1]

LEMMA 1.1 (Kruskal's lemma). Let $\mathcal{A} = \sum_{i=1}^r \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i$ be an $n_1 \times n_2 \times n_3$ tensor. Let k_A denote the largest integer such that every subset of $k \leqslant k_A$ vectors from $\{\mathbf{a}_1, \ldots, \mathbf{a}_r\}$ is linearly independent. Let k_B and k_C be defined analogously for the sets $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_r\}$, respectively. If $k_A, k_B, k_C \geqslant 2$ and $r \leqslant \frac{1}{2}(k_A + k_B + k_C - 2)$, then $\mathcal{A}$ is r-identifiable.

It is important to stress the radical departure that r-identifiability of higher-order tensors with $d \geqslant 3$ represents from the case of matrices. Indeed, low-rank matrix decompositions, i.e., (CPD) with $d = 2$ and $n_1, n_2 \geqslant 2$ are *never* r-identifiable for $r \geqslant 2$. This is intuitively clear from the fact that

$$\mathcal{A} = \sum_{i=1}^r \mathbf{a}_i \otimes \mathbf{b}_i = \sum_{i=1}^r \mathbf{a}_i \mathbf{b}_i^{\mathrm{T}} = AB^{\mathrm{T}} = (AX^{-1})(BX^{\mathrm{T}})^{\mathrm{T}},$$

where $A = [\mathbf{a}_i]_i$, $B = [\mathbf{b}_i]_i$ and X is an invertible $r \times r$ matrix. For a general choice[2] of X, the CPDs $\{\mathbf{a}_1 \otimes \mathbf{b}_1, \ldots, \mathbf{a}_r \otimes \mathbf{b}_r\}$ and $\{(AX^{-1})_1 \otimes (BX^{\mathrm{T}})_1, \ldots, (AX^{-1})_r \otimes (BX^{\mathrm{T}})_r\}$ will be distinct, where $(Y)_i$ denotes the ith column of a matrix Y. The precise reason why CPDs of matrices are not r-identifiable follows from a classic result from algebraic geometry; see Harris (1992, Example 12.1).

Recent research by Strassen (1983), Chiantini & Ottaviani (2012), Bocci *et al.* (2014), Chiantini *et al.* (2014) and Domanov & De Lathauwer (2015) has shown that 'most complex rank-r tensors are r-identifiable and hence have a *unique* CPD'. They prove that on an open dense subset of the set of complex rank-r tensors, r-identifiability holds provided that r is sufficiently small; see the aforementioned references for the precise statements. Qi *et al.* (2016) showed that analogous statements

[1] Sidiropoulos & Bro (2000) derived a generalization to order-d tensors, but it is more common to apply Kruskal's lemma to a *reshaped* tensor as this always extends the range of applicability and is usually computationally more attractive; see Chiantini *et al.* (2017) for more details.

[2] Specifically, if an invertible X cannot be factored as ΛP, where P is a permutation matrix and Λ a diagonal matrix.

are valid for real tensors. For a summary of the conjecturally complete picture of these so-called *generic r-identifiability* results, see Chiantini *et al.* (2017, Section 3).

As most rank-r higher-order tensors with $d \geqslant 3$ are known to admit only a finite number of distinct CPDs, usually only 1, the question of the sensitivity of the rank-1 summands in (CPD) naturally enters the picture. Indeed, the tensors originating from applications are almost invariably perturbed by noise and measurement errors. Hence, for unambiguously interpreting the rank-1 tensors $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$, which oftentimes reveal domain-specific information, it is required that they are not too sensitive to perturbations of the tensor $\mathcal{A}$. For measuring the sensitivity of a computational problem to perturbations in the data, a standard technique in numerical analysis is to investigate the *condition number* (Higham, 2002; Bürgisser & Cucker, 2013). The condition number of a function $f : D \to \mathbb{R}^n$ at $x \in D \subset \mathbb{R}^m$ is defined to be

$$\kappa(x) := \lim_{\epsilon \to 0} \sup_{\substack{y \in D, \\ \|x-y\| \leqslant \epsilon}} \frac{\|f(x) - f(y)\|}{\|x - y\|}, \tag{1.2}$$

where the norms are the usual Euclidean norms. Note that if D is a submanifold and f is a differentiable function, then the above reduces to $\kappa(x) = \|\mathrm{d}_x f\|_2$, where $\mathrm{d}_x f : \mathbb{R}^m \to \mathbb{R}^n$ is the derivative of f at x; see Bürgisser & Cucker (2013, Section 14.1).

Earlier theoretical work by the authors introduced two related condition numbers for two distinct computational problems related to computing a CPD from a given tensor. Vannieuwenhoven (2017) investigated the sensitivity of the factor matrices $A^k = [\mathbf{a}_i^k]_{i=1}^r$ that represent a CPD $\{\mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d\}_{i=1}^r$ of an order-d tensor. Factor matrices are not locally unique, making the analysis very technical. An alternative approach was considered by Breiding & Vannieuwenhoven (2018c), based on the geometric framework of a condition due to Rice (1966), Blum *et al.* (1998) and Bürgisser & Cucker (2013). This condition number measures the sensitivity of the set of rank-1 terms that represent a CPD. The foregoing two condition numbers are intimately related to the convergence (and so computational complexity) of certain Gauss–Newton optimization algorithms for approximating a tensor by a rank-r CPD; see Breiding & Vannieuwenhoven (2018a, Section 3.4) for details. In this paper we continue the study of the condition number from Breiding & Vannieuwenhoven (2018c), as it is simpler to analyse and it governs the convergence of the state-of-the-art optimization method of Breiding & Vannieuwenhoven (2018a).

We recall the condition number of the problem of computing a CPD from Breiding & Vannieuwenhoven (2018c). The set of all rank-1 tensors $\mathscr{S}_{n_1,\ldots,n_d}$, which excludes zero, is the cone over a smooth projective variety; see, e.g., Harris (1992, Example 2.11) or Landsberg (2012, Section 4.3). This means that $\mathscr{S}_{n_1,\ldots,n_d}$ is a smooth submanifold of $\mathbb{R}^{n_1 \times \cdots \times n_d}$; its tangent space $\mathrm{T}_{\mathcal{A}}\mathscr{S}_{n_1,\ldots,n_d}$ at $\mathcal{A}$ is the usual one (see Lee, 2013, Chapter 3). By construction, the set of tensors of rank bounded by r is the image of

$$\Phi_r : \mathscr{S}_{n_1,\ldots,n_d}^{\times r} = \mathscr{S}_{n_1,\ldots,n_d} \times \cdots \times \mathscr{S}_{n_1,\ldots,n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}, \quad (\mathcal{A}_1, \ldots, \mathcal{A}_r) \mapsto \mathcal{A}_1 + \cdots + \mathcal{A}_r. \tag{1.3}$$

The problem of computing an (ordered) CPD takes as input a rank-r tensor $\mathcal{A}$ and should compute a CPD; i.e., the preimage Φ_r^{-1}. However, immediately a complication arises: which (ordered) CPD should be computed? Namely, Φ_r^{-1} is not an inverse *function* of Φ_r.[3] This issue is resolved in Breiding

[3] Under the stronger assumption of generic r-identifiability, it is possible to make restrictions such that Φ_r becomes a global diffeomorphism, but the full power of this construction from Beltrán *et al.* (2019) is not required for the purpose of this paper.

& Vannieuwenhoven (2018c) by observing that while a global inverse function Φ_r^{-1} does not exist, there can be local inverse functions. That is, if $\mathcal{A} = \Phi_r(\mathcal{A}_1, \ldots, \mathcal{A}_r)$, there could exist an open neighbourhood $\mathcal{E}$ of $\mathcal{A} \in \mathrm{Im}(\Phi_r)$, an open neighbourhood $\mathcal{F}$ of $(\mathcal{A}_1, \ldots, \mathcal{A}_r) \in \mathscr{S}_{n_1,\ldots,n_d}^{\times r}$ and a local inverse function $\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} : \mathcal{E} \to \mathcal{F}$ such that $\Phi_r \circ \Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} = \mathrm{Id}_{\mathcal{E}}$ and $\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} \circ \Phi_r = \mathrm{Id}_{\mathcal{F}}$. The standard definition (1.2) then applies to $\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1}$. This yields the natural condition number of the ordered CPD $(\mathcal{A}_1, \ldots, \mathcal{A}_r)$:

$$\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) := \lim_{\epsilon \to 0} \sup_{\substack{\mathcal{B} \text{ has rank } r, \\ \|\mathcal{A}-\mathcal{B}\| < \epsilon}} \frac{\|\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1}(\mathcal{A}) - \Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1}(\mathcal{B})\|}{\|\mathcal{A} - \mathcal{B}\|},$$

where the norms are the Euclidean norms induced by the ambient spaces of the domain and image of Φ_r. Note that Φ_r is a smooth map between manifolds. The existence of a local inverse function is readily established by invoking the full power of the inverse function theorem (see, e.g., Lee, 2013, Theorem 4.5); if the derivative $\mathrm{d}_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}\Phi_r$ is an invertible linear map at $(\mathcal{A}_1, \ldots, \mathcal{A}_r) \in \mathscr{S}_{n_1,\ldots,n_d}^{\times r}$, then a local inverse function $\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1}$ as above exists. Let $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r$. Note that $\Phi_r \circ \Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} = \mathrm{Id}_{\mathcal{E}}$ implies that $\mathrm{d}_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}\Phi_r \circ \mathrm{d}_{\mathcal{A}}\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} = \mathrm{Id}$, and likewise for $\mathrm{d}_{\mathcal{A}}\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} \circ \mathrm{d}_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}\Phi_r = \mathrm{Id}$. It follows that $\mathrm{d}_{\mathcal{A}}\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1} = (\mathrm{d}_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}\Phi_r)^{-1}$, so that we obtain the simplification

$$\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) = \|\mathrm{d}_{\mathcal{A}}\Phi_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}^{-1}\|_2 = \|(\mathrm{d}_{(\mathcal{A}_1,\ldots,\mathcal{A}_r)}\Phi_r)^{-1}\|_2, \tag{1.4}$$

which is the spectral characterization in Breiding & Vannieuwenhoven (2018c, Theorem 1.1).

By construction, the above condition number measures the sensitivity of the *ordered* CPD $(\mathcal{A}_1, \ldots, \mathcal{A}_r)$ to (structured) perturbations of the tensor $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r$. Moreover, Breiding & Vannieuwenhoven (2018c) showed that $\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) = \kappa(\mathcal{A}_{\pi_1}, \ldots, \mathcal{A}_{\pi_r})$ for every permutation π of $\{1, \ldots, r\}$. This means that κ is truly the condition number of the CPD $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$. We could emphasize this notationally by writing $\kappa(\{\mathcal{A}_1, \ldots, \mathcal{A}_r\})$, but for brevity we will refrain from doing so.

Remark 1.2 It is important to note that the condition number (1.4) is an intrinsic property of the CPD $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \in (\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d})^{\times r}$ of $\mathcal{A} \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$, where $\otimes$ denotes the abstract tensor product (see Greub, 1978) and $\mathbb{R}^{n_k}$ is the Euclidean space of dimension n_k, rather than an extrinsic property of the concrete set of d-arrays $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \in (\mathbb{R}^{n_1 \times \cdots \times n_d})^{\times r}$ that represent these abstract tensors in coordinates. Indeed, to see that this is true, it suffices to show that the condition number of a coordinate representation of a CPD $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \in \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ of $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r$ is constant on $O_{n_1} \otimes \cdots \otimes O_{n_d}$ orbits, where O_n is the orthogonal group of size n. This is exactly what Breiding & Vannieuwenhoven (2018c, Proposition 5.1) states. To compute the condition number in practice it is nevertheless recommended that we work with a coordinate representation with respect to an orthogonal basis, so that (1.4) can be computed via the singular value decomposition of a matrix; see Breiding & Vannieuwenhoven (2018c, Section 5.1).

Remark 1.3 If $\mathcal{A}$ is r-identifiable and $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$ is its CPD, then $\mu(\mathcal{A}) := \kappa(\{\mathcal{A}_1, \ldots, \mathcal{A}_r\})$ can safely be called *the* condition number of computing *the* CPD of $\mathcal{A}$. However, if $\mathcal{A}$ has several distinct CPDs, then each CPD can have a different sensitivity to perturbations of $\mathcal{A}$. This is not at all exceptional: think of the conditioning of the solutions of a square system of polynomial equations where there are several distinct solutions, each of which has a different sensitivity to perturbations in the coefficients of the polynomials; see, e.g., Bürgisser & Cucker (2013, Chapter 16). This example is apt, because decomposing a rank-r tensor into its CPD can be modelled as a system of polynomial equations. Hence, different CPDs are *expected* to have different condition numbers.

The topic of this paper is the first *probabilistic analysis* of the condition number of the CPD; see, e.g., Bürgisser & Cucker (2013) and Cucker (2016). We focus on average analysis and compute the expected value of powers of the condition number for *random rank*-1 *tuples* $(\lambda_1\mathcal{A}_1,\ldots,\lambda_r\mathcal{A}_r)$ of length r, where the $\lambda_i \in \mathbb{R}\setminus\{0\}$ are arbitrary and $\mathcal{A}_i := \mathbf{a}_i^1\otimes\cdots\otimes\mathbf{a}_i^d$ in which the $\mathbf{a}_i^j \in \mathbb{R}^{n_j}$ have independently and identically distributed (i.i.d.) standard normal entries. This distribution is very relevant for scientific research, as samples from it are often used to test the effectiveness of algorithms for computing CPDs.

The considered problem is not vacuous. Indeed, the spectral characterization (1.4), Beltrán *et al.* (2019, Proposition 4.5) and the results of Chiantini *et al.* (2017, Section 3) entail that[4] 'most rank-r higher-order tensors with $d \geqslant 3$ have a finite condition number.' By contrast, for matrices ($d = 2$) the condition number is always ∞. The reason is that the rank-r CPD of a matrix $\mathcal{A} \in \mathbb{R}^{m\times n}$ is a factorization $\mathcal{A} = AB^{\mathrm{T}}$ with $A \in \mathbb{R}^{m\times r}$ and $B \in \mathbb{R}^{n\times r}$, which always has a positive-dimensional family of alternative CPDs, namely $(AX^{-1})(BX^{\mathrm{T}})^{\mathrm{T}}$ for every invertible $X \in \mathbb{R}^{r\times r}$; see Harris (1992, Example 12.1). Whenever there is such a positive-dimensional family, Φ_r is not locally invertible and by Breiding & Vannieuwenhoven (2018c, Corollary 2.3) its condition number is ∞. Indeed, *no* perturbation of $\mathcal{A}$ is required to obtain a completely different set of rank-1 matrices whose sum is $\mathcal{A}$.

In Breiding & Vannieuwenhoven (2018c, Proposition 5.1) we showed that the condition number is invariant under scaling of the rank-1 tensors $\mathcal{A}_i$:

$$\kappa(\lambda_1\mathcal{A}_1,\ldots,\lambda_r\mathcal{A}_r) = \kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r). \tag{1.5}$$

For this reason we assume, without loss of generality, that $\lambda_1 = \cdots = \lambda_r = 1$ in the remainder of this paper. One of the main results we will prove is the following statement.

Corollary 1.4 Let $(\mathcal{A}_1,\ldots,\mathcal{A}_r) \in \mathscr{S}^{\times r}$ be a random rank-1 tuple in $\mathbb{R}^{n_1\times n_2\times n_3}$, where $n_1 \geqslant n_2 \geqslant n_3 \geqslant 2$ and $r \geqslant 2$. Then we have $\mathbb{E}\left[\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r)^c\right] = \infty$ for all $c \geqslant n_3 - 1$.

In particular, the corollary states that the expected value of the condition number—without a power—of random rank-1 tuples in $\mathbb{R}^{n_1\times n_2\times 2}$ is ∞. This result provides further concrete evidence that the problem of computing a CPD can have a high condition number with a nonnegligible probability. See, for example, the curve $n = 2$ in Fig. 3, which shows the complementary cumulative distribution function of the condition number of random rank-1 tuples of length 7 in $\mathbb{R}^{7\times 7\times 2}$. It shows that there is a 10% chance that the condition number is greater than 10^4, and a 1% chance that it is greater than $4\cdot 10^5$. In many applications where the CPD is employed, the measurement errors are not sufficiently small to compensate such high condition numbers.

Corollary 1.4 is a contribution to a body of research illustrating that computing CPDs can be a very challenging problem. The result of Håstad (1990) is often cited in this regard. Håstad reduces 3SAT to computing the rank of a tensor, which shows that the latter problem is NP-complete in the Turing machine computational model. However, this does not entail that computing a *typical* CPD is a difficult problem. Another oft-cited result by de Silva & Lim (2008) relates to the difficulty of approximating CPDs; they proved that the problem of computing the best rank-2 approximation is ill posed on an open set in $\mathbb{R}^{n_1\times n_2\times n_3}$. Further evidence originates from the sensitivity to perturbations of the CPD:

[4] Specifically, there is a proper closed semialgebraic subset of $\Phi_r(\mathscr{S}^{\times r}_{n_1,\ldots,n_d})$ where r-identifiability fails.

Vannieuwenhoven (2017) illustrated numerically that the norm-balanced condition number can blow up near the ill-posed locus of de Silva & Lim (2008); subsequently, Breiding & Vannieuwenhoven (2018c) proved that the geometric condition number will diverge to infinity when approaching the ill-posed locus. Recall from Breiding & Vannieuwenhoven (2018b, Theorem 1) that the condition number appears in estimates of the rate of convergence and radii of attraction of Riemannian Gauss–Newton methods for computing a best rank-r approximation of a tensor, such as those in Breiding & Vannieuwenhoven (2018a,b). Corollary 1.4 thus not only shows that computing CPDs is a difficult problem, but also reinforces the result about the high computational complexity of computing low-rank approximations. Nevertheless, the present article is the first to study average complexity.

There are two new key insights that this paper offers. The first is decidedly negative: the average condition number of random rank-1 tuples of length r in $\mathbb{R}^{n_1 \times n_2 \times 2}$ is infinite, implying that sampling a CPD with a high condition number is *probable*; see Section 5.2. However, the second one is considerably more positive: our inability to reduce the value of c in Corollary 1.4 to $c = 1$, or even any value less than $n_3 - 1$, in our analysis should, in combination with the empirical evidence in Section 5.2 and the impossibility result in Proposition 3.8, be taken as clear evidence for the following conjecture.

CONJECTURE 1.5 There exists an integer $2 \leqslant r^\star \leqslant \frac{n_1 n_2 n_3}{n_1+n_2+n_3-2}$ such that for all $1 \leqslant r \leqslant r^\star$ the expected condition number of random rank-1 tuples of length r in $\mathbb{R}^{n_1 \times n_2 \times n_3}$ with $n_1 \geqslant n_2 \geqslant n_3 \geqslant 3$ is finite.

This would suggest that there exists a gap in sensitivity (which is one measure of complexity, as explained above) between $n_1 \times n_2 \times 2$ tensors or pairs of $n_1 \times n_2$ matrices, where the average condition number is proved to be ∞, and more general $n_1 \times n_2 \times n_3$ tensors with $n_1, n_2, n_3 > 2$, where all empirical and theoretical evidence points to a finite average condition number. This is similar to the gap in classic complexity between order-2 tensors and order-d tensors with $d \geqslant 3$ for computing the tensor rank. It is noteworthy that increasing the size of the tensor seems to decrease the complexity of computing the CPD.

Statement of the technical contributions

We proved in Breiding & Vannieuwenhoven (2018c, Theorem 1.3) that the condition number of the CPD is equal to the distance to ill-posedness in an *auxiliary space*; according to the theorem the condition number of the CPD $\kappa(\mathcal{A}_1, \dots, \mathcal{A}_r)$ at a decomposition $(\mathcal{A}_1, \dots, \mathcal{A}_r) \in \mathscr{S}^{\times r}$ is equal to the inverse distance of the tuple of tangent spaces $(\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \dots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S})$ to ill-posedness:

$$\kappa(\mathcal{A}_1, \dots, \mathcal{A}_r) = \frac{1}{\mathrm{dist}_{\mathrm{P}}((\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \dots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S}), \Sigma_{\mathrm{Gr}})}, \tag{1.6}$$

where Σ_{Gr} and the distance $\mathrm{dist}_{\mathrm{P}}$ are defined as follows. Let $n := \dim \mathscr{S}$ and write $\Pi := n_1 \cdots n_d$ for the dimension of $\mathbb{R}^{n_1 \times \cdots \times n_d}$. Denote by $\mathrm{Gr}(\Pi, n)$ the *Grassmann manifold* of n-dimensional linear spaces in the space of tensors $\mathbb{R}^{n_1 \times \cdots \times n_d} \cong \mathbb{R}^{\Pi}$. Then the tuple of tangent spaces to $\mathscr{S}$ at the decomposition $(\mathcal{A}_1, \dots, \mathcal{A}_r)$ is an element in the product of Grassmannians: $(\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \dots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S}) \in \mathrm{Gr}(\Pi, n)^{\times r}$. The set Σ_{Gr} in (1.6) is then defined as the r-tuples of linear spaces that are not in general position. In formulas,

$$\Sigma_{\mathrm{Gr}} := \left\{(W_1, \dots, W_r) \in \mathrm{Gr}(\Pi, n)^{\times r} \mid \dim(W_1 + \cdots + W_r) < rn\right\}. \tag{1.7}$$

The distance measure in (1.6) is given by the *projection distance* on $\mathrm{Gr}(\Pi, n)$. It is defined as $\|\mathrm{P}_V - \mathrm{P}_W\|$, where P_V and P_W are the orthogonal projections on the spaces V and W, respectively, and $\| \cdot \|$

is the spectral norm. From a different perspective, $\mathrm{Gr}(\Pi, n) \cong \{P \in \mathbb{R}^{\Pi \times \Pi} : P^2 = P, P^T P = \mathrm{Id}, \mathrm{rk}\,(P) = n\}$ can be seen as the set of all orthogonal projections to an n-dimensional linear space in $\mathbb{R}^{\Pi}$. In this interpretation, the projection distance is just the distance in the Frobenius norm. Nevertheless, the abstract formulation using linear spaces will be useful later. The projection distance is then extended to $\mathrm{Gr}(\Pi, n)^{\times r}$ in the usual way:

$$\mathrm{dist}_{\mathrm{P}}((V_1, \ldots, V_r), (W_1, \ldots, W_r)) := \sqrt{\sum_{i=1}^{r} \|\mathrm{P}_{V_i} - \mathrm{P}_{W_i}\|^2}. \tag{1.8}$$

The decomposition $(\mathcal{A}_1, \ldots, \mathcal{A}_r)$ whose corresponding tangent space lies in Σ_{Gr} is *ill posed* in the following sense. It was shown in Breiding & Vannieuwenhoven (2018c, Corollary 1.2) that whenever there is a smooth curve $\gamma(t) = (\mathcal{A}_1(t), \ldots, \mathcal{A}_r(t))$ such that $\mathcal{A} = \sum_{i=1}^{r} \mathcal{A}_i(t)$ is constant, even though $\gamma'(0) \neq 0$, then all of the decompositions $(\mathcal{A}_1(t), \ldots, \mathcal{A}_r(t))$ of $\mathcal{A}$ have condition number ∞. Note that in this case, the tensor $\mathcal{A}$ thus has a family of decompositions running through $(\mathcal{A}_1(0), \ldots, \mathcal{A}_r(0))$. We say that $\mathcal{A}$ is not *locally r-identifiable*. Since tensors are expected to admit only a finite number of decompositions, generically (for the precise statements see, e.g., Abo *et al.*, 2009; Chiantini & Ottaviani, 2012; Bocci *et al.*, 2014; Chiantini *et al.*, 2014), tensors that are not locally r-identifiable are very special as their parameters cannot be identified uniquely. Ill-posed decompositions are exactly those that, *using only first-order information, are indistinguishable from decompositions that are not locally r-identifiable*.

In this article we relate the condition number to a metric on the data space $\mathcal{S}^{\times r}$; see Theorem 1.6. Following Demmel (1987a) we then use this result and show in Theorem 1.7 that the expected value of the condition number is infinite whenever the ill-posed locus in $\mathcal{S}^{\times r}$ is of codimension 1. To describe the condition number as an inverse distance to ill-posedness on $\mathcal{S}^{\times r}$ we need to consider an angular distance. This is why the main theorem of this article, Theorem 1.6, is naturally stated in projective space.

THEOREM 1.6 Denote by $\pi : \mathbb{R}^{n_1 \times \cdots \times n_d} \backslash \{0\} \to \mathbb{P}(\mathbb{R}^{n_1 \times \cdots \times n_d})$ the canonical projection onto projective space. We put $\mathbb{P}\mathcal{S} := \pi(\mathcal{S})$ and for nonzero tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ we denote the corresponding class in projective space by $[\mathcal{A}] := \pi(\mathcal{A})$. Let $(\mathcal{A}_1, \ldots, \mathcal{A}_r) \in \mathcal{S}^{\times r}$. Then

$$\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) \geqslant \frac{1}{\mathrm{dist}_{\mathrm{w}}(([\mathcal{A}_1], \ldots, [\mathcal{A}_r]), \Sigma_{\mathbb{P}})},$$

where

$$\Sigma_{\mathbb{P}} = \left\{([\mathcal{A}_1], \ldots, [\mathcal{A}_r]) \in (\mathbb{P}\mathcal{S})^{\times r} \mid \kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) = \infty\right\}$$

and the distance $\mathrm{dist}_{\mathrm{w}}$ is defined in Definition 2.1.

This characterization of a condition number as an inverse distance to ill-posedness is called a *condition number theorem* in the literature and it provides a geometric interpretation of the complexity of a computational problem. Demmel (1987a) advocates this characterization as it may be used to 'compute the probability distribution of the distance from a "random" problem to the set [of ill-posedness].' Condition number theorems were, for instance, derived for matrix inversion (Eckart & Young, 1939; Kahan, 1966; Demmel, 1987b), polynomial zero finding (Hough, 1977; Demmel, 1987b) and computing eigenvalues (Wilkinson, 1972; Demmel, 1987b). For a comprehensive overview see Bürgisser & Cucker (2013, pages 10, 16, 125, 204). We use the above partial condition number theorem to derive a result on the average condition number of CPDs.

Theorem 1.7 Let $(\mathcal{A}_1,\ldots,\mathcal{A}_r)\in\mathscr{S}^{\times r}$, $r\geqslant 2$ be a random r-tuple of rank-1 tensors in $\mathbb{R}^{n_1\times\cdots\times n_d}$. Let $e\geqslant c\geqslant 1$. If $\varSigma_{\mathbb{P}}$ contains a manifold of codimension 0 or c in $\mathscr{S}^{\times r}$, then $\mathbb{E}\left[\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r)^e\right]=\infty$.

In Section 3 we prove that for the format $n_1\times n_2\times n_3$, $n_1\geqslant n_2\geqslant n_3\geqslant 2$, the ill-posed locus $\varSigma_{\mathbb{P}}$ contains a submanifold that is of codimension n_3-1 in $\mathscr{S}^{\times r}$. Hence, the aforementioned Corollary 1.4 is obtained as a consequence of Theorem 1.7.

Remark 1.8 The statement of Corollary 1.4 can easily be strengthened as follows. It is known from dimensionality arguments about fibres of projections of projective varieties that there exists an integer critical value $r^\star\leqslant\frac{\dim\mathbb{R}^{n_1\times\cdots\times n_d}}{\dim\mathscr{S}}$ such that every tensor of rank $r>r^\star$ has at least a one-dimensional variety of rank decompositions in $\mathscr{S}^{\times r}$ (see, e.g., Harris, 1992; Abo *et al.*, 2009; Landsberg, 2012). Specifically, $r^\star$ is the smallest value such that the dimension of the projective $(r^\star+1)$-secant variety of $\mathbb{P}\mathscr{S}$ is strictly less than $(r^\star+1)\dim\mathscr{S}-1$. It follows then from Breiding & Vannieuwenhoven (2018c, Corollary 1.2) that the condition number $\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r)=\infty$ for *all* decompositions $(\mathcal{A}_1,\ldots,\mathcal{A}_r)$ when $r>r^\star$. For smaller values of r we can prove only the statement in Corollary 1.4.

Structure of the article

The rest of this paper is structured as follows. In the next section we recall some preliminary material on Riemannian geometry. We start by proving the main contribution in Section 3, namely Theorem 1.7, because its proof is less technical. Section 4 is devoted to the proof of the condition number theorem, namely Theorem 1.6. In Section 5 we present some numerical experiments and computer algebra computations illustrating the main contributions. Finally, the paper is concluded in Section 6.

2. Preliminaries and notation

We denote the standard Euclidean inner product on $\mathbb{R}^m$ by $\langle\cdot,\cdot\rangle$. The real projective space of dimension $m-1$ is denoted by $\mathbb{P}(\mathbb{R}^m)$ and the unit sphere of dimension $m-1$ is denoted by $\mathbb{S}(\mathbb{R}^m)$. Points in linear spaces are typeset in boldface lower-case symbols like $\mathbf{a},\mathbf{x}$. Points in projective space or other manifolds are typeset in lowercase letters like a,x. The *orthogonal complement* of a point $\mathbf{x}\in\mathbb{R}^m$ is $\mathbf{x}^\perp:=\{\mathbf{y}\in\mathbb{R}^m\mid\langle\mathbf{x},\mathbf{y}\rangle=0\}$. We write $\mathscr{S}$ for the Segre manifold in $\mathbb{R}^{n_1\times\cdots\times n_d}$. If it is necessary to clarify the parameters we also write $\mathscr{S}_{n_1,\ldots,n_d}$. Throughout this paper, n denotes the dimension of $\mathscr{S}$:

$$n:=\dim\mathscr{S}_{n_1,\ldots,n_d}=1-d+\sum_{i=1}^d n_i; \tag{2.1}$$

see Harris (1992) and Landsberg (2012). The projective Segre map is

$$\sigma:\mathbb{P}(\mathbb{R}^{n_1})\times\cdots\times\mathbb{P}(\mathbb{R}^{n_d})\to\mathbb{P}\mathscr{S},\quad([\mathbf{a}^1],\ldots,[\mathbf{a}^d])\mapsto[\mathbf{a}^1\otimes\cdots\otimes\mathbf{a}^d]; \tag{2.2}$$

see Landsberg (2012, Section 4.3.4.).

Let $(\mathscr{M},g)$ be a Riemannian manifold. For $x\in\mathscr{M}$ we write $\mathrm{T}_x\mathscr{M}$ for the tangent space of $\mathscr{M}$ at x. For $\gamma:(-1,1)\to\mathscr{M}$ a smooth curve in $\mathscr{M}$ we will use the shorthand notation $\gamma'(0):=\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\gamma(t)$ for the tangent vector in $\mathrm{T}_{\gamma(0)}\mathscr{M}$ and $\gamma'(t):=\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)$. Recall that the *Riemannian distance* between two points $p,q\in\mathscr{M}$ is $\mathrm{dist}_{\mathscr{M}}(p,q)=\inf\{l(\gamma)\mid\gamma(0)=p,\gamma(1)=q\}$. The infimum is over all piecewise differentiable curves $\gamma : [0,1] \to \mathscr{M}$ and the length of a curve is $l(\gamma) = \int_0^1 g(\gamma'(t), \gamma'(t))^{\frac{1}{2}}\,\mathrm{d}t$. The distance $\mathrm{dist}_{\mathscr{M}}$ makes $\mathscr{M}$ a metric space (do Carmo, 1992, Proposition 2.5).

We use the symbol $|\omega|$ to denote the *density* on $\mathscr{M}$ given by g (Lee, 2013, Proposition 16.45). For densities with finite volume, i.e., $\int_{\mathscr{M}} |\omega| < \infty$, this defines the *uniform distribution*:

$$\underset{X \text{ uniformly in } \mathscr{M}}{\mathrm{Prob}} \{X \in N\} := \frac{1}{\int_{\mathscr{M}} |\omega|} \int_N |\omega| \quad \text{where } N \subset \mathscr{M}.$$

A particularly important manifold in the context of this article is the projective space $\mathbb{P}(\mathbb{R}^{m+1})$. An atlas for $\mathbb{P}(\mathbb{R}^{m+1})$ is, for instance, given by the affine charts (U_i, φ_i) with $U_i = \{(x_0 : \ldots : x_m) \mid x_i \neq 0\}$ and $\varphi_i(x_0 : \ldots : x_m) = (\frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_m}{x_i})$. A Riemannian structure on $\mathbb{P}(\mathbb{R}^n)$ is the *Fubini–Study metric* (see, e.g., Bürgisser & Cucker, 2013, Section 14.2.2); the tangent space to x can be identified with

$$\mathrm{T}_x\mathbb{P}(\mathbb{R}^n) \cong \mathbf{x}^\perp, \quad \text{where } \mathbf{x} \in x \text{ is a representative,} \tag{2.3}$$

and through this identification the Fubini–Study metric on $\mathbf{x}^\perp$ is $\langle \mathbf{y}_1, \mathbf{y}_2 \rangle_{\mathbf{x}} := \frac{\langle \mathbf{y}_1, \mathbf{y}_2 \rangle}{\|\mathbf{x}\|^2}$. The Fubini–Study distance $d_{\mathbb{P}}$ is the distance associated to the Fubini–Study metric. For points $x, y \in \mathbb{P}(\mathbb{R}^n)$ the formula is

$$d_{\mathbb{P}}(x,y) = \arccos\left(\frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{x}\|\,\|\mathbf{y}\|}\right), \quad \text{where } \mathbf{x} \in x, \mathbf{y} \in y \text{ are representatives.}$$

For the Fubini–Study distance in $\mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$ we write

$$\mathrm{dist}_{\mathbb{P}}((x_1, \ldots, x_d), (y_1, \ldots, y_d)) := \sqrt{\sum_{i=1}^d d_{\mathbb{P}}(x_i, y_i)^2}. \tag{2.4}$$

The weighted distance, which is the protagonist of Theorem 1.6, is introduced next.

Definition 2.1 (Weighted distance). The *weighted distance* between two points $p = (p_1, \ldots, p_d)$ and $q = (q_1, \ldots, q_d) \in \mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$ is defined as

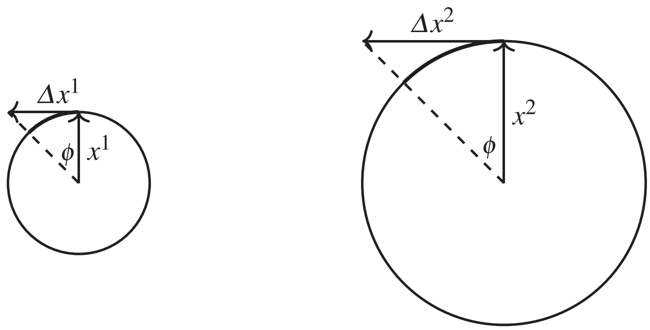
$$d_{\mathrm{w}}(p,q) := \sqrt{\sum_{i=1}^d (n - n_i) d_{\mathbb{P}}(p_i, q_i)^2},$$

where, as before, $n = \dim \mathscr{S}$. The weighted distance on $\mathscr{S}^{\times r}$ then is defined as

$$\mathrm{dist}_{\mathrm{w}}((\mathcal{A}_1, \ldots, \mathcal{A}_r), (\mathcal{B}_1, \ldots, \mathcal{B}_r)) := \sqrt{\sum_{i=1}^r d_{\mathrm{w}}(\sigma^{-1}(\mathcal{A}_i), \sigma^{-1}(\mathcal{B}_i))^2},$$

where σ^{-1} is the inverse of the projective Segre map from (2.2).

For $n_1 > n_2$ the relative errors in the factor $\mathbb{P}(\mathbb{R}^{n_2})$ weigh more than relative errors in the factor $\mathbb{P}(\mathbb{R}^{n_1})$ when the measure is the weighted distance d_{w}; this is illustrated in Fig. 1.



$$\tan\phi = \frac{\|\Delta x^1\|}{\|x^1\|} = \frac{\|\Delta x^2\|}{\|x^2\|}$$

FIG. 1. The picture depicts relative errors in the weighted distance, where $x^1 \in \mathbb{P}(\mathbb{R}^{n_1})$ and $x^2 \in \mathbb{P}(\mathbb{R}^{n_2})$ with $n_1 > n_2$. The relative errors of the tangent directions Δx^1 and Δx^2 are both equal to $\tan\phi$, but the contribution to the weighted distance is larger for the large circle, which corresponds to the smaller projective space $\mathbb{P}(\mathbb{R}^{n_2})$.

REMARK 2.2 One may wonder why the weighted distance is defined as the weighted pull-back of the Fubini-Study metric, and not of the usual Euclidean metric that $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ inherits from the Euclidean metrics of its factors. The reason is equation (1.5). The condition number is invariant under scaling of the individual rank-1 summands. This means that, ultimately, the condition number is a property of rank-1 tuples in $(\mathbb{P}\mathscr{S})^{\times r}$, rather than of tuples in $\mathscr{S}^{\times r}$. Still, we regard the latter as the *output data space* and define the weighted distance there as well. One should have in mind that in this space, the weighted distance is not truly a distance, since $d_w((\lambda_1\mathcal{A}_1, \ldots, \lambda_r\mathcal{A}_r), (\mathcal{A}_1, \ldots, \mathcal{A}_r)) = 0$ for all nonzero $\lambda_1, \ldots, \lambda_r$ and for all $(\mathcal{A}_1, \ldots, \mathcal{A}_r) \in \mathscr{S}^{\times r}$.

3. The expected value of the condition number

Before proving Theorem 1.7 we need four auxiliary lemmata. The first provides a deterministic lower bound of the condition number.

LEMMA 3.1 Let $r \geqslant 1$. For all rank-1 tuples $(\mathcal{A}_1, \ldots, \mathcal{A}_r)$ in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ we have $\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) \geqslant 1$.

Proof. The condition number equals the inverse of the smallest singular value of a matrix all of whose columns are of unit length by Breiding & Vannieuwenhoven (2018c, Theorem 1.1). The result follows from the min-max characterization of the smallest singular value. □

The next lemma is a basic computation in Riemannian geometry.

LEMMA 3.2 Let $\mathscr{M}$ be a Riemannian manifold, and $\mathscr{N}$ a codimension c submanifold of $\mathscr{M}$. Let $\mathrm{dist}_{\mathscr{M}}$ denote the Riemannian distance on $\mathscr{M}$ and $|\omega|$ be the density on $\mathscr{M}$. Then,

$$\int_{x\in\mathscr{M}} \left(\frac{1}{\mathrm{dist}_{\mathscr{M}}(x, \mathscr{N})}\right)^c |\omega| = \infty.$$

Proof. Let m, k be the dimensions of $\mathscr{M}, \mathscr{N}$ and let $y \in \mathscr{N}$ be any point. Let $\epsilon > 0$. From the definition of being a submanifold, there exists an open neighbourhood $\mathscr{U}$ of y in $\mathscr{M}$ and a diffeomorphism $\phi : \mathscr{U} \to B_\epsilon(\mathbb{R}^k) \times B_\epsilon(\mathbb{R}^{m-k})$, such that $\mathscr{N} \cap \mathscr{U} = \phi^{-1}(B_\epsilon(\mathbb{R}^k) \times \{0\})$, where $B_\epsilon(\mathbb{R}^m)$ is the open ball of radius ϵ in $\mathbb{R}^m$. By compactness, choosing ϵ small enough, we can assume that there are positive constants C, D such that the derivative of ϕ satisfies $\|\mathrm{d}_x\phi\| \leqslant C$, $\|\mathrm{d}_x\phi^{-1}\| \leqslant C$ and $|\det(\mathrm{d}_x\phi)| \geqslant D$ for all $x \in \mathscr{U}$. In particular, the length L of a curve in $\mathscr{U}$ and the length L' of its image under ϕ satisfy $L \leqslant CL'$. Writing $(\mathbf{x}_1, \mathbf{x}_2) := \phi(x)$ for the image of x under ϕ we thus have $\mathrm{dist}_{\mathscr{M}}(x, \mathscr{N}) \leqslant C\|\mathbf{x}_2\|$. The change of variables theorem, i.e., Spivak (1965, Theorem 3-13), gives

$$\int_{x\in\mathscr{U}} \left(\frac{1}{\mathrm{dist}_{\mathscr{M}}(x,\mathscr{N})}\right)^c |\omega| = \int_{x\in\mathscr{U}} \left(\frac{1}{\mathrm{dist}_{\mathscr{M}}(x,\mathscr{N})}\right)^{m-k} |\omega|$$
$$\geqslant \frac{D}{C^{m-k}} \int_{(\mathbf{x}_1,\mathbf{x}_2)\in B_\epsilon(\mathbb{R}^k)\times B_\epsilon(\mathbb{R}^{m-k})} \frac{1}{\|\mathbf{x}_2\|^{m-k}}\, \mathrm{d}\mathbf{x}_1 \mathrm{d}\mathbf{x}_2.$$

Up to positive constants, using Fubini's theorem, i.e., Spivak (1965, Theorem 3–10), and passing to polar coordinates, this last integral equals

$$\int_{\mathbf{x}_1\in B_\epsilon(\mathbb{R}^k)} \int_0^\epsilon \frac{t^{m-k-1}}{t^{m-k}}\, \mathrm{d}t\, \mathrm{d}\mathbf{x}_1 = \infty.$$

The lower bound for the integral in the lemma then follows from

$$\int_{x\in\mathscr{M}} \left(\frac{1}{\mathrm{dist}_{\mathscr{M}}(x,\mathscr{N})}\right)^c |\omega| \geqslant \int_{x\in\mathscr{U}} \left(\frac{1}{\mathrm{dist}_{\mathscr{M}}(x,\mathscr{N})}\right)^c |\omega|.$$

This finishes the proof. □

Inspecting Theorem 1.6 we see that combining it with the above lemma contains the key idea for proving that the expected value of the condition number can be infinite. However, to use these results in our proof of Theorem 1.7, we need to ensure that Lemma 3.2 applies. Theorem 1.6 uses the weighted distance from Definition 2.1 and it is not immediately evident whether it is induced by a Riemannian metric on $\mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$. Fortunately, the next lemma shows that it is.

LEMMA 3.3 Let $\langle\cdot,\cdot\rangle$ be the Fubini–Study metric on $\mathbb{P}(\mathbb{R}^{n_i})$. We define the weighted inner product $\langle\cdot,\cdot\rangle_{\mathrm{w}}$ on the tangent space at $p \in \mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$ as follows. For all $\mathbf{u}, \mathbf{v} \in \mathrm{T}_p(\mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d}))$, where $\mathbf{u} = (\mathbf{u}^1, \dots, \mathbf{u}^d)$ and $\mathbf{v} = (\mathbf{v}^1, \dots, \mathbf{v}^d)$, we define $\langle \mathbf{u}, \mathbf{v}\rangle_{\mathrm{w}} := \sum_{i=1}^d (n - n_i)\langle \mathbf{u}^i, \mathbf{v}^i\rangle_{p_i}$. Then the distance on $\mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$ corresponding to $\langle\cdot,\cdot\rangle_{\mathrm{w}}$ is d_{w} from Definition 2.1.

Proof. Let $\gamma(t) = (\gamma_1(t), \dots, \gamma_d(t))$ be a piecewise continuous curve in $\mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$ connecting $p, q \in \mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$, such that the distance between p, q given by $\langle\cdot,\cdot\rangle_{\mathrm{w}}$ is

$$\int_0^1 \langle \gamma'(t), \gamma'(t)\rangle_{\mathrm{w}}^{\frac{1}{2}}\, \mathrm{d}t = \int_0^1 \left(\sum_{i=1}^d (n - n_i)\langle \gamma_i'(t), \gamma_i'(t)\rangle\right)^{\frac{1}{2}} \mathrm{d}t.$$

Because $(n - n_i)\langle \gamma_i'(t), \gamma_i'(t)\rangle = \langle \sqrt{n - n_i}\, \gamma_i'(t), \sqrt{n - n_i}\, \gamma_i'(t)\rangle$ and because we have the identity of tangent spaces $\mathrm{T}_{\gamma_i(t)}\mathbb{P}(\mathbb{R}^{n_i}) = \mathrm{T}_{\gamma_i(t)}\mathbb{S}(\mathbb{R}^{n_i})$ for all i and t, we may view γ as the shortest path between two points on a product of d spheres with radii $\sqrt{n - n_1}, \dots, \sqrt{n - n_d}$. The length of this path is $d_{\mathrm{w}}(p, q)$. □

Let σ be the projective Segre map from (2.2). By Landsberg (2012, Section 4.3.4.), σ is a diffeomorphism and we define a Riemannian metric g on $\mathbb{P}\mathscr{S}$ to be the pull-back metric of $\langle\cdot,\cdot\rangle_{\mathrm{w}}$ under σ^{-1}; see Lee (2013, Proposition 13.9). Then, by construction, we have the following result.

COROLLARY 3.4 The weighted distance $\mathrm{dist}_{\mathrm{w}}$ on $(\mathbb{P}\mathscr{S})^{\times r}$ is given by the Riemannian metric g.

The last technical lemma we need is the following.

LEMMA 3.5 Consider the projective Segre map $\sigma : \mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d}) \to \mathbb{P}\mathscr{S}$ from (2.2). For any point $p = ([\mathbf{a}_1], \ldots, [\mathbf{a}_d]) \in \mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d})$ we have $|\det(\mathrm{d}_p\sigma)| = 1$.

Proof. We denote by $\mathbf{e}_i^j$ the ith standard basis vector of $\mathbb{R}^{n_j}$; i.e., $\mathbf{e}_i^j$ has zeros everywhere except for the ith entry, where it has a 1. To ease notation, let us assume $\mathbf{e}_i^j$ to be a row vector. Because each point in $\mathbb{P}(\mathbb{R}^{n_j})$ is in an orbit of $[\mathbf{e}_1^j]$ under the orthogonal group, it suffices to show the claim for $p = ([\mathbf{e}_1^1], \ldots, [\mathbf{e}_1^d])$. By (2.3), an orthonormal basis for the tangent space $\mathrm{T}_{[\mathbf{e}_1^j]}\mathbb{P}(\mathbb{R}^{n_j})$ is $\{\mathbf{e}_2^j, \ldots, \mathbf{e}_{n_j}^j\}$. Hence, an orthonormal basis for $\mathrm{T}_p(\mathbb{P}(\mathbb{R}^{n_1}) \times \cdots \times \mathbb{P}(\mathbb{R}^{n_d}))$ is

$$\bigcup_{j=1}^{d} \{(\underbrace{0, \ldots, 0}_{j-1 \text{ times}}, \mathbf{e}_i^j, \underbrace{0, \ldots, 0}_{d-j \text{ times}}) \mid 2 \leqslant i \leqslant n_j\}.$$

Fix $1 \leqslant j \leqslant d$ and $2 \leqslant i \leqslant n_j$. Then, by the product rule, we have

$$\mathrm{d}_p\sigma(0, \ldots, 0, \mathbf{e}_i^j, 0, \ldots, 0) = \mathbf{e}_1^1 \otimes \cdots \otimes \mathbf{e}_1^{j-1} \otimes \mathbf{e}_i^j \otimes \mathbf{e}_1^{j+1} \otimes \cdots \otimes \mathbf{e}_1^d.$$

It is easily verified that $\{\mathbf{e}_1^1 \otimes \cdots \otimes \mathbf{e}_1^{j-1} \otimes \mathbf{e}_i^j \otimes \mathbf{e}_1^{j+1} \otimes \cdots \otimes \mathbf{e}_1^d \mid 1 \leqslant j \leqslant d, 2 \leqslant i \leqslant n_j\}$ is an orthonormal basis of $\mathrm{T}_{\sigma(p)}\mathbb{P}\mathscr{S}$ (for instance, by using Lemma A1 below). This shows that $\mathrm{d}_p\sigma$ maps an orthonormal basis to an orthonormal basis. Hence, $|\det(\mathrm{d}_p\sigma)| = 1$. □

REMARK 3.6 In fact, the proof of the foregoing lemma shows more than $|\det(\mathrm{d}_p\sigma)| = 1$. Namely, it shows that σ is an *isometry* in the sense of Definition 4.1.

Now we have gathered all the ingredients to prove Theorem 1.7.

Proof of Theorem 1.7. First we use that the condition number is *scale invariant*. That is, for all $t_1, \ldots, t_r \in \mathbb{R}\backslash\{0\}$ we have by Breiding & Vannieuwenhoven (2018c, Proposition 4.4),

$$\kappa(t_1\mathcal{A}_1, \ldots, t_r\mathcal{A}_r) = \kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) =: \kappa([\mathcal{A}_1], \ldots, [\mathcal{A}_r]),$$

where $\mathcal{A}_i$ are rank-1 tensors. This implies that the random variable under consideration is independent of the scaling of the factors $\mathbf{a}_i^j$ and, consequently, we have (see, e.g., Bürgisser & Cucker, 2013, Remark 2.24)

$$\underset{\substack{\mathbf{a}_i^j \in \mathbb{R}^{n_j}, 1\leqslant j\leqslant d, 1\leqslant i\leqslant r \\ \text{standard normal i.i.d.}}}{\mathbb{E}} \left[\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)^c\right] = \underset{\substack{\mathbf{a}_i^j \in \mathbb{P}(\mathbb{R}^{n_j}), 1\leqslant j\leqslant d, 1\leqslant i\leqslant r \\ \text{uniformly i.i.d.}}}{\mathbb{E}} \left[\kappa([\mathcal{A}_1], \ldots, [\mathcal{A}_r])^c\right],$$

where $\mathcal{A}_i = \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$. Let $|\omega|$ denote the density on $(\mathbb{P}\mathcal{S})^{\times r} = (\mathbb{P}\mathcal{S}_{n_1,\ldots,n_d})^{\times r}$. By Lemma 3.4, the Jacobian of the change of variables via the projective Segre map σ is constant and equal to 1. Hence,

$$\underset{\substack{\mathbf{a}_i^j \in \mathbb{P}(\mathbb{R}^{n_j}), 1\leqslant j\leqslant d, 1\leqslant i\leqslant r\\ \text{uniformly i.i.d.}}}{\mathbb{E}} \left[\kappa([\mathcal{A}_1],\ldots,[\mathcal{A}_r])^c\right] = \frac{1}{C}\int_{([\mathcal{A}_1],\ldots,[\mathcal{A}_r])\in(\mathbb{P}\mathcal{S})^{\times r}} \kappa([\mathcal{A}_1],\ldots,[\mathcal{A}_r])^c\,|\omega|,$$

where $C = \int_{(\mathbb{P}\mathcal{S})^{\times r}} |\omega| < \infty$, because $(\mathbb{P}\mathcal{S})^{\times r}$ is compact. For brevity we write $\mathfrak{p} = ([\mathcal{A}_1],\ldots,[\mathcal{A}_r]) \in (\mathbb{P}\mathcal{S})^{\times r}$. Then, by Theorem 1.6, we have

$$\int_{\mathfrak{p}\in(\mathbb{P}\mathcal{S})^{\times r}} \kappa(\mathfrak{p})^c\,|\omega| \geqslant \int_{\mathfrak{p}\in(\mathbb{P}\mathcal{S})^{\times r}} \left(\frac{1}{\operatorname{dist}_{\mathrm{w}}(\mathfrak{p},\Sigma_{\mathbb{P}})}\right)^c |\omega|.$$

We cannot directly apply Lemma 3.7 here, because the weighted distance $\operatorname{dist}_{\mathrm{w}}$ *is not given by the product Fubini–Study metric*. However, from the definitions of the weighted distance and the Fubini–Study distance (2.4), we find $\operatorname{dist}_{\mathrm{w}}(\mathfrak{p},\Sigma_{\mathbb{P}}) \leqslant \sqrt{n}\,\operatorname{dist}_{\mathbb{P}}(\mathfrak{p},\Sigma_{\mathbb{P}})$. Therefore, we have

$$\int_{\mathfrak{p}\in(\mathbb{P}\mathcal{S})^{\times r}} \left(\frac{1}{\operatorname{dist}_{\mathrm{w}}(\mathfrak{p},\Sigma_{\mathbb{P}})}\right)^c |\omega| \geqslant n^{-\frac{c}{2}} \int_{\mathfrak{p}\in(\mathbb{P}\mathcal{S})^{\times r}} \left(\frac{1}{\operatorname{dist}_{\mathbb{P}}(\mathfrak{p},\Sigma_{\mathbb{P}})}\right)^c |\omega|.$$

By assumption, there is a manifold $\mathcal{U} \subset \Sigma_{\mathbb{P}}$ of codimension c in $\mathcal{S}^{\times r}$. Applying Lemma 3.2 to this manifold we have

$$\int_{\mathfrak{p}\in(\mathbb{P}\mathcal{S})^{\times r}} \left(\frac{1}{\operatorname{dist}_{\mathbb{P}}(\mathfrak{p},\Sigma_{\mathbb{P}})}\right)^c |\omega| \geqslant \int_{\mathfrak{p}\in(\mathbb{P}\mathcal{S})^{\times r}} \left(\frac{1}{\operatorname{dist}_{\mathbb{P}}(\mathfrak{p},\mathcal{U})}\right)^c |\omega| = \infty.$$

Putting all the equalities and inequalities together we therefore get

$$\underset{\substack{\mathbf{a}_i^j \in \mathbb{R}^{n_j}, 1\leqslant j\leqslant d, 1\leqslant i\leqslant r\\ \text{standard normal i.i.d.}}}{\mathbb{E}} \left[\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r)^c\right] = \infty.$$

By Lemma 3.1, the condition number satisfies $\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r) \geqslant 1$ for every $(\mathcal{A}_1,\ldots,\mathcal{A}_r) \in \mathcal{S}^{\times r}$. This, together with the foregoing equation, implies for $c \leqslant e$,

$$\underset{\substack{\mathbf{a}_i^j \in \mathbb{R}^{n_j}, 1\leqslant j\leqslant d, 1\leqslant i\leqslant r\\ \text{standard normal i.i.d.}}}{\mathbb{E}} \left[\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r)^e\right] = \infty.$$

The proof is finished. □

Next we investigate a particular corollary of the foregoing result. We will show that for third-order tensors $\mathbb{R}^{n_1\times n_2\times n_3}$, $n_1 \geqslant n_2 \geqslant n_3 \geqslant 2$, the expected value of the (n_3-1)th power of the condition number of random rank-r tensors is indeed ∞. The following is the key ingredient.

LEMMA 3.7 Let $\mathcal{S}$ be the Segre manifold in $\mathbb{R}^{n_1\times n_2\times n_3}$, $n_1 \geqslant n_2 \geqslant n_3 \geqslant 2$, and let $\Sigma_{\mathbb{P}} \subset (\mathbb{P}\mathcal{S})^{\times r}$ be the ill-posed locus. Then there is a subvariety $\mathcal{V} \subset \Sigma_{\mathbb{P}}$ of codimension $n_3 - 1$ in $(\mathbb{P}\mathcal{S})^{\times r}$.

Proof. Consider the regular map

$$\psi : \left(\mathbb{P}(\mathbb{R}^{n_1}) \times \mathbb{P}(\mathbb{R}^{n_2}) \times \mathbb{P}(\mathbb{R}^{n_3})\right)^{\times r-1} \times \mathbb{P}(\mathbb{R}^{n_1}) \times \mathbb{P}(\mathbb{R}^{n_2}) \to (\mathbb{P}\mathscr{S})^{\times r},$$
$$\left(([\mathbf{a}_i],[\mathbf{b}_i],[\mathbf{c}_i])_{i=1}^{r-1},([\mathbf{a}_r],[\mathbf{b}_r])\right) \mapsto \left(([\mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i])_{i=1}^{r-1},[\mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_1]\right).$$

The image of ψ, write $\mathscr{V} = \mathrm{Im}(\psi)$, is a projective variety by Harris (1992, Theorem 3.13). Because the projective Segre map from (2.2) is a bijection, the fibre of ψ at any point in $\mathscr{V}$ consists of precisely one point. As a result, by Harris (1992, Theorem 11.12), dim $\mathscr{V}$ equals the dimension of the source, which is seen to be $r(\dim \mathbb{P}\mathscr{S}) - n_3 + 1$, i.e., $\mathrm{codim}(\mathscr{V}) = n_3 - 1$.

Next we show that $\mathscr{V} \subset \Sigma_{\mathbb{P}}$, which then concludes the proof. Let $[\mathcal{A}_i] = [\mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i]$ be such that $([\mathcal{A}_1],\ldots,[\mathcal{A}_r]) \in \mathscr{V}$. Thus, $[\mathbf{c}_r] = [\mathbf{c}_1]$. Consider the (affine) tangent spaces

$$\mathrm{T}_{\mathcal{A}_1}\mathscr{S} = \mathrm{T}_{\mathbf{a}_1\otimes\mathbf{b}_1\otimes\mathbf{c}_1}\mathscr{S} = \mathbb{R}^{n_1} \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \mathbf{a}_1 \otimes (\mathbf{b}_1)^{\perp} \otimes \mathbf{c}_1 + \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes (\mathbf{c}_1)^{\perp} \text{ and}$$
$$\mathrm{T}_{\mathcal{A}_r}\mathscr{S} = \mathrm{T}_{\mathbf{a}_r\otimes\mathbf{b}_r\otimes\mathbf{c}_1}\mathscr{S} = (\mathbf{a}_r)^{\perp} \otimes \mathbf{b}_r \otimes \mathbf{c}_1 + \mathbf{a}_r \otimes \mathbb{R}^{n_2} \otimes \mathbf{c}_1 + \mathbf{a}_r \otimes \mathbf{b}_r \otimes (\mathbf{c}_1)^{\perp}.$$

They intersect at least in the one-dimensional subspace $\{\alpha\, \mathbf{a}_r \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 \mid \alpha \in \mathbb{R}\}$. This means that

$$\mathrm{dist}_{\mathrm{P}}((\mathrm{T}_{\mathcal{A}_1}\mathscr{S},\ldots,\mathrm{T}_{\mathcal{A}_r}\mathscr{S}),\Sigma_{\mathrm{Gr}}) = 0;$$

hence, by (1.6), $\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r) = \infty$ and so $([\mathcal{A}_1],\ldots,[\mathcal{A}_r]) \in \Sigma_{\mathbb{P}}$. □

We can now wrap up the proof of Corollary 1.4.

Proof of Corollary 1.4. Lemma 3.7 shows there is a subvariety $\mathscr{V} \subset \Sigma_{\mathbb{P}}$ with codimension equal to $n_3 - 1$. Let p be any smooth point in this subvariety, and consider a neighbourhood $\mathscr{U}$ of p in $\mathscr{V}$ such that all points in $\mathscr{U}$ are smooth points of $\mathscr{V}$. Then $\mathscr{U}$ is a submanifold of $\Sigma_{\mathbb{P}}$ that has codimension $n_3 - 1$ in $(\mathbb{P}\mathscr{S})^{\times r}$. Hence, Theorem 1.7 applies and Corollary 1.4 is proven. □

Lemma 3.7 still leaves some doubt over the precise codimension of $\Sigma_{\mathbb{P}}$ in other tensor formats than $n_1 \times n_2 \times 2$. It might be possible to sharpen Corollary 1.4. Namely, if there exists a submanifold $\mathscr{M}$ of codimension $k < n_3 - 1$ in $(\mathbb{P}\mathscr{S})^{\times r}$ with $\mathscr{M} \subset \Sigma_{\mathbb{P}}$, then we also have $\mathbb{E}[\kappa(\mathcal{A}_1,\ldots,\mathcal{A}_r)^k] = \infty$. For small tensors we can compute the codimension of the ill-posed locus using computer algebra software. Employing Macaulay2 (Grayson & Stillman, 2018) we were able to show that Lemma 3.7 cannot be improved for small tensors with rank $r = 2$.

PROPOSITION 3.8 Let $\mathscr{S}$ be the Segre manifold in $\mathbb{R}^{n_1\times n_2\times n_3}$, $10 \geqslant n_1 \geqslant n_2 \geqslant n_3 \geqslant 2$, and let $\Sigma_{\mathbb{P}} \subset (\mathbb{P}\mathscr{S})^{\times 2}$ be the ill-posed locus. Then there is no subvariety $\mathscr{V} \subset \Sigma_{\mathbb{P}}$ of codimension $k < n_3 - 1$.

Proof. It is an exercise to verify that the Segre manifold $\mathscr{S}$ is covered by the charts $(U_{i,j},\phi_{i,j}^{-1})$, defined uniquely as follows: $U_{i,j} := \mathrm{Im}(\phi_{i,j})$ and $\phi_{i,j}$ is the diffeomorphism

$$\phi_{i,j} : \mathbb{R}^{n_1-1} \times \mathbb{R}^{n_2-1} \times \mathbb{R}^{n_3} \to \mathbb{R}^{n_1\times n_2\times n_3},$$
$$(\mathbf{x},\mathbf{y},\mathbf{z}) \mapsto (x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_{n_1-1}) \otimes (y_1,\ldots,y_{j-1},1,y_{j+1},\ldots,y_{n_2-1}) \otimes \mathbf{z}.$$

Let $\mathcal{A}_1 \in U_{i_1,j_1}$ and $\mathcal{A}_2 \in U_{i_2,j_2}$ and $p_1 = \phi_{i_1,j_1}^{-1}(\mathcal{A}_1)$, $p_2 = \phi_{i_2,j_2}^{-1}(\mathcal{A}_2)$. The corresponding rank-2 tensor is $\Phi(\mathcal{A}_1, \mathcal{A}_2) = \mathcal{A}_1 + \mathcal{A}_2$. By definition of the derivative of the addition map Φ, its matrix with respect to an orthonormal basis for $\phi_{i_1,j_1}^{-1}(U_{i_1,j_1}) \times \phi_{i_2,j_2}^{-1}(U_{i_2,j_2})$ and the standard basis on $\mathbb{R}^{n_1\times n_2\times n_3} \simeq \mathbb{R}^{n_1n_2n_3}$ is the Jacobian of the transformation $\Phi \circ (\phi_{i_1,j_1}^{-1} \times \phi_{i_2,j_2}^{-1})$ (Lee, 2013, pages 55–65). For example, if $i_1 = j_1 = i_2 = j_2$ and $n_1 = n_2 = n_3 = 2$, then the derivative $\mathrm{d}_{(\mathcal{A}_1,\mathcal{A}_2)}\Phi$ is represented in bases as the 8×8 Jacobian matrix of the map from $(\mathbb{R}^1 \times \mathbb{R}^1 \times \mathbb{R}^2) \times (\mathbb{R}^1 \times \mathbb{R}^1 \times \mathbb{R}^2) \to \mathbb{R}^8$ taking

$$(a_2, b_2, c_1, c_2) \times (x_2, y_2, z_1, z_2) \mapsto \begin{bmatrix} 1 \\ a_2 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ b_2 \end{bmatrix} \otimes \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} 1 \\ x_2 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ y_2 \end{bmatrix} \otimes \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

The ill-posed locus is then the projectivization of the locus where these Jacobian matrices have linearly dependent columns. Note that the codimension of $\Sigma_{\mathbb{P}} \subset (\mathbb{P}\mathcal{S})^{\times 2}$ is the same as the codimension in $\mathcal{S}^{\times 2}$ of the affine cone over $\Sigma_{\mathbb{P}}$. The codimension of the variety where these Jacobian matrices are not injective is the number we need to compute. This variety is given by the vanishing of all maximal minors.

Let $s = n_1 + n_2 + n_3 - 2 = \dim \mathcal{S}$. Computing all $\binom{n_1n_2n_3}{2s}$ maximal minors of a Jacobian matrix J is too expensive. Instead we proceed as follows. Note that we can perform all computations over $\mathbb{Q}$, because the Jacobian matrix is given by polynomials with integer coefficients. By homogeneity we can always assume that the first rank-1 tensor is $\mathcal{A}_1 = \mathbf{e}_1^1 \otimes \mathbf{e}_1^2 \otimes \mathbf{e}_1^3 \in \mathcal{S}$, where $\mathbf{e}_1^j \in \mathbb{Q}^{n_j}$ is the first standard basis vector. For each chart on the second copy of $\mathcal{S}$ we then take $\mathcal{A}_2 \in U_{i_2,j_2}$ and construct the Jacobian matrix J. We then multiply it with the column vector $\mathbf{k} = (k_1, k_2, \ldots, k_s) \in \mathbb{Q}^s \setminus \{0\}$ consisting of free variables; note that $\mathbb{Q}^s \setminus \{0\}$ should be covered by charts V_i for this. Now the condition number $\kappa(\mathcal{A}_1, \mathcal{A}_2) = \infty$ if $\mathbf{v} := J\mathbf{k}$ is zero, as then there would be a nontrivial kernel. It follows that the ideal generated by the maximal minors of J is then equal to the elimination ideal obtained by eliminating the k_i's from the ideal generated by the $n_1n_2n_3$ components of $\mathbf{v}$. This can be computed more efficiently in Macaulay2 than generating all maximal minors. The ideal thus obtained is the same ideal as the one that would have been obtained by performing all computations over $\mathbb{R}$, by the elementary properties of computing Gröbner bases (Cox *et al.*, 2015, Chapters 2–3). Performing this computation in all charts and taking the minimum of the computed codimensions, we found in all cases the value $n_3 - 1$. The code we used can be found at https://arxiv.org/abs/1801.01673. □

4. The condition number and distance to ill-posedness

In the course of establishing that the expected value of powers of the condition number can be infinite, i.e., Theorem 1.7, we relied on the unproved Theorem 1.6. The overall goal of this section is to prove Theorem 1.6. We start with a short detour and recall some results from Riemannian geometry.

4.1 *Isometric immersions*

Recall that a smooth map $f : \mathcal{M} \to \mathcal{N}$ between manifolds $\mathcal{M}, \mathcal{N}$ is called a *smooth immersion* if the derivative $\mathrm{d}_p f$ is injective for all $p \in \mathcal{M}$; see Lee (2013, Chapter 4). This requires $\dim \mathcal{M} \leqslant \dim \mathcal{N}$.

DEFINITION 4.1 A differentiable map $f : \mathcal{M} \to \mathcal{N}$ between Riemannian manifolds $(\mathcal{M}, g)$, $(\mathcal{N}, h)$ is called an *isometric immersion* if f is a smooth immersion and, furthermore, for all $p \in \mathcal{M}$ and $u, v \in \mathrm{T}_p\mathcal{M}$ it holds that $g_p(u, v) = h_{f(p)}(\mathrm{d}_p f(u), \mathrm{d}_p f(v))$. If in addition f is a diffeomorphism then it is called an *isometry*.

We will need the following lemma.

LEMMA 4.2 Let $\mathscr{M}, \mathscr{N}, \mathscr{N}$ be Riemannian manifolds and $f : \mathscr{M} \to \mathscr{N}$ and $g : \mathscr{N} \to \mathscr{N}$ be differentiable maps.

(1) Assume f is an isometry. Then $g \circ f$ is an isometric immersion if and only if g is an isometric immersion.

(2) Assume g is an isometry. Then $g \circ f$ is an isometric immersion if and only if f is an isometric immersion.

(3) If f is an isometric immersion, then for all $p, q \in \mathscr{M}$, we have $\operatorname{dist}_{\mathscr{M}}(p, q) \geqslant \operatorname{dist}_{\mathscr{N}}(f(p), f(q))$.

Proof. Let $p \in \mathscr{M}$. By the chain rule we have $\mathrm{d}_p(g \circ f) = \mathrm{d}_{f(p)} g\ \mathrm{d}_p f$. Hence, for all $\mathbf{u}, \mathbf{v} \in \mathrm{T}_p\mathscr{M}$ we have $\langle \mathrm{d}_p(g \circ f)\ \mathbf{u}, \mathrm{d}_p(g \circ f)\ \mathbf{v}\rangle = \langle \mathrm{d}_{f(p)} g\ \mathrm{d}_p f\ \mathbf{u}, \mathrm{d}_{f(p)} g\ \mathrm{d}_p f\ \mathbf{v}\rangle$. We prove (1): if g is isometric, the foregoing equation simplifies to $\langle \mathrm{d}_p(g \circ f)\ \mathbf{u}, \mathrm{d}_p(g \circ f)\ \mathbf{v}\rangle = \langle \mathrm{d}_p f\ \mathbf{u}, \mathrm{d}_p f\ \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$. Hence, $g \circ f$ is isometric. By the same argument, if $g \circ f$ is isometric, $g = g \circ f \circ f^{-1}$ is isometric. The second assertion is proved similarly. Finally, the last assertion is immediately clear from the definition of Riemannian distance. □

4.2 *Proof of Theorem* 1.6

In the introduction we recalled, in (1.6), that the condition number is equal to the inverse distance of the tuple of tangent spaces to the tuples of linear spaces not in general position. The idea to prove Theorem 1.6 is to make use of Lemma 4.1(3) from the previous subsection. This lemma lets us compare Riemannian distances between two manifolds. However, the projection distance from (1.8) is not given by some Riemannian metric on $\mathrm{Gr}(\Pi, n)$. In fact, up to scaling there is a unique orthogonally invariant metric on $\mathrm{Gr}(\Pi, n)$ when $\Pi > 4$; see Leichtweiss (1961). The usual choice of scaling is such that the distance associated to the metric is given by $d(V, W) = \sqrt{\theta_1^2 + \cdots + \theta_n^2}$, where $\theta_1, \ldots, \theta_n$ are the *principal angles* between V and W (Björck & Golub, 1973). Let us call this choice of metric the *standard metric* on $\mathrm{Gr}(\Pi, n)$. From this we construct the following distance function on $\mathrm{Gr}(\Pi, n)^{\times r}$:

$$\operatorname{dist}_{\mathrm{R}}((V_i)_{i=1}^r, (W_i)_{i=1}^r) := \sqrt{\sum_{i=1}^r d(V_i, W_i)^2}. \tag{4.1}$$

We can also express the projection distance in terms of the principal angles between the linear spaces V and W: $\|\mathrm{P}_V - \mathrm{P}_W\| = \max_{1 \leqslant i \leqslant n} |\sin \theta_i|$, where P_A denotes the projection onto the linear subspace A; see, e.g., Ye & Lim (2016, Table 2). Since, for all $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ we have $|\sin(\theta)| \leqslant |\theta|$, this shows that

$$\operatorname{dist}_{\mathrm{P}}((V_i)_{i=1}^r, (W_i)_{i=1}^r) \leqslant \operatorname{dist}_{\mathrm{R}}((V_i)_{i=1}^r, (W_i)_{i=1}^r). \tag{4.2}$$

This is an important inequality because it allows us to prove Theorem 1.6 by replacing $\operatorname{dist}_{\mathrm{P}}$ by $\operatorname{dist}_{\mathrm{R}}$. The second key result for the proof of Theorem 1.6 is the following.

PROPOSITION 4.3 We consider $\mathbb{P}\mathscr{S}$ to be endowed with the weighted metric from Definition 2.1 and $\mathrm{Gr}(\Pi, n)$ to be endowed with the standard metric. Then $\phi : \mathbb{P}\mathscr{S} \to \mathrm{Gr}(\Pi, n), [\mathscr{A}] \mapsto \mathrm{T}_{\mathscr{A}}\mathscr{S}$ is an isometric immersion in the sense of Definition 4.1.

REMARK 4.4 Note that ϕ in the proposition is *not* the Gauss map $\mathbb{P}\mathscr{S} \to \mathrm{Gr}(n-1, \mathbb{P}\mathbb{R}^{\Pi}), [\mathcal{A}] \mapsto \mathrm{T}_{[\mathcal{A}]}\mathbb{P}\mathscr{S}$, which maps a tensor to a *projective* subspace of $\mathbb{P}\mathbb{R}^{\Pi}$ of dimension $n-1 = \dim \mathbb{P}\mathscr{S}$.

Proposition 4.3 lies at the heart of this section, but its proof is quite technical and is therefore delayed until Appendix A below. First we use it to give a proof of Theorem 1.6.

Proof of Theorem 1.6. Assume that $\mathrm{Gr}(\Pi, n)^{\times r}$ is endowed with the standard metric on $\mathrm{Gr}(\Pi, n)$. Since ϕ is a isometric immersion, it follows from the definitions of the product metrics on the r-fold products of the smooth manifolds $\mathbb{P}\mathscr{S}$ and $\mathrm{Gr}(\Pi, n)$, respectively, that the r-fold product

$$\phi^{\times r} : (\mathbb{P}\mathscr{S})^{\times r} \to \mathrm{Gr}(\Pi, n)^{\times r},\ ([\mathcal{A}_1], \ldots, [\mathcal{A}_r]) \mapsto (\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \ldots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S})$$

is an isometric immersion. The associated distance on $\mathrm{Gr}(\Pi, n)^{\times r}$ is dist_R from (4.1). By Lemma 4.2(3) this implies

$$\mathrm{dist}_{\mathrm{w}}\big(([\mathcal{A}_1], \ldots, [\mathcal{A}_r]), \Sigma_{\mathbb{P}}\big) \geqslant \mathrm{dist}_{\mathrm{R}}\big((\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \ldots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S}), \phi^{\times r}(\Sigma_{\mathbb{P}})\big).$$

Recall from (1.7) the definition of Σ_{Gr} and note that $\phi^{\times r}(\Sigma_{\mathbb{P}}) \subset \Sigma_{\mathrm{Gr}}$ by construction. Consequently,

$$\mathrm{dist}_{\mathrm{w}}\big(([\mathcal{A}_1], \ldots, [\mathcal{A}_r]), \Sigma_{\mathbb{P}}\big) \geqslant \mathrm{dist}_{\mathrm{R}}\big((\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \ldots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S}), \Sigma_{\mathrm{Gr}}\big),$$

so that, by (4.2),

$$\mathrm{dist}_{\mathrm{w}}\big(([\mathcal{A}_1], \ldots, [\mathcal{A}_r]), \Sigma_{\mathbb{P}}\big) \geqslant \mathrm{dist}_{\mathrm{P}}\big((\mathrm{T}_{\mathcal{A}_1}\mathscr{S}, \ldots, \mathrm{T}_{\mathcal{A}_r}\mathscr{S}), \Sigma_{\mathrm{Gr}}\big).$$

By (1.6), the latter equals $\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)^{-1}$, which proves the assertion. □

5. Numerical experiments

In this section we perform a few numerical experiments in MATLAB R2017b for illustrating Theorems 1.6 and 1.7 and Corollary 1.4.

5.1 *Distance to ill-posedness*

To illustrate Theorem 1.6 we performed the following experiment with tensors in $\mathbb{R}^{11\times 10\times 5}$. Note that the generic rank in that space is 23. For each $2 \leqslant r \leqslant 5$ we select an ill-posed tensor decomposition $A := (\mathcal{A}_1, \ldots, \mathcal{A}_r) \in \mathscr{S}^{\times r}$ as explained next. First we sample a random rank-1 tuple $(\mathcal{A}_1, \ldots, \mathcal{A}_{r-1})$ in $\mathbb{R}^{11\times 10\times 5}$. Suppose that $\mathcal{A}_1 = \mathbf{a}_1^1 \otimes \mathbf{a}_1^2 \otimes \mathbf{a}_1^3$. Then we take $\mathcal{A}_r := \mathbf{a}_1^1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$, where the components of $\mathbf{x}_i$ are sampled from $N(0, 1)$. Now

$$\mathcal{A}_1 + \mathcal{A}_r = \mathbf{a}_i^1 \otimes (\mathbf{a}_i^2 \otimes \mathbf{a}_i^3 + \mathbf{x}_2 \otimes \mathbf{x}_3),$$

and since a rank-2 matrix decomposition is never unique, it follows from Landsberg (2012, Proposition 5.3.1.4) that $\mathcal{A}_1 + \mathcal{A}_r$ has a strictly positive-dimensional family of decompositions and, hence, so does $\mathcal{A}_1 + \cdots + \mathcal{A}_r$. Then it follows from Breiding & Vannieuwenhoven (2018c, Corollary 1.2) that $\kappa(A) = \infty$ and hence $A \in \Sigma_{\mathbb{P}}$. Finally we generate a neighbouring tensor decomposition $B := (\mathcal{B}_1, \ldots, \mathcal{B}_r) \in \mathscr{S}^{\times r}$

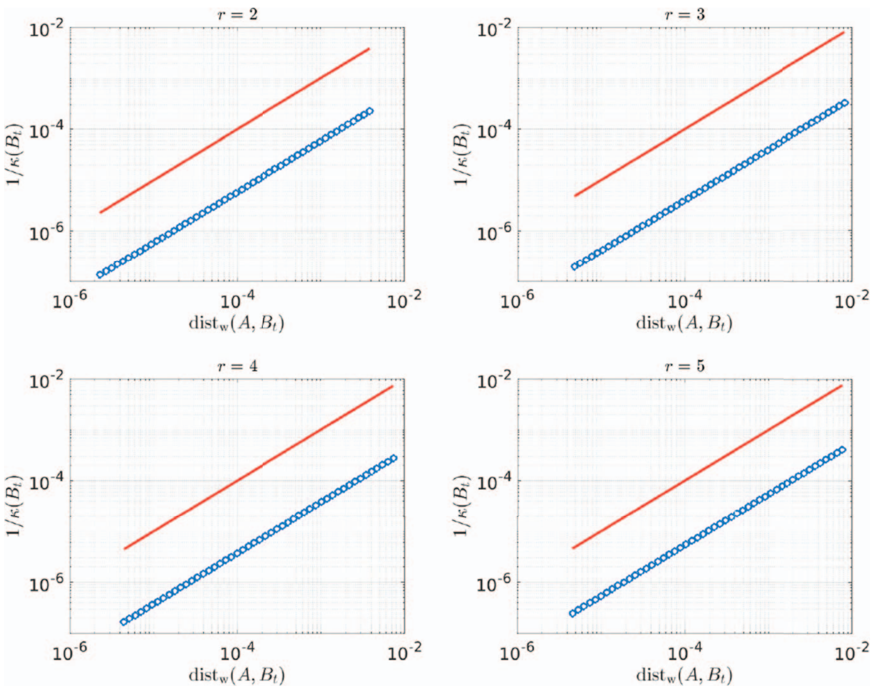


FIG. 2. The data points compare the inverse condition number and the estimate of the weighted distance to the locus of ill-posed CPDs for the tensors described in Section 5.1. The line above the data points illustrates where the data points would lie if the inequality in Theorem 1.6 were an equality. The gap between the line and the data points is thus a measure of the sharpness of the bound in Theorem 1.6.

by perturbing A as follows. Let $\mathcal{A}_i = \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \mathbf{a}_i^3$, and then we set $\mathcal{B}_i = (\mathbf{a}_i^1 + 10^{-2} \cdot \mathbf{x}_i^1) \otimes (\mathbf{a}_i^2 + 10^{-2} \cdot \mathbf{x}_i^2) \otimes (\mathbf{a}_i^3 + 10^{-2} \cdot \mathbf{x}_i^3)$, where the elements of $\mathbf{x}_i^k$ are randomly drawn from $N(0, 1)$.

Denote by $(0, 1) \to \mathscr{S}^{\times r}, t \mapsto B_t$ a curve between A and B whose length is $\mathrm{dist}_{\mathrm{w}}(A, B)$. Then, for all t, we have $\mathrm{dist}_{\mathrm{w}}(B_t, \Sigma_{\mathbb{P}}) \leqslant \mathrm{dist}_{\mathrm{w}}(A, B_t)$ and hence, by Theorem 1.6,

$$\frac{1}{\kappa(B_t)} \leqslant \mathrm{dist}_{\mathrm{w}}(A, B_t). \tag{5.1}$$

We expect for small t that $\mathrm{dist}_{\mathrm{w}}(A, B_t) \approx \mathrm{dist}_{\mathrm{w}}(B_t, \Sigma_{\mathbb{P}})$ and so (5.1) is a good substitute for the true inequality from Theorem 1.6.

The data points in the plots in Fig. 2 show, for each experiment, $\mathrm{dist}_{\mathrm{w}}(A, B_t)$ on the x-axis and $\frac{1}{\kappa(B_t)}$ on the y-axis. Since all the data points are below the red line, it is clearly visible that (5.1) holds. Moreover, since the data points (approximately) lie on a line parallel to the red line, the plots suggest, at least in the cases covered by the experiments, that for decompositions $A = (\mathcal{A}_1, \ldots, \mathcal{A}_r)$ close to $\Sigma_{\mathbb{P}}$ the reverse of Theorem 1.6 could hold as well, i.e., $\mathrm{dist}_{\mathrm{w}}(([\mathcal{A}_1], \ldots, [\mathcal{A}_r]), \Sigma_{\mathbb{P}}) \leqslant c \frac{1}{\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)}$, for some constant $c > 0$ that might depend on A. For completeness, in the experiments shown in Fig. 2, such a bound seems to hold for $c = 17, 25, 27, 19$, respectively, in the cases $r = 2, 3, 4, 5$.

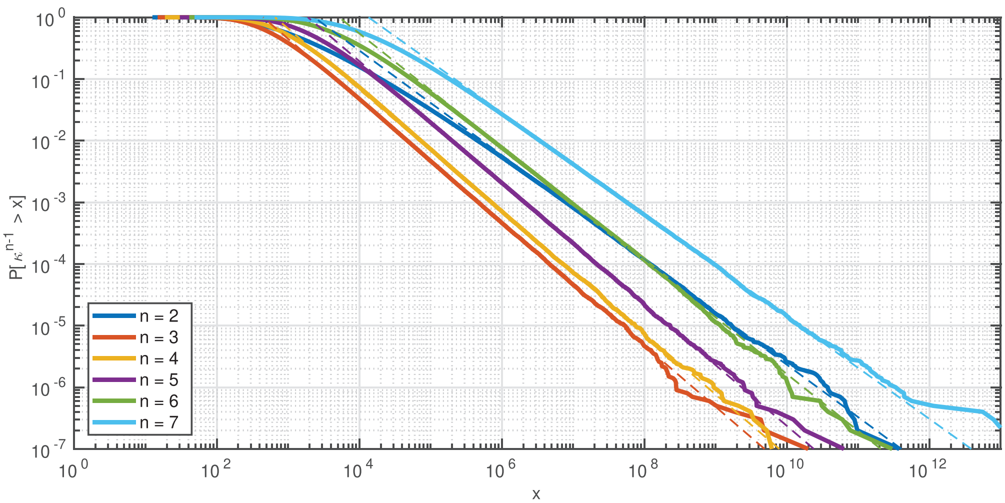


FIG. 3. A log–log plot of the empirical complementary cumulative distribution function of the $(n-1)$th power of the condition number of random rank-1 tuples $(\mathcal{A}_1, \ldots, \mathcal{A}_7)$ in the space $\mathbb{R}^{7\times 7\times n}$ for $n = 2, 3, \ldots, 7$, computed from 10^7 samples. The dashed lines represent approximations of the form $a_n x^{-b_n}$ of the empirical ccdf for $i = 2, 3, \ldots, 7$; the parameters (a_n, b_n) for each case are given in Table 1.

5.2 *Distribution of the condition number*

We perform Monte Carlo experiments to provide additional numerical evidence for Theorem 1.7 and Corollary 1.4. To this end we randomly sampled 10^7 random rank-1 tuples $(\mathcal{A}_1, \ldots, \mathcal{A}_7)$ in $\mathbb{R}^{7\times 7\times n}$, where $n = 2, 3, \ldots, 7$, and computed their condition numbers. We will abbreviate the random variable $\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_7)$ to κ from now onwards. These condition numbers are computed by constructing the $49n \times 7(12+n)$ block matrix $T = [U_i]_{i=1}^{7}$ from Breiding & Vannieuwenhoven (2018c, Theorem 1.1), where the individual blocks U_i are those from Breiding & Vannieuwenhoven (2018c, equation (5.1)), and then computing the inverse of the least (i.e., the $7(12+n)$th) singular value of T. The outcome of this experiment is summarized in Fig. 3, where we plot the complementary cumulative distribution function (ccdf) of the $(n-1)$th power of the condition number; recall that we know from Corollary 1.4 that $\mathbb{E}[\kappa^{n-1}] = \infty$.

It may appear at first glance that κ^{n-1} behaves very erratically near the tails of the ccdfs in Fig. 3. This phenomenon is entirely due to the sample error. Indeed, as we took 10^7 samples, this means that in the empirical ccdf, there are 10^k data points between $10^{-7} \leqslant \mathrm{P}[\kappa^{n-1} > x] \leqslant 10^{-7+k}$. For $k = 1$ or 2, the resulting sample error is visually evident.

It is particularly noteworthy that all of the ccdfs in Fig. 3 roughly appear to be shifted by a constant; the slopes of the curves look rather similar. In the figure, there are additional dashed lines that appear to capture the asymptotic behaviour of the ccdfs of κ^{n-1} quite well. These straight lines in the log–log plot correspond to a hypothesized model $a_n x^{-b_n}$ with $a_n, b_n \geqslant 0$. In Table 1 we give the (rounded) parameter values for these dashed lines in Fig. 3. By taking a log transformation, fitting the model becomes a linear least squares problem, which was solved exactly. To avoid overfitting we leave out the $9.9 \cdot 10^6$ smallest condition numbers, that is, all data above the horizontal line $\mathrm{P}[\kappa^{n-1} > x] = 10^{-2}$, as well as the 100 largest condition numbers, i.e., the data below the horizontal line $\mathrm{P}[\kappa^{n-1} > x] = 10^{-5}$. The motivation for this is as follows: the right tails of the ccdfs are corrupted by sampling errors, while for the left tails

TABLE 1 *Parameters (n, a_n, b_n) of the model $a_n x^{-b_n}$ fitted to the empirical cumulative distribution function described in Fig. 3. The row R^2 reports the coefficient of determination of the linear regression model* $\log(a_n) - b_n \log(x)$ *on the log-transformed empirical data; $R^2 = 1$ means the model perfectly predicts the data*

n	2	3	4	5	6	7
a_n	2328.45	447.54	656.27	1902.08	5210.73	13485.19
b_n	1.17713	1.00514	1.01091	1.01415	1.08573	1.20828
R^2	0.99994	0.99987	0.99975	0.99988	0.99940	0.99972

the model is clearly not valid. Based on Theorem 1.6 we are convinced that the hypothesized model is the correct one for large condition numbers. Theorem 1.6, shows that a small distance from the ill-posed locus $\Sigma_{\mathbb{P}}$ the condition number grows at least like 1 over the distance, and the experiments from Section 5.1, which show that close to the ill-posed locus the growth of the condition number appears also to be bounded by a constant times the inverse distance to $\Sigma_{\mathbb{P}}$. In other words, close to $\Sigma_{\mathbb{P}}$, the condition number behaves, as determined experimentally, asymptotically as $\kappa(A) = \mathscr{O}\big((\mathrm{dist}_w(A, \Sigma_{\mathbb{P}}))^{-1}\big)$.

From the above discussion we can conclude that for sufficiently large x, say $x \geqslant \kappa_0$, the true cdf of κ^{n-1}, i.e., $F(x) = \mathrm{P}[\kappa^{n-1} \leqslant x] = 1 - \mathrm{P}[\kappa^{n-1} \geqslant x]$ is very well approximated by $1 - a_n x^{-b_n} = \widetilde{F}(x)$. We can now employ the estimated cdfs to estimate the expected value of the kth power of the condition number κ in the unknown cases $n = 3, 4, \ldots, 7$ and $1 \leqslant k \leqslant n-2$. We are unable to compute these cases analytically because, first, we do not know whether the codimension of $\overline{\Sigma_{\mathbb{P}}}$ is 1, and, secondly, the techniques in this paper can prove only lower bounds on the condition number. We compute

$$\mathbb{E}[\kappa^k] = \mathbb{E}[(\kappa^{n-1})^{\frac{k}{n-1}}] = \int_0^\infty x^{\frac{k}{n-1}}\,\mathrm{d}F(x) = C + \int_{\kappa_0}^\infty x^{\frac{k}{n-1}} F'(x)\,\mathrm{d}x \approx C' + \int_{\kappa_0}^\infty x^{\frac{k}{n-1}} \widetilde{F}'(x)\,\mathrm{d}x,$$

where in the last step we assume that the error term $E(x) = F'(x) - \widetilde{F}'(x)$ integrated against $x^{\frac{k}{n-1}}$ is at most a constant; this requires that the hypothesized model is asymptotically correct as $x \to \infty$, which seems reasonable based on the above experiments. So it follows that

$$\mathbb{E}[\kappa^k] \approx C' + \int_{\kappa_0}^\infty a_n b_n x^{-b_n - 1 + \frac{k}{n-1}}\,\mathrm{d}x.$$

Note that the critical value for obtaining a finite integral is $k < (n-1)b_n$. Incidentally, the integral computed from the hypothesized model is finite for $n = 2$, as $1 < 1.17713$, but we attribute this 17% error in b_n to the sample variance, as we proved in Corollary 1.4 that the true integral is infinity. For $n \geqslant 3$, all of the hypothesized integrals with $1 \leqslant k \leqslant n-2$ integrate to constants; the computed values b_n would have to be off by 27% before the case $n = 5$ with $k = 3$ integrates to infinity. This provides some indications that the expected value of the condition number κ will be *finite* for $n_1 \times n_2 \times n_3$ tensors, provided that all $n_i \geqslant 3$. It is therefore unlikely that Corollary 1.4 may be improved by the techniques considered in this paper.

6. Conclusions

We presented a technique for establishing whether the average condition number of CPDs is infinite, namely Theorem 1.7. This is based on the partial condition number theorem, Theorem 1.6, which bounds the inverse condition number by the distance to the locus of ill-posed CPDs. Using this strategy we showed that the average of powers of the condition numbers of random rank-1 tuples of length r can be infinite in Corollary 1.4, depending on the codimension of the ill-posed locus. In particular, it was proved that the average condition number for $n_1 \times n_2 \times 2$ tensors is infinite. We are convinced that the inability to reduce the power in Corollary 1.4 to 1 for $n_1 \times n_2 \times n_3$ tensors with $2 \leqslant n_1, n_2, n_3 \leqslant 10$, as shown in Proposition 3.8, along with the numerical experiments in Section 5.2, are a strong indication that the average condition number is finite for tensors for which $n_1, n_2, n_3 \geqslant 3$.

The large gap in sensitivity between the case of $n_1 \times n_2 \times 2$ tensors and larger tensors has negative implications for the numerical stability of algorithms for computing CPDs based on a generalized eigendecomposition (such as those by Sands & Young, 1980; Lorber, 1985; Sanchez & Kowalski, 1990; Leurgans *et al.*, 1993), as is shown by Beltrán *et al.* (2019).

The strategy presented in this article cannot prove that the average condition number is finite. However, we believe that the main components of our approach can be adapted to prove upper bounds on the average condition number, provided that one can establish a local converse to Theorem 1.6.

Acknowledgements

The editor and reviewers are thanked for carefully checking the manuscript and providing many helpful suggestions for improvement. We thank C. Beltrán for pointing out Lemma 3.2 to us, so that we could use Theorem 1.6 to obtain Theorem 1.7. We wish to thank P. Bürgisser for carefully reading through the proof of Proposition 4.3. A. Seigal is thanked for discussions relating to Lemma 3.7, which she discovered independently. Some parts of this work are also part of the Ph.D. thesis Breiding (2017) of the first author.

Funding

Deutsche Forschungsgemeinschaft (DFG) research grant (BU 1371/2-2 to P.B.); Postdoctoral Fellowship of the Research Foundation–Flanders (FWO) (12E8116N and 12E8119N to N.V.)

References

Abo, H., Ottaviani, G. & Peterson, C. (2009) Induction for secant varieties of Segre varieties. *Trans. Amer. Math. Soc.*, **361**, 767–792.

Allman, E. S., Matias, C. & Rhodes, J. A. (2009) Identifiability of parameters in latent structure models with many observed variables. *Ann. Stat.*, **37**, 3099–3132.

Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M. & Telgarsky, M. (2014) Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.*, **15**, 2773–2832.

Beltrán, C., Breiding, P. & Vannieuwenhoven, N. (2019) Pencil-based algorithms for tensor rank decomposition are not stable. To appear in *SIAM J. Matrix Anal. Appl.* DOI: 10.1137/18M1200531.

Björck, A. & Golub, G. H. (1973) Numerical methods for computing angles between linear subspaces. *Math. Comput.*, **27**, 579–594.

Blum, L., Cucker, F., Shub, M. & Smale, S. (1998) *Complexity and Real Computation*. New York: Springer.

Bocci, C., Chiantini, L. & Ottaviani, G. (2014) Refined methods for the identifiability of tensors. *Ann. Mat. Pura Appl.*, **193**, 1691–1702.

Breiding, P. (2017) Numerical and statistical aspects of tensor decompositions. *Ph.D. Thesis*, TU Berlin. http://dx.doi.org/10.14279/depositonce-6148.

Breiding, P. & Vannieuwenhoven, N. (2018a) A Riemannian trust region method for the canonical tensor rank approximation problem. *SIAM J. Optim.*, **28**, 2435–2465.

Breiding, P. & Vannieuwenhoven, N. (2018b) Convergence analysis of Riemannian Gauss–Newton methods and its connection with the geometric condition number. *Appl. Math. Lett.*, **78**, 42–50.

Breiding, P. & Vannieuwenhoven, N. (2018c) The condition number of join decompositions. *SIAM J. Matrix Anal. Appl.*, **39**, 287–309.

Bürgisser, P., Clausen, M. & Shokrollahi, M. A. (1997) *Algebraic Complexity Theory*. Grundlehren der Mathematischen Wissenshaften, vol. 315. Berlin, Germany: Springer.

Bürgisser, P. & Cucker, F. (2013) *Condition: The Geometry of Numerical Algorithms*. Grundlehren der Mathematischen Wissenschaften, vol. 349. Berlin: Springer.

do Carmo, M. (1992) *Riemannian Geometry*. Basel: Birhäuser Boston.

Chiantini, L. & Ottaviani, G. (2012) On generic identifiability of 3-tensors of small rank. *SIAM J. Matrix Anal. Appl.*, **33**, 1018–1037.

Chiantini, L., Ottaviani, G. & Vannieuwenhoven, N. (2014) An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM J. Matrix Anal. Appl.*, **35**, 1265–1287.

Chiantini, L., Ottaviani, G. & Vannieuwenhoven, N. (2017) Effective criteria for specific identifiability of tensors and forms. *SIAM J. Matrix Anal. Appl.*, **38**, 656–681.

Comon, P. (1994) Independent component analysis, a new concept? *Signal Proc.*, **36**, 287–314.

Comon, P. & Jutten, C. (2010) *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Oxford: Elsevier.

Cox, D., Little, J. & O'Shea, D. (2015) *Ideals, Varieties, and Algorithms, 4th edn.* Undergraduate Texts in Mathematics. Berlin: Springer.

Cucker, F. (2016) Probabilistic analyses of condition numbers. *Acta Numer.*, **25**, 321–382.

Demmel, J. W. (1987a) The geometry of ill-conditioning. *J. Complexity*, **3**, 201–229.

Demmel, J. W. (1987b) On condition numbers and the distance to the nearest ill-posed problem. *Numer. Math.*, **51**, 251–289.

Domanov, I. & De Lathauwer, L. (2015) Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL. *SIAM J. Matrix Anal. Appl.*, **36**, 1567–1589.

Eckart, C. & Young, G. (1939) A principal axis transformation for non-Hermitian matrices. *Bull. Am. Math. Soc.*, **45**, 118–121.

Fuchs, D. B. (2004) Classical manifolds. *Topology II*, vol. 24. Berlin, Heidelberg: Springer, pp. 199–251.

Gelfand, I., Kapranov, M. M. & Zelevinsky, A. V. (1994) *Discriminants, Resultants and Multidimensional Determinants*. Modern Birkhäuser Classics. Boston, MA: Birkhäuser.

Grayson, D. & Stillman, M. (2018) *Macaulay2, a Software System for Research in Algebraic Geometry*. Available at www.math.uiuc.edu/Macaulay2. Accessed 16 November 2018.

Greub, W. H. (1978) *Multilinear Algebra*. New York: Springer.

Griffiths, P. (1974) On Cartan's method of Lie groups and moving frames as applied to uniqueness and existence questions in differential geometry. *Duke Math. J.*, **41**, 775–814.

Hackbusch, W. (2012) *Tensor Spaces and Numerical Tensor Calculus*. Springer Series in Computational Mathematics, vol. 42. Berlin: Springer.

Harris, J. (1992) *Algebraic Geometry, A First Course*. Graduate Text in Mathematics, vol. 133. New York: Springer.

Håstad, J. (1990) Tensor rank is NP-complete. *J. Algorithms*, **11**, 644–654.

Higham, N. J. (2002) *Accuracy and Stability of Numerical Algorithms*, 2nd edn. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Hitchcock, F. L. (1927) The expression of a tensor or a polyadic as a sum of products. *J. Math. Phys.*, **6**, 164–189.

Hough, D. (1977) Explaining and ameliorating the condition of zeros of polynomials. *Ph.D. Thesis*, University of California, Berkeley.

Kahan, W. (1966) Numerical linear algebra. *Canad. Math. Bull.*, **9**, 757–801.

Kroonenberg, P. M. (2008) *Applied Multiway Data Analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley.

Kruskal, J. B. (1977) Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.*, **18**, 95–138.
Landsberg, J. M. (2012) *Tensors: Geometry and Applications*. Graduate Studies in Mathematics, vol. 128. Providence, RI: AMS.
Lee, J. M. (2013) *Introduction to Smooth Manifolds, 2nd edn*. Graduate Texts in Mathematics, vol. 218. New York: Springer.
Leichtweiss, K. (1961) Zur riemannschen Geometrie in grassmannschen Mannigfaltigkeiten. *Math. Z.*, **76**, 334–366.
Leurgans, S. E., Ross, R. T. & Abel, R. B. (1993) A decomposition for three-way arrays. *SIAM J. Matrix Anal. Appl.*, **14**, 1064–1083.
Lorber, A. (1985) Features of quantifying chemical composition from two-dimensional data array by the rank annihilation factor analysis method. *Anal. Chem.*, **57**, 2395–2397.
McCullagh, P. (1987) *Tensor Methods in Statistics*. Monographs on Statistics and Applied Probability. New York: Chapman and Hall.
Qi, Y., Comon, P. & Lim, L.-H. (2016) Semialgebraic geometry of nonnegative tensor rank. *SIAM J. Matrix Anal. Appl.*, **37**, 1556–1580.
Rice, J. R. (1966) A theory of condition. *SIAM J. Numer. Anal.*, **3**, 287–310.
Sanchez, E. & Kowalski, B. R. (1990) Tensorial resolution: a direct trilinear decomposition. *J. Chemom.*, **4**, 29–45.
Sands, R. & Young, F. W. (1980) Component models for three-way data: an alternating least squares algorithm with optimal scaling features. *Psychometrika*, **45**, 39–67.
Sidiropoulos, N. D. & Bro, R. (2000) On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemom.*, **14**, 229–239.
de Silva, V. & Lim, L.-H. (2008) Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, **30**, 1084–1127.
Smilde, A., Bro, R. & Geladi, P. (2004) *Multi-way Analysis: Applications in the Chemical Sciences*. Hoboken, NJ: John Wiley.
Spivak, M. (1965) *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calcalus*. Boca Raton: CRC Press.
Strassen, V. (1983) Rank and optimal computation of generic tensors. *Linear Algebra Appl.*, **52–53**, 645–685.
Vannieuwenhoven, N. (2017) A condition number for the tensor rank decomposition. *Linear Algebra Appl.*, **535**, 35–86.
Wilkinson, J. H. (1972) Note on matrices with a very ill-conditioned eigenproblem. *Numer. Math.*, **19**, 176–178.
Ye, K. & Lim, L.-H. (2016) Schubert varieties and distances between subspaces of different dimensions. *SIAM J. Matrix Anal. Appl.*, **37**, 1176–1197.

A. Proof of Proposition 4.3

In this section we prove Proposition 4.3 to complete our study. We abbreviate $\mathbb{P}^{m-1} := \mathbb{P}(\mathbb{R}^m)$ in the following. Consider the following commutative diagram:

$$\begin{tikzcd}
\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1} \arrow[r, "\sigma"] \arrow[d, "\psi:=\iota\circ\phi\circ\sigma"'] & \mathbb{P}\mathscr{S} \arrow[d, "\phi"] \\
\mathbb{P}(\wedge^n \mathbb{R}^{\Pi}) & \mathrm{Gr}(\Pi, n) \arrow[l, hook', "\iota"]
\end{tikzcd}.$$

Herein, σ as defined in (2.2) is an isometry by the definition, ϕ is defined as in the statement of the proposition and ι is the Plücker embedding (Gelfand *et al.*, 1994, Chapter 3.1.), which maps into the space of *alternating tensors* $\mathbb{P}(\wedge^n \mathbb{R}^{\Pi})$. Recall from Landsberg (2012, Section 2.6) that alternating tensors are linear combinations of alternating rank-1 tensors like

$$\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_d := \frac{1}{d!} \sum_{\pi \in \mathfrak{S}_d} \operatorname{sgn}(\pi) \mathbf{x}_{\pi_1} \otimes \mathbf{x}_{\pi_2} \otimes \cdots \otimes \mathbf{x}_{\pi_d},$$

where $\mathfrak{S}_d$ is the permutation group on $\{1, \ldots, d\}$.

The image of the Plücker embedding $\mathscr{P} := \iota(\mathrm{Gr}(\Pi, n)) \subset \mathbb{P}\left(\wedge^n \mathbb{R}^\Pi\right)$ is a smooth variety called the *Plücker variety*. The Fubini–Study metric on $\mathbb{P}\left(\wedge^n \mathbb{R}^\Pi\right)$ makes $\mathscr{P}$ a Riemannian manifold. The Plücker embedding is an isometry; see, e.g., Griffiths (1974, Section 2) or Fuchs (2004, Chapter 3, Section 1.3).

Since σ and ι are isometries, it follows from Lemma 4.2 that ϕ is an isometric immersion if and only if $\psi := \iota \circ \phi \circ \sigma$ is an isometric immersion. We proceed by proving the latter. According to Definition 4.1, we have to prove that for all $p \in \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}$ and for all $x, y \in \mathrm{T}_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$ we have

$$\langle x, y \rangle_{\mathrm{w}} = \langle (\mathrm{d}_p \psi)(x), (\mathrm{d}_p \psi)(y) \rangle.$$

However, the equality $2\langle x, y \rangle = \langle x + y, x + y \rangle - \langle x, x \rangle - \langle y, y \rangle$ shows that it suffices to prove

$$\text{for all } p \in \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}, \text{ for all } x \in \mathrm{T}_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}), \langle x, x \rangle_{\mathrm{w}} = \langle (\mathrm{d}_p \psi)(x), (\mathrm{d}_p \psi)(x) \rangle. \tag{A.1}$$

To show this, let $p \in \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}$ and $x \in \mathrm{T}_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$ be fixed and consider any smooth curve $\Gamma : (-1, 1) \to \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}$ with $\Gamma(0) = p$ and $\Gamma'(0) = x$. The action of the differential is computed as follows according to Lee (2013, Corollary 3.25):

$$(\mathrm{d}_p \psi)(x) = \mathrm{d}_0(\psi \circ \Gamma).$$

We compute the right-hand side of that equation. However, before taking derivatives, we first compute an expression for $(\psi \circ \Gamma)(t)$.

Because $\mathrm{T}_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}) = \mathrm{T}_{p_1}\mathbb{P}^{n_1-1} \times \cdots \times \mathrm{T}_{p_d}\mathbb{P}^{n_d-1}$, we can write $x = (x_1, \ldots, x_d)$ with $x_i \in \mathrm{T}_{p_i}\mathbb{P}^{n_i-1}$. For each i we denote by $\mathbf{a}_i \in \mathbb{S}(\mathbb{R}^{n_i})$ a unit-norm representative for p_i, i.e., $p_i = [\mathbf{a}_i]$ with $\|\mathbf{a}_i\| = 1$ in the Euclidean norm. Letting $\mathbf{a}_i^\perp = \left\{\mathbf{u} \in \mathbb{R}^{n_i} \mid \langle \mathbf{u}, \mathbf{a}_i \rangle = 0\right\}$ denote the orthogonal complement of $\mathbf{a}_i$ in $\mathbb{R}^{n_i}$, we can then identify $\mathbf{a}_i^\perp = \mathrm{T}_{p_i}\mathbb{P}^{n_i-1}$ by (2.3). Moreover, because $\mathbf{a}_i$ is of unit norm, the Fubini–Study metric on $\mathrm{T}_{p_i}\mathbb{P}^{n_i-1}$ is given by the Euclidean inner product on the linear subspace $\mathbf{a}_i^\perp$. Now let $\mathbf{x}_i$ denote the unique vector in $\mathbf{a}_i^\perp$ corresponding to x_i. The sphere $\mathbb{S}(\mathbb{R}^{n_i})$ is a smooth manifold, so we find a curve $\gamma_i : (-1, 1) \to \mathbb{S}(\mathbb{R}^{n_i})$ with $\gamma_i(0) = \mathbf{a}_i$ and $\gamma_i'(0) = \mathbf{x}_i$. Without loss of generality we assume that each γ_i is the exponential map (Lee, 2013, Chapter 20). We claim that we can write the curve Γ as $\Gamma(t) = (\pi_1 \circ \gamma_1(t), \ldots, \pi_d \circ \gamma_d(t))$, where $\pi_i : \mathbb{S}(\mathbb{R}^{n_i}) \to \mathbb{P}^{n_i-1}$ is the canonical projection. Indeed, we have $\Gamma(0) = ([\mathbf{a}_1], \ldots, [\mathbf{a}_d]) = p$ and

$$\begin{aligned}\Gamma'(0) = \left((\pi_1 \circ \gamma_1)'(0), \ldots, (\pi_d \circ \gamma_d)'(0)\right) &= \left(\mathrm{P}_{(\mathbf{a}_1^\perp)}\gamma_1'(0), \ldots, \mathrm{P}_{(\mathbf{a}_d^\perp)}\gamma_d'(0)\right) \\ &= \left(\mathrm{P}_{(\mathbf{a}_1^\perp)}\mathbf{x}_1, \ldots, \mathrm{P}_{(\mathbf{a}_d^\perp)}\mathbf{x}_d\right) = \left(\mathbf{x}_1, \ldots, \mathbf{x}_d\right) = x,\end{aligned}$$

where P_A denotes the orthogonal projection onto the linear space A, where the second equality is due to Bürgisser & Cucker (2013, Lemma 14.8) and where the last step is due to the identification $\mathbf{a}_i^\perp \simeq \mathrm{T}_{p_i}\mathbb{P}^{n_i-1}$. This shows $(\psi \circ \Gamma)(t) = \psi(\pi_1 \circ \gamma_1(t), \ldots, \pi_d \circ \gamma_d(t))$. Recall that $\psi = \iota \circ \phi \circ \sigma$ and that

$$(\phi \circ \sigma \circ \Gamma)(t) = \mathrm{T}_{\gamma_1(t) \otimes \cdots \otimes \gamma_d(t)}\, \mathcal{S}.$$

Hence, $(\psi \circ \Gamma)(t) = \iota(\mathrm{T}_{\gamma_1(t)\otimes\cdots\otimes\gamma_d(t)}\mathcal{S})$. To compute the latter we must give a basis for the tangent space $\mathrm{T}_{\gamma_1(t)\otimes\cdots\otimes\gamma_d(t)}\mathcal{S}$. To do so, let us denote by $\{\mathbf{u}_1^i(t), \mathbf{u}_2^i(t), \ldots, \mathbf{u}_{n_i-1}^i(t)\}$ an orthonormal basis for the orthogonal complement of $\gamma_i(t)$; such a moving orthonormal basis is called an *orthonormal frame*. Then, by Landsberg (2012, Section 4.6.2), a basis for $\mathrm{T}_{\gamma_1(t)\otimes\cdots\otimes\gamma_d(t)}\mathcal{S}$ is given by

$$\mathscr{B}(t) = \{\mathcal{A}(t)\} \cup \left\{\mathcal{A}_{(i,j)}(t) \mid 1 \leqslant i \leqslant d, 1 \leqslant j \leqslant n_i - 1\right\},$$

where

$$\mathcal{A}(t) := \gamma_1(t) \otimes \cdots \otimes \gamma_d(t) \text{ and } \mathcal{A}_{(i,j)}(t) = \gamma_1(t) \otimes \cdots \otimes \gamma_{i-1}(t) \otimes \mathbf{u}_j^i(t) \otimes \gamma_{i+1}(t) \otimes \cdots \otimes \gamma_d(t). \tag{A.2}$$

If we let π denote the canonical projection $\pi : \wedge^n\mathbb{R}^\Pi \to \mathbb{P}\left(\wedge^n\mathbb{R}^\Pi\right)$, then we find

$$(\psi \circ \Gamma)(t) = \iota(\operatorname{span}\mathscr{B}(t)) = \pi\left(\mathcal{A}(t) \wedge \left(\bigwedge_{i=1}^{d}\bigwedge_{j=1}^{n_i-1}\mathcal{A}_{(i,j)}(t)\right)\right); \tag{A.3}$$

see Gelfand *et al.* (1994, Chapter 3.1.C). Note in particular that the right-hand side of (A.3) is independent of the specific choice of the orthonormal bases $\mathscr{B}(t)$, because the exterior product of another basis is just a scalar multiple of the basis we chose; below we make a specific choice of $\mathscr{B}(t)$ that simplifies subsequent computations. In the following let

$$\mathbf{g}(t) := \mathcal{A}(t) \wedge \left(\bigwedge_{i=1}^{d}\bigwedge_{j=1}^{n_i-1}\mathcal{A}_{(i,j)}(t)\right).$$

We are now prepared to compute the derivative of $(\psi \circ \Gamma)(t) = (\pi \circ \mathbf{g})(t) = [\mathbf{g}(t)]$. According to Bürgisser & Cucker (2013, Lemma 14.8) we have

$$\mathrm{d}_0(\psi \circ \Gamma) = \mathrm{P}_{(\mathbf{g}(0))^\perp}\frac{\mathbf{g}'(0)}{\|\mathbf{g}(0)\|}.$$

We will first prove that $\|\mathbf{g}(t)\| = 1$, which entails that $\mathbf{g}(t) \subset \mathbb{S}(\wedge^n\mathbb{R}^\Pi)$ so that

$$\mathrm{d}_0(\psi \circ \Gamma) = \mathrm{P}_{(\mathbf{g}(0))^\perp}\mathbf{g}'(0) = \mathbf{g}'(0) = \mathrm{d}_0\mathbf{g},$$

as $\mathbf{g}'(t)$ would in this case be contained in the tangent space to the sphere over $\wedge^n\mathbb{R}^\Pi$. We now need the following standard result.

LEMMA A1 We have the following:

(1) For $1 \leqslant k \leqslant d$, let $\mathbf{x}_k, \mathbf{y}_k \in \mathbb{R}^{n_k}$, and let $\langle\cdot,\cdot\rangle$ denote the standard Euclidean inner product. Then, the inner product of rank-1 tensors satisfies $\langle\mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_d, \mathbf{y}_1 \otimes \cdots \otimes \mathbf{y}_d\rangle = \prod_{j=1}^d\langle\mathbf{x}_j, \mathbf{y}_j\rangle$.

(2) Let $\mathbf{x}_1, \ldots, \mathbf{x}_d, \mathbf{y}_1, \ldots, \mathbf{y}_d \in \mathbb{R}^m$. Let $\langle\cdot,\cdot\rangle$ be the standard Euclidean inner product. Then, the inner product of skew-symmetric rank-1 tensors satisfies $\langle\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_d, \mathbf{y}_1 \wedge \cdots \wedge \mathbf{y}_d\rangle = \det\big([\langle\mathbf{x}_i, \mathbf{y}_j\rangle]_{i,j=1}^d\big)$.

(3) Whenever $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$ is a linearly dependent set we have $\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_d = 0$.

Proof. For the first point see, e.g., Hackbusch (2012, Section 4.5). For the second see, e.g., Greub (1978, Section 4.8) or Lee (2013, Proposition 14.11). The third is a consequence of the second point. □

Using the computation rules for inner products from Lemma A1 we find

$$\langle \mathcal{A}(t), \mathcal{A}(t)\rangle = \prod_{i=1}^{d} \langle \gamma_i(t), \gamma_i(t)\rangle = 1, \tag{A.4}$$

$$\langle \mathcal{A}(t), \mathcal{A}_{(i,j)}(t)\rangle = \langle \gamma_i(t), \mathbf{u}_j^i(t)\rangle \prod_{k\neq i} \langle \gamma_k(t), \gamma_k(t)\rangle = 0, \tag{A.5}$$

$$\langle \mathcal{A}_{(i,j)}(t), \mathcal{A}_{(k,\ell)}(t)\rangle = \begin{cases} 1 & \text{if } (i,j) = (k,\ell), \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

In other words, $\mathscr{B}(t)$ is an *orthonormal basis* for $\mathrm{T}_{\mathcal{A}(t)}\mathcal{S} = \mathrm{T}_{\gamma_1(t)\otimes\cdots\otimes\gamma_d(t)}\mathcal{S}$. By Lemma A1 we have

$$\langle \mathbf{g}(t), \mathbf{g}(t)\rangle = \det \begin{bmatrix} \langle \mathcal{A}(t), \mathcal{A}(t)\rangle & \langle \mathcal{A}(t), \mathcal{A}_{(1,1)}(t)\rangle & \cdots & \langle \mathcal{A}(t), \mathcal{A}_{(d,n_d)}(t)\rangle \\ \langle \mathcal{A}_{(1,1)}(t), \mathcal{A}(t)\rangle & \langle \mathcal{A}_{(1,1)}(t), \mathcal{A}_{(1,1)}(t)\rangle & \cdots & \langle \mathcal{A}_{(1,1)}(t), \mathcal{A}_{(d,n_d)}(t)\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathcal{A}_{(d,n_d)}(t), \mathcal{A}(t)\rangle & \langle \mathcal{A}_{(d,n_d)}(t), \mathcal{A}_{(1,1)}(t)\rangle & \cdots & \langle \mathcal{A}_{(d,n_d)}(t), \mathcal{A}_{(d,n_d)}(t)\rangle \end{bmatrix},$$

which equals $\det I_n = 1$.

It now only remains to compute $\mathrm{d}_0\mathbf{g}$. For this we have the following result.

LEMMA A2 Let

$$\mathcal{A} := \mathcal{A}(0) = \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_d \text{ and } \mathcal{A}_{(i,j)} := \mathcal{A}_{(i,j)}(0) = \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes \mathbf{u}_j^i \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d$$

and write

$$\mathbf{f}_{(i,j)} := \mathcal{A} \wedge \mathcal{A}_{(1,1)} \wedge \cdots \wedge \mathcal{A}_{(i,j-1)} \wedge \mathcal{A}'_{(i,j)}(0) \wedge \mathcal{A}_{(i,j+1)} \wedge \cdots \wedge \mathcal{A}_{(p,n_d-1)}.$$

The differential satisfies $\mathrm{d}_0\mathbf{g} = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \mathbf{f}_{(i,j)}$, where $\left\langle \mathbf{f}_{(i,j)}, \mathbf{f}_{(k,\ell)}\right\rangle = \delta_{ik}\delta_{j\ell}\sum_{1\leqslant\lambda\neq i\leqslant d}\langle \mathbf{x}_\lambda, \mathbf{x}_\lambda\rangle$, where δ_{ij} is the Kronecker delta, which is 1 if $i = j$ and 0 otherwise.

We prove this lemma at the end of this section. We can now prove (A.1). From Lemma A2 we find

$$\langle (\mathrm{d}_p\psi)(x), (\mathrm{d}_p\psi)(x)\rangle = \langle \mathrm{d}_0\mathbf{g}, \mathrm{d}_0\mathbf{g}\rangle = \left\langle \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \mathbf{f}_{(i,j)}, \sum_{k=1}^{d}\sum_{\ell=1}^{n_k-1} \mathbf{f}_{(k,\ell)} \right\rangle = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1}\sum_{1\leqslant\lambda\neq i\leqslant d} \langle \mathbf{x}_\lambda, \mathbf{x}_\lambda\rangle.$$

Reordering the terms, one finds

$$\langle (\mathrm{d}_p\psi)(x), (\mathrm{d}_p\psi)(x)\rangle = \sum_{i=1}^{d} \langle \mathbf{x}_i, \mathbf{x}_i\rangle \sum_{1\leqslant\lambda\neq i\leqslant d}\sum_{j=1}^{n_\lambda-1} 1 = \sum_{i=1}^{d} \langle \mathbf{x}_i, \mathbf{x}_i\rangle \cdot (n - n_i) = \langle \mathbf{x}, \mathbf{x}\rangle_{\mathrm{w}},$$

where the penultimate equality follows from the formula $n = 1 + \sum_{i=1}^{d}(n_i - 1)$ in (2.1). This proves (A.1) so that ϕ is an isometric map.

Finally, (A.1) also entails that ϕ is an immersion. Indeed, for an immersion it is required that $\mathrm{d}_p\psi$ is injective. Suppose that this is false; then there is a nonzero $x \in \mathrm{T}_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$ with corresponding nonzero $\mathbf{x}$ such that

$$0 = \langle 0, 0\rangle = \langle (\mathrm{d}_p\psi)(x), (\mathrm{d}_p\psi)(x)\rangle = \langle \mathbf{x}, \mathbf{x}\rangle_{\mathrm{w}} > 0,$$

which is a contraction. Consequently, ϕ is an isometric immersion, concluding the proof. □

It remains to prove Lemma A2.

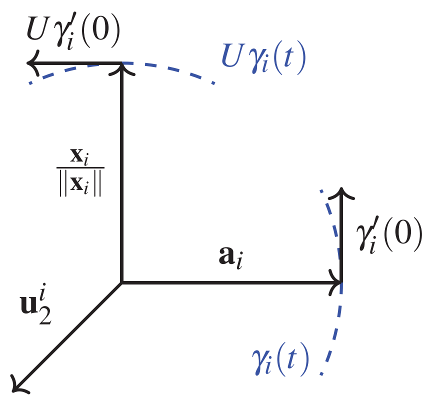


FIG. A1. A sketch of the orthonormal frame $\{\gamma_i(t), U\gamma_i(t), \mathbf{u}_2^i(t), \ldots, \mathbf{u}_{n_i-1}^i(t)\}$.

Proof of Lemma A.2. Recall that we have put $\mathbf{a}_i := \gamma_i(0) \in \mathbb{S}(\mathbb{R}^{n_i})$ and $\mathbf{x}_i := \gamma_i'(0) \in \mathrm{T}_{\mathbf{a}_i}\mathbb{S}(\mathbb{R}^{n_i})$ for $1 \leqslant i \leqslant d$. Without restriction we can assume that γ_i is contained in the great circle through $\mathbf{a}_i$ and $\mathbf{x}_i$. As argued above, we have the freedom of choice of an orthonormal basis of each $\gamma_i(t)^\perp$. To simplify computations we make the following choice.

For all i, let $\mathbf{u}_2^i, \ldots, \mathbf{u}_{n_i-1}^i$ be an orthonormal basis for $\mathbf{a}_i^\perp \cap \mathbf{x}_i^\perp$ and consider the orthogonal transformation U that rotates $\mathbf{a}_i$ to $\|\mathbf{x}_i\|^{-1}\mathbf{x}_i$, $\mathbf{x}_i$ to $-\|\mathbf{x}_i\|\mathbf{a}_i$ and leaves $\left\{\mathbf{u}_2^i, \ldots, \mathbf{u}_{n-1}^i\right\}$ fixed; cf. Fig. A1. Then we define the following curves (which except for the first one are all constant):

$$\mathbf{u}_1^i(t) := U\gamma_i(t), \quad \mathbf{u}_2^i(t) := \mathbf{u}_2^i, \quad \ldots \quad \mathbf{u}_{n_i-1}^i(t) := \mathbf{u}_{n_i-1}^i.$$

By construction $\{\mathbf{u}_1^i(t), \mathbf{u}_2^i(t), \ldots, \mathbf{u}_{n-1}^i(t)\}$ is an orthonormal basis for the orthogonal complement of $\gamma_i(t)$ for all t. We have

$$\mathrm{d}_0\mathbf{u}_1^i(t) = U\gamma_i'(0) = -\|\mathbf{x}_i\|\mathbf{a}_i, \quad \mathrm{d}_0\mathbf{u}_2^i(t) = \cdots = \mathrm{d}_0\mathbf{u}_{n_i-1}^i(t) = 0. \tag{A.7}$$

We will use this choice of orthonormal bases for the remainder of the proof. By the definition of $\mathbf{g}(t)$ and the product rule of differentiation, the first term of $\mathrm{d}_0\mathbf{g}$ is $\mathcal{A}'(0) \wedge \bigwedge_{i=1}^{d} \bigwedge_{j=1}^{n_i-1} \mathcal{A}_{(i,j)}$. We have

$$\mathcal{A}'(0) = \sum_{\lambda=1}^{d} \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{\lambda-1} \otimes \mathbf{x}_\lambda \otimes \mathbf{a}_{\lambda+1} \otimes \cdots \otimes \mathbf{a}_d = \sum_{\lambda=1}^{d} \|\mathbf{x}_\lambda\| \mathcal{A}_{(\lambda,1)}. \tag{A.8}$$

Hence, from the multilinearity of the exterior product it follows that the first term of $\mathrm{d}_0\mathbf{g}$ is

$$\sum_{\lambda=1}^{d} \|\mathbf{x}^\lambda\| \left(\mathcal{A}_{(\lambda,1)} \wedge \mathcal{A}_{(1,1)} \wedge \cdots \wedge \mathcal{A}_{(d,n_d-1)}\right) = \sum_{\lambda} 0 = 0.$$

This implies that all of the terms of $\mathrm{d}_0\mathbf{g}$ involve $\mathcal{A}_{(i,j)}'(0)$ for some (i,j). From (A.2) we find

$$\mathcal{A}_{(i,j)}'(0) = \sum_{\lambda=1}^{d} \mathcal{A}_{(i,j)}^{\lambda}, \tag{A.9}$$

where, using the shorthand notation $\mathbf{u}_j^i = \mathbf{u}_j^i(0)$, we have put

$$\mathcal{A}_{(i,j)}^{\lambda} := \begin{cases} \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{\lambda-1} \otimes \mathbf{x}_\lambda \otimes \mathbf{a}_{\lambda+1} \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes \mathbf{u}_j^i \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d & \text{if } \lambda \neq i, \\ \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes \mathrm{d}_0\mathbf{u}_j^i(t) \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d & \text{otherwise.} \end{cases}$$

Recall from (A.7) that $\mathrm{d}_0\mathbf{u}_1^i(t) = -\|\mathbf{x}_i\|\mathbf{a}_i$, while for $j > 1$ we have $\mathrm{d}_0\mathbf{u}_j^i(t) = 0$. Hence,

$$\mathcal{A}_{(i,j)}^{\lambda} := \begin{cases} \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{\lambda-1} \otimes \mathbf{x}_\lambda \otimes \mathbf{a}_{\lambda+1} \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes \mathbf{u}_j^i \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d & \text{if } \lambda \neq i, \\ \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes (-\|\mathbf{x}_i\|\mathbf{a}_i) \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d & \text{if } (\lambda, j) = (i, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Let $I_{i,j}$ be the ordered sequence $((1,1), \ldots, (d, n_d - 1))$ from which the tuple (i,j) was removed. Then

$$\begin{aligned} \mathbf{f}_{(i,j)} &= s_{(i,j)}\, \mathcal{A} \wedge \mathcal{A}'_{(i,j)}(0) \wedge \bigwedge_{(\alpha,\beta)\in I_{i,j}} \mathcal{A}_{(\alpha,\beta)} \\ &= s_{(i,j)} \sum_{1\leqslant\lambda\neq i\leqslant d} \mathcal{A} \wedge \mathcal{A}_{(i,j)}^{\lambda} \wedge \bigwedge_{(\alpha,\beta)\in I_{i,j}} \mathcal{A}_{(\alpha,\beta)} =: s_{(i,j)} \sum_{1\leqslant\lambda\neq i\leqslant d} \mathbf{f}_{(i,j)}^{\lambda}, \end{aligned} \tag{A.10}$$

where $s_{(i,j)} \in \{-1, 1\}$ is the sign of the permutation for moving $\mathcal{A}'_{(i,j)}(0)$ to the second position in the exterior product; note that in the second step we used (A.9) and the multilinearity of the exterior product. We continue by computing for $\lambda \neq i$ and $\mu \neq k$ the value

$$\langle \mathbf{f}_{(i,j)}^{\lambda}, \mathbf{f}_{(k,\ell)}^{\mu} \rangle = \det\bigl(B_{(i,j),\lambda}^{\mathrm{T}} B_{(k,\ell),\mu}\bigr), \quad \text{where } B_{(i,j),\lambda} := \Bigl[\mathcal{A} \ \ \mathcal{A}_{(i,j)}^{\lambda} \ \ [\mathcal{A}_{(\alpha,\beta)}]_{(\alpha,\beta)\in I_{i,j}}\Bigr];$$

herein, the column vectors should be interpreted as vectorized tensors. Recall that $\langle \mathbf{a}_i, \mathbf{x}_i \rangle = 0$ and that $\langle \mathbf{a}_i, \mathbf{u}_j^i \rangle = 0$ for all i, j. Then it follows from Lemma A1 and direct computations that for $\lambda \neq i$ and $\mu \neq k$, and for all $1 \leqslant \alpha \leqslant d$ and $1 \leqslant \beta < n_\alpha$, we have

$$\langle \mathcal{A}, \mathcal{A}_{(k,\ell)}^{\mu} \rangle = \langle \mathcal{A}, \mathcal{A}_{(k,\ell)} \rangle = 0, \quad \langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A}_{(k,\ell)}^{\mu} \rangle = \delta_{ik}\delta_{j\ell}\delta_{\lambda\mu}\|\mathbf{x}_\lambda\|^2 \quad \text{and} \quad \langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A}_{(\alpha,\beta)} \rangle = 0.$$

We distinguish between two cases. If $(i,j) \neq (k,\ell)$, $\lambda \neq i$ and $\mu \neq k$, it follows from the above equations that the row of $(B_{(i,j),\lambda})^T B_{(k,\ell),\mu}$ consisting of

$$\Bigl[\langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A} \rangle \ \ \langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A}_{(k,\ell)}^{\mu} \rangle \ \ [\langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A}_{(\alpha,\beta)} \rangle]_{(\alpha,\beta)\in I_{k,\ell}}\Bigr]$$

is a zero row, which implies that $\bigl\langle \mathbf{f}_{(i,j),\lambda}, \mathbf{f}_{(k,\ell),\mu} \bigr\rangle = 0$. On the other hand, if $(i,j) = (k,\ell)$, $\lambda \neq i$ and $\mu \neq k$, then it follows from the above equations that $B_{(i,j),\lambda}^T B_{(i,j),\mu}$ is a diagonal matrix, namely

$$B_{(i,j),\lambda}^{\mathrm{T}} B_{(i,j),\mu} = \operatorname{diag}(1, \langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A}_{(i,j)}^{\mu} \rangle, 1, \ldots, 1).$$

Its determinant is then $\langle \mathcal{A}_{(i,j)}^{\lambda}, \mathcal{A}_{(i,j)}^{\mu} \rangle = \delta_{\lambda\mu}\|\mathbf{x}_\lambda\|^2$. Therefore,

$$\langle \mathbf{f}_{(i,j)}^{\lambda}, \mathbf{f}_{(k,\ell)}^{\mu} \rangle = \delta_{ik}\delta_{j\ell}\delta_{\lambda\mu}\|\mathbf{x}_\lambda\|^2. \tag{A.11}$$

Finally we can compute $\langle \mathbf{f}_{(i,j)}, \mathbf{f}_{(k,\ell)} \rangle$. From (A.10),

$$\langle \mathbf{f}_{(i,j)}, \mathbf{f}_{(k,\ell)} \rangle = s_{(i,j)} s_{(k,\ell)} \left\langle \sum_{1\leqslant\lambda\neq i\leqslant d} \mathbf{f}^{\lambda}_{(i,j)}, \sum_{1\leqslant\mu\neq k\leqslant d} \mathbf{f}^{\mu}_{(k,\ell)} \right\rangle = s_{(i,j)} s_{(k,\ell)} \delta_{ik} \delta_{j\ell} \sum_{1\leqslant\lambda\neq i\leqslant d} \|\mathbf{x}_\lambda\|^2,$$

which is zero unless $(i,j) = (k,\ell)$. For $(i,j) = (k,\ell)$ we find

$$\|\mathbf{f}_{(i,j)}\|^2 = s^2_{(i,j)} \sum_{1\leqslant\lambda\neq i\leqslant d} \|\mathbf{x}_\lambda\|^2 = \sum_{1\leqslant\lambda\neq i\leqslant d} \|\mathbf{x}_\lambda\|^2,$$

proving the result. □