Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Towards a condition number theorem for the tensor rank decomposition

by

Paul Breiding and Nick Vannieuwenhoven

Preprint no.: 3 2018

TOWARDS A CONDITION NUMBER THEOREM FOR THE TENSOR RANK DECOMPOSITION

PAUL BREIDING AND NICK VANNIEUWENHOVEN

Abstract. We show that a natural weighted distance from a tensor rank decomposition to the locus of ill-posed decompositions (i.e., decompositions with unbounded geometric condition number, derived in [P. Breiding and N. Vannieuwenhoven, *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl. (2018)]) is bounded from below by the inverse of this condition number. That is, we prove one inequality towards a condition number theorem for the tensor rank decomposition. Numerical experiments suggest that the other inequality could also hold (at least locally).

1. Introduction

Whenever data depends on several variables, it may be stored as a d-array

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

For the purpose of our exposition, this d-array is informally called a *tensor*. Due to the curse of dimensionality, storing all this data in a tensor is neither feasible nor insightful. Fortunately, the data of interest often admit additional structure that can be exploited. One particular tensor decomposition that arises in several applications is the *tensor rank decomposition*, or *canonical polyadic decomposition* (CPD). It was proposed by Hitchcock [27] and it expresses a tensor $\mathfrak{A} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ as a minimum-length linear combination of *pure tensors*:

$$\mathfrak{A} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d, \quad \mathbf{a}_i^k \in \mathbb{R}^{n_k}, \tag{CPD}$$

where $\otimes$ is the tensor product:

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

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

In several applications, the CPD of a tensor reveals domain-specific information that is of interest, such as in psychometrics [30], chemical sciences [37], theoretical computer science [8], signal processing [13, 14, 36], statistics [2, 35] and machine learning [3, 36]. In most of these applications, the data that the tensor represents is corrupted by measurement errors, which will cause the CPD computed from the measured data to differ from the CPD of the true, uncorrupted data. For measuring the sensitivity of the CPD to perturbations in the data, the standard technique in numerical analysis consists of computing the *condition number* [9, 26] of

---

2010 *Mathematics Subject Classification.* Primary 49Q12, 53B20, 15A69; Secondary 14P10, 65F35, 14Q20.

Key words and phrases. tensor rank decomposition problem; condition number; distance to ill-posedness; ill-posed problems; CPD .

PB: Max-Planck-Institute for Mathematics in the Sciences Leipzig, breiding@mis.mpg.de. Partially supported by DFG research grant BU 1371/2-2.

NV: KU Leuven, Department of Computer Science. Supported by a Postdoctoral Fellowship of the Research Foundation–Flanders (FWO).

the CPD. Earlier theoretical work by the authors introduced two related condition numbers for the computational problem of computing a CPD from a given tensor; see [6, 38].

The topic of this paper is a further characterization of the geometric condition number of the CPD from [6][1] as an *inverse distance to ill-posedness*. The characterization of a condition number as an inverse distance to ill-posedness is called *condition number theorem* in the literature and it provides a geometric interpretation of complexity of a computational problem. Demmel [17] 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 [10, 18, 29], polynomial zero finding [18, 28] or computing eigenvalues [18, 39]. Sometimes a condition number is also defined as inverse distance to ill-posedness; e.g., for the problem of computing an optimal basis in linear programming [15, 16]. For a comprehensive overview see also [9, pages 10, 16, 125, 204].

However, an interpretation of condition numbers as inverse distance to ill-posedness is usually understood as a distance *in the data space*. On the contrary, the authors proved in [6] that the condition number for the CPD is equal to the distance to ill-posedness in an *auxiliary space*, which is a product of Grassmann manifolds. To be precise, recall that the set of pure tensors, or rank-1 tensors, is a smooth manifold, called the *Segre manifold*. We will denote it by $\mathcal{S} := \mathcal{S}_{n_1,\ldots,n_k} := \left\{\mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d \mid \mathbf{a}^k \in \mathbb{R}^{n_k} \setminus \{0\}\right\}$, assuming that $n_1, \ldots, n_k$ were fixed. According to [6, Theorem 1.3], the condition number of the CPD at a decomposition $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r) \in \mathcal{S}^{\times r}$ can then be expressed as the inverse distance of the tuple of tangent spaces $(\mathrm{T}_{\mathfrak{A}_1}\mathcal{S}, \ldots, \mathrm{T}_{\mathfrak{A}_r}\mathcal{S})$ to ill-posedness:

$$\kappa(\mathfrak{A}_1, \ldots, \mathfrak{A}_r) = \frac{1}{\mathrm{dist}_{\mathrm{P}}((\mathrm{T}_{\mathfrak{A}_1}\mathcal{S}, \ldots, \mathrm{T}_{\mathfrak{A}_r}\mathcal{S}), \Sigma_{\mathrm{Gr}})}, \tag{1.2}$$

where Σ_{Gr} and the distance $\mathrm{dist}_{\mathrm{P}}$ are defined as below.

From (1.2) we only see that the condition number tends to infinity as $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$ approaches an ill-posed decomposition, but do now know *how fast* this happens. In other words, the characterization (1.2) only gives a qualitative answer to the question

(1.3) *"If $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$ is close to an ill-posed decompositions, then what is $\kappa(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$?"*

In this article we make a first advance for giving a quantitative answer to question (1.3) by relating the condition number to a metric on the data space $\mathcal{S}^{\times r}$. Our main theorem is Theorem 1.1 below. Before we state it, though, we recall the definitions of Σ_{Gr} and $\mathrm{dist}_{\mathrm{P}}$ from [6].

Let $n := \dim \mathcal{S}$ and $\Pi := n_1 \cdots n_d$. Denote by $\mathrm{Gr}(\Pi, n)$ the Grassmann manifold of n-dimensional linear spaces in the space of tensors $\mathbb{R}^{n_1 \times \cdots \times n_d} \cong \mathbb{R}^{\Pi}$ and observe that the tangent space to $\mathcal{S}$ at the decomposition $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$ is $(\mathrm{T}_{\mathfrak{A}^i}\mathcal{S})_{i=1}^r \in \mathrm{Gr}(\Pi, n)^{\times r}$. The *projection distance* on $\mathrm{Gr}(\Pi, n)$ is given by $\|\pi_V - \pi_W\|$, where π_V and π_W are the orthogonal projections on the spaces V and W respectively, and $\|\cdot\|$ is the spectral norm. This distance measure is extended to $\mathrm{Gr}(\Pi, n)^{\times r}$ in the usual way:

$$\mathrm{dist}_{\mathrm{P}}((V_i)_{i=1}^r, (W_i)_{i=1}^r) := \Big(\sum_{i=1}^r \|\pi_{V_i} - \pi_{W_i}\|^2\Big)^{\frac{1}{2}}.$$

The set $\Sigma_{\mathrm{Gr}} \subset \mathrm{Gr}(\Pi, n)^{\times r}$ is defined as

$$\Sigma_{\mathrm{Gr}} := \left\{(W_1, \ldots, W_r) \in \mathrm{Gr}(\Pi, n)^{\times r} \mid \dim(W_1 + \cdots + W_r) < rn\right\}. \tag{1.4}$$

The decomposition $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$ whose corresponding tangent space lies in Σ_{Gr} is *ill-posed* in the following sense. It was shown in [6, Corollary 1.2] that whenever there is a smooth curve $\gamma(t) = (\mathfrak{A}_1(t), \ldots, \mathfrak{A}_r(t))$ such that $\mathfrak{A} = \sum_{i=1}^r \mathfrak{A}_i(t)$ is constant, even though $\gamma'(0) \neq 0$, then

[1]The topic of [6] are join decompositions, of which the CPD is a special case; see also [6, Section 7].

all of the decompositions $(\mathfrak{A}_1(t),\ldots,\mathfrak{A}_r(t))$ of $\mathfrak{A}$ are ill-posed decompositions. Note that in this case, the tensor $\mathfrak{A}$ thus has a family of decompositions running through $(\mathfrak{A}_1(0),\ldots,\mathfrak{A}_r(0))$. We say that $\mathfrak{A}$ is not *locally r-identifiable.* Since tensors are expected to admit only a finite number of decompositions generically when $r(1+\sum_{k=1}^d(n_k-1)) < \prod_{k=1}^d n_k$, see, e.g., [1, 5, 11, 12], 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 [6, Proposition 7.1] we have shown that the condition number is invariant under scaling of the rank-one tensors $\mathfrak{A}_i$. Hence, to describe the condition number as an inverse distance to ill-posedness on $\mathcal{S}^{\times r}$ we must consider some sort of angular distance. This is why the main theorem of this article (Theorem 1.1) is stated in projective space.

Theorem 1.1 (A condition number theorem for the CPD)**.** *Let $\Pi > 4$, and denote the canonical projection onto projective space by $\pi : \mathbb{R}^\Pi\backslash\{0\} \to \mathbb{P}(\mathbb{R}^\Pi)$. We put $\mathbb{P}\mathcal{S} := \pi(\mathcal{S})$ and for points $\mathfrak{A} \in \mathbb{R}^\Pi$ we denote $[\mathfrak{A}] := \pi(\mathfrak{A})$. Let $(\mathfrak{A}_1,\ldots,\mathfrak{A}_r) \in \mathcal{S}^{\times r}$. Then,*

$$\frac{1}{\kappa(\mathfrak{A}_1,\ldots,\mathfrak{A}_r)} = \mathrm{dist}_\mathrm{P}((\mathrm{T}_{\mathfrak{A}_1}\mathcal{S},\ldots,\mathrm{T}_{\mathfrak{A}_r}\mathcal{S}),\Sigma_{\mathrm{Gr}}) \le \mathrm{dist}_\mathrm{w}(([\mathfrak{A}_1],\ldots,[\mathfrak{A}_r]),\Sigma_\mathbb{P}),$$

where

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

and the distance dist_w *is defined in Definition 1.2 below.*

Remark. The experiments in Section 4 suggest that the reverse inequality of Theorem 1.1 could also be true (at least locally). In all of the experiments we find for decompositions $(\mathfrak{A}_1,\ldots,\mathfrak{A}_r)$ close to $\Sigma_\mathbb{P}$ that there is a constant $c>0$ such that $\mathrm{dist}_\mathrm{w}(([\mathfrak{A}_1],\ldots,[\mathfrak{A}_r]),\Sigma_\mathbb{P}) \le c\frac{1}{\kappa(\mathfrak{A}_1,\ldots,\mathfrak{A}_r)}$.

It is important to note that the lower bound in the theorem can be computed efficiently via the spectral characterization of the condition number $\kappa(\mathfrak{A}_1,\ldots,\mathfrak{A}_r)$ from [6, Theorem 1.1].

We prove Theorem 1.1 in Section 3. The weighted distance is introduced next.

Definition 1.2 (Weighted distance)**.** Let $\langle\cdot,\cdot\rangle$ denote the Fubini-Study metric on $\mathbb{P}(\mathbb{R}^{n_i})$ and let $d_\mathbb{P}$ be the corresponding distance on $\mathbb{P}(\mathbb{R}^{n_i})$; see, e.g., [9, Section 14.2.2]. The *weighted distance* between the points $p=(p_1,\ldots,p_d), q=(q_1\ldots,q_d) \in \mathbb{P}(\mathbb{R}^{n_1})\times\cdots\times\mathbb{P}(\mathbb{R}^{n_d})$ is defined

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

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

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

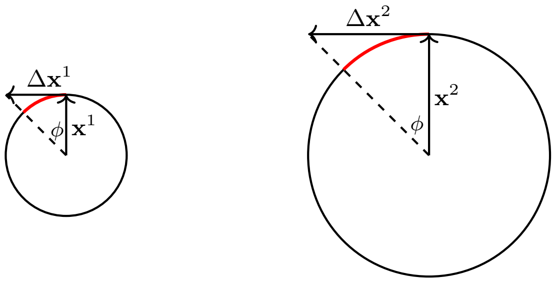
where σ^{-1} is the inverse of the projective Segre map

$$\sigma : \mathbb{P}(\mathbb{R}^{n_1})\times\cdots\times\mathbb{P}(\mathbb{R}^{n_d}) \to \mathbb{P}\mathcal{S}, ([\mathbf{a}^1],\ldots,[\mathbf{a}^d]) \mapsto [\mathbf{a}^1\otimes\cdots\otimes\mathbf{a}^d], \quad (1.5)$$

see [31, Section 4.3.4.].

Note that for $n_1 > n_2$ relative errors in the factor $\mathbb{P}(\mathbb{R}^{n_2})$ weigh more than relative errors in the factor $\mathbb{P}(\mathbb{R}^{n_1})$; this is illustrated in Figure 1.1.

The rest of this paper is structured as follows. In the next section, we recall some preliminary material on inner products of rank-1 tensors and rank-1 alternating tensors, as well as elementary results about Riemannian isometric immersions. Section 3 is entirely devoted to the proof of the main theorem, i.e., Theorem 1.1. In Section 4 we present numerical experiments.



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

FIGURE 1.1. The picture depicts relative errors in the weighted distance, where $\mathbf{x}^1 \in \mathbb{P}(\mathbb{R}^{n_1})$ and $\mathbf{x}^2 \in \mathbb{P}(\mathbb{R}^{n_2})$ with $n_1 > n_2$. The relative errors of the tangent directions $\Delta\mathbf{x}^1$ and $\Delta\mathbf{x}^2$ are both equal to $\tan\phi$, but the contribution to the weighted distance marked in red is larger for the large circle, which corresponds to the smaller projective space $\mathbb{P}(\mathbb{R}^{n_2})$.

Acknowledgements. We like to thank P. Bürgisser for carefully reading through the proof of Proposition 3.3. This work is part of the PhD thesis [7] of the first author .

2. Preliminaries

2.1. **Notation.** The real projective space of dimension $n-1$ is denoted $\mathbb{P}(\mathbb{R}^n)$ and the unit sphere of dimension $n-1$ is denoted $\mathbb{S}(\mathbb{R}^n)$.

Throughout this paper, n denotes the dimension of the (affine) Segre variety $\mathcal{S}$ [25,31], i.e.,

$$n := \dim\mathcal{S} = 1 - d + \sum_{i=1}^{d} n_i; \tag{2.1}$$

Letting $\gamma : (-1,1) \to M$ be a smooth curve in a manifold M, we will use the shorthand notations $\gamma'(0) := \frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\gamma(t)$ for the tangent vector in $\mathrm{T}_{\gamma(0)}M$ and $\gamma'(t) := \frac{\mathrm{d}}{\mathrm{d}t}\gamma(t)$.

2.2. **Inner products of rank-one tensors.** The following lemmas will be useful.

Lemma 2.1. *For $1 \le k \le d$, 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, $\langle \mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_d, \mathbf{y}_1 \otimes \cdots \otimes \mathbf{y}_d\rangle = \prod_{j=1}^d \langle \mathbf{x}_j, \mathbf{y}_j\rangle$.*

Proof. See, e.g., [24, Section 4.5]. □

Let $\mathfrak{S}_d$ be the permutation group on $1,\ldots,d$, and let $\mathrm{sgn}(\pi)$ denote the sign of the permutation $\pi \in \mathfrak{S}_d$. Recall that the exterior product on $\mathbb{R}^n$ can be defined as

$$\wedge : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \wedge^d\mathbb{R}^n, (\mathbf{x}_1,\ldots,\mathbf{x}_d) \mapsto \frac{1}{d!}\sum_{\pi\in\mathfrak{S}_d} \mathrm{sgn}(\pi)\mathbf{x}_{\pi_1}\otimes\mathbf{x}_{\pi_2}\otimes\cdots\otimes\mathbf{x}_{\pi_d};$$

in this definition, $\wedge^d\mathbb{R}^n \subset \mathbb{R}^n \otimes\cdots\otimes \mathbb{R}^n$ is to be interpreted as the linear subspace generated by the image of $\wedge$. The elements of $\wedge^d\mathbb{R}^n$ are then called *alternating tensors* [31, Section 2.6]. It is standard to use the shorthand $\mathbf{x}_1\wedge\cdots\wedge\mathbf{x}_d$ for $\wedge(\mathbf{x}_1,\ldots,\mathbf{x}_d)$. The next result is well known.

Lemma 2.2. *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 induced inner product on $\wedge^d\mathbb{R}^m$ satisfies $\langle \mathbf{x}_1\wedge\cdots\wedge\mathbf{x}_d, \mathbf{y}_1\wedge\cdots\wedge\mathbf{y}_d\rangle = \det\big([\langle\mathbf{x}_i,\mathbf{y}_j\rangle]_{i,j=1}^d\big)$.*

Proof. See, e.g., [22, Section 4.8] or [32, Proposition 14.11]. □

As a corollary of this result one immediately finds the standard fact that

$$\|\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_d\| = 0, \text{ and so } \mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_d = 0,$$

whenever $\{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$ is a linearly dependent set.

2.3. **Isometric immersions.** Let $(M, \langle \cdot, \cdot \rangle)$ be a Riemannian manifold. Recall that the *Riemannian distance* between two points $p, q \in M$ is defined as

$$\operatorname{dist}_M(p, q) = \inf \{l(\gamma) \mid \gamma(0) = p, \gamma(1) = q\}, \tag{2.2}$$

where the infimum is over all piecewise differentiable curves $\gamma : [0, 1] \to M$ and the length of a curve is defined as $l(\gamma) = \int_0^1 \langle \gamma'(t), \gamma'(t) \rangle^{\frac{1}{2}} \,\mathrm{d}t$. The distance dist_M makes the manifold M a metric space [19, Proposition 2.5].

Recall that a smooth map $f : M \to N$ between manifolds M, N is called a *smooth immersion* if the derivative $\mathrm{d}_p f$ is injective for all $p \in M$; see [32, Chapter 4]. Hence, $\dim M \leq \dim N$.

Definition 2.3. A differentiable map $f : M \to N$ between Riemannian manifolds $(M, g), (N, h)$ is called an *isometric immersion* if f is a smooth immersion and for all $p \in M$ and $u, v \in \mathrm{T}_p M$ it holds that $g_p(u, v) = h_{f(p)}(\mathrm{d}_p f(u), \mathrm{d}_p f(v))$. We also say that f is *isometric*. If in addition f is a diffeomorphism then it is called an *isometry*.

Note that if f is an isometry then $\dim M = \dim N$.

Lemma 2.4. *Let M, N, P be Riemannian manifolds and $f : M \to N$ and $g : N \to P$ be differentiable maps.*

1. *Assume that f is an isometry. Then, $g \circ f$ is isometric if and only if g is isometric.*
2. *Assume that g is an isometry. Then, $g \circ f$ is isometric if and only if f is isometric.*

Proof. Let $p \in M$. By the chain rule we have $\mathrm{d}_p(g \circ f) = \mathrm{d}_{f(p)} g \, \mathrm{d}_p f$. Hence, for all $\mathbf{u}, \mathbf{v} \in \mathrm{T}_p M$ 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, then we have $\langle \mathrm{d}_p(g \circ f)\, \mathbf{u}, \mathrm{d}_p(g \circ f)\, \mathbf{v} \rangle = \langle \mathrm{d}_p f \, \mathbf{u}, \mathrm{d}_p f \, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$ and hence $g \circ f$ is isometric. If $g \circ f$ is isometric, by the foregoing argument, $g = g \circ f \circ f^{-1}$ is isometric. The second assertion is proved similarly. □

Isometries between manifolds are *distance preserving* while isometric immersions are *path-length preserving*. We make this precise in the following lemma, which is straightforward to prove.

Lemma 2.5. *Let $f : M \to N$ be a differentiable map between Riemannian manifolds M, N.*

1. *If f is an isometric immersion, then for each piecewise differentiable curve $\gamma : [0, 1] \to M$ we have for the length $l(\gamma) = l(f \circ \gamma)$. In particular, for all $p, q \in M$ we have $\operatorname{dist}_M(p, q) \geq \operatorname{dist}_N(f(p), f(q))$.*
2. *If f is an isometry, for all $p, q \in M$ we have $\operatorname{dist}_M(p, q) = \operatorname{dist}_N(f(p), f(q))$.*

We close this subsection with a lemma that is useful when it comes to proving isometric properties of linear maps.

Lemma 2.6. *Let $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear form and $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Then the following holds: $\forall \, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : \langle A\mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ if and only if $\forall \, \mathbf{x} \in \mathbb{R}^n : \langle A\mathbf{x}, A\mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle$.*

Proof. The claim follows from $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2} \left(\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle - \langle \mathbf{x}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle \right)$. □

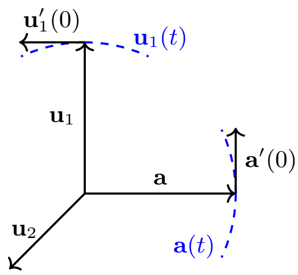


FIGURE 2.1. A sketch of the construction made in the proof of Proposition 2.7.

2.4. **Orthonormal frames.** An orthonormal frame in $\mathbb{R}^n$ is an ordered orthonormal basis of $\mathbb{R}^n$. We write orthonormal frames as ordered tuples $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$. The following proposition will be useful.

Proposition 2.7. *Let* $\gamma : (-1, 1) \to \mathbb{S}(\mathbb{R}^n)$ *be a curve with* $\mathbf{a} := \gamma(0)$ *and* $\mathbf{x} := \gamma'(0) \neq 0$. *Then, there exists a curve* $\Gamma(t) = (\mathbf{u}_1(t), \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}, \mathbf{a}(t))$ *in the set of orthonormal frames satisfying the following properties:*

(1) $\mathbf{u}_1(0) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$*;*
(2) $\mathbf{u}_1'(0) = -\|\mathbf{x}\|\,\mathbf{a}$*;*
(3) $\langle \mathbf{u}_1'(0), \mathbf{u}_1(0)\rangle = 0$*;*
(4) For all $2 \leq j \leq n-1$*:* $\langle \mathbf{u}_1'(0), \mathbf{u}_j\rangle = 0$*;*
(5) For all $2 \leq j \leq n-1$*:* $\langle \mathbf{x}, \mathbf{u}_j\rangle = 0$*;*

Proof. We construct $\Gamma(t)$ explicitly. Let $\mathbf{u}_1 := \frac{\mathbf{x}}{\|\mathbf{x}\|}$. Since $\mathrm{T}_{\mathbf{a}}\mathbb{S}(\mathbb{R}^n) = \{\mathbf{w} \in \mathbb{R}^n \mid \langle \mathbf{w}, \mathbf{a}\rangle = 0\}$, we have $\langle \mathbf{a}, \mathbf{u}_1\rangle = 0$. We can thus complete $\{\mathbf{a}, \mathbf{u}_1\}$ to an orthonormal basis $\{\mathbf{a}, \mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}\}$ of $\mathbb{R}^n$. Consider the orthogonal transformation $U = \mathbf{u}_1\mathbf{a}^T - \mathbf{a}\mathbf{u}_1^T + \sum_{j=2}^{n-1} \mathbf{u}_j\mathbf{u}_j^T$ that rotates $\mathbf{a}$ to $\mathbf{u}_1$, $\mathbf{u}_1$ to $-\mathbf{a}$ and leaves $\{\mathbf{u}_2, \ldots, \mathbf{u}_{n-1}\}$ fixed. Let $\mathbf{a}(t) := \gamma(t)$ and $\mathbf{u}_1(t) := U\,\mathbf{a}(t)$; see Figure 2.1 for a sketch of this construction. Now take $\Gamma(t) = (\mathbf{u}_1(t), \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}, \mathbf{a}(t))$. By construction, conditions (1) and (5) hold. Moreover, $\mathbf{u}_1'(0) = U\,\mathbf{a}'(0) = U\mathbf{x} = \|\mathbf{x}\|U\mathbf{u}_1 = -\|\mathbf{x}\|\mathbf{a}$, which implies (2), (3) and (4). This finishes the proof. □

3. PROOF OF THE MAIN THEOREM

Recall from the introduction the projection distance that was defined on $\mathrm{Gr}(\Pi, n)$: If the subspaces $V, W \subset \mathbb{R}^\Pi$ are of dimension n, the projection distance between them is $\|\pi_V - \pi_W\|$.

The projection distance, however, is not given by some Riemannian metric on $\mathrm{Gr}(\Pi, n)$. In fact, there is a unique orthogonally invariant Riemannian metric on $\mathrm{Gr}(\Pi, n)$ when $\Pi > 4$; see [33]. The associated distance is given by $d(V, W) = \sqrt{\sum_{i=1}^n \theta_i^2}$, where $\theta_1, \ldots, \theta_n$ are the *principal angles* [4] between V and W. From this we construct the following distance function on $\mathrm{Gr}(\Pi, n)^{\times r}$:

$$\mathrm{dist}_{\mathrm{R}}((V_i)_{i=1}^r, (W_i)_1^r) := \sqrt{\sum_{i=1}^r d(V_i, W_i)^2}. \tag{3.1}$$

We can also express the projection distance in terms of the principal angles between V and W: $\|\pi_V - \pi_W\| = \max_{1\leq i\leq n} |\sin\theta_i|$; see, e.g., [40, Table 2]. Since, for all $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ we have

$|\sin(\theta)| \le |\theta|$, this shows that

$$\mathrm{dist}_{\mathrm{P}}((V_i)_{i=1}^r, (W_i)_{i=1}^r) \le \mathrm{dist}_{\mathrm{R}}((V_i)_{i=1}^r, (W_i)_{i=1}^r) \tag{3.2}$$

This inequality is important as it allows us to prove the inequality from Theorem 1.1 by replacing $\mathrm{dist}_{\mathrm{R}}$ by $\mathrm{dist}_{\mathrm{P}}$. The advantage of using $\mathrm{dist}_{\mathrm{R}}$ is that it comes from a Riemannian metric, so that we may use the framework from Section 2.3 to prove inequalities between distances defined on different manifolds. It turns out that the weighted distance is given by a Riemannian metric, as well. This we prove next.

Lemma 3.1. *Let $\langle\cdot,\cdot\rangle$ denote the Fubini-Study metric on $\mathbb{P}(\mathbb{R}^{n_i})$ and let the weighted inner product $\langle\cdot,\cdot\rangle_{\mathrm{w}}$ on the tangent space to a point $p \in \mathbb{P}(\mathbb{R}^{n_1})\times\cdots\times\mathbb{P}(\mathbb{R}^{n_d})$ be defined 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 $\mathbf{u} = (\mathbf{u}^1,\ldots,\mathbf{u}^d)$ and $\mathbf{v} = (\mathbf{v}^1,\ldots,\mathbf{v}^d)$, we define $\langle\mathbf{u},\mathbf{v}\rangle_{\mathrm{w}} := \sum_{i=1}^d (n-n_i)\langle\mathbf{u}^i,\mathbf{v}^i\rangle$. 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}.*

Proof. Let $\gamma(t) = (\gamma_1(t),\ldots,\gamma_d(t))$ 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 the curve γ as the shortest path between two points on a products of d spheres with radii $\sqrt{n-n_1},\ldots,\sqrt{n-n_d}$. The length of this shortest path is $d_{\mathrm{w}}(p,q)$. □

Let σ be the projective Segre map from (1.5). By [31, Section 4.3.4.], σ is a diffeomorphism and we define a Riemannian metric g on $\mathbb{P}\mathcal{S}$ to be the pull-back metric of $\langle\cdot,\cdot\rangle_{\mathrm{w}}$ under σ^{-1}; see [32, Proposition 13.9]. Then, by construction, we have the following result.

Corollary 3.2. *The weighted distance $\mathrm{dist}_{\mathrm{w}}$ on $\mathbb{P}\mathcal{S}^{\times r}$ is given by the Riemannian metric g.*

The proof of Theorem 1.1 uses the following important result. It allows us to compare distances on $\mathcal{S}^{\times r}$ and $\mathrm{Gr}(\Pi,n)$ using Lemma 2.5.

Proposition 3.3. *We consider to $\mathbb{P}\mathcal{S}$ to be endowed with the weighted metric from Definition 1.2 and $\mathrm{Gr}(\Pi,n)$ to be endowed with the unique orthogonal invariant metric on the Grassmannian. Then, the map $\phi: \mathbb{P}\mathcal{S} \to \mathrm{Gr}(\Pi,n), [\mathfrak{A}] \mapsto \mathrm{T}_{\mathfrak{A}}\mathcal{S}$ is an isometric immersion.*

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

Proposition 3.3 lies at the heart of this section. We postpone its quite technical proof until after the proof of Theorem 1.1, which we present next.

Proof of Theorem 1.1. Assume that $\mathrm{Gr}(\Pi,n)^{\times r}$ is endowed with the product metric of the unique orthogonally invariant 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 $\mathbb{P}\mathcal{S}$ and $\mathrm{Gr}(\Pi,n)$, respectively, that the r-fold product

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

is an isometric immersion. The associated distance on $\mathrm{Gr}(\Pi,n)^{\times r}$ is dist_R from (3.1). By Lemma 2.5 (1) this implies that

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

Recall from (1.4) the definition of Σ_{Gr} and note that $\phi^{\times r}(\Sigma_{\mathbb{P}}) \subset \Sigma_{\mathrm{Gr}}$ by construction. Consequently,

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

so that, by (3.2),

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

By (1.2), the latter equals $1/\kappa(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$, which proves the assertion. □

Having shown that proving Proposition 3.3 suffices for concluding the proof of the main theorem, we now focus on proving that ϕ is an isometric immersion.

Proof of Proposition 3.3. In the remainder of this proof, we abbreviate $\mathbb{P}^{m-1} := \mathbb{P}(\mathbb{R}^m)$. Consider the following commutative diagram:

$$\begin{array}{ccc} \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1} & \xrightarrow{\ \sigma\ } & \mathbb{P}\mathcal{S} \\ {\scriptstyle \psi := \iota\circ\phi\circ\sigma}\downarrow & & \downarrow{\scriptstyle \phi} \\ \mathbb{P}(\wedge^n \mathbb{R}^\Pi) & \xleftarrow[\ \iota\ ]{} & \mathrm{Gr}(\Pi, n) \end{array}$$

Herein, σ as defined in (1.5) is an isometry by the definition, ϕ is defined as in the statement of the proposition, and ι is the Plücker embedding [21, Chapter 3.1.]. 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. It is known that the Plücker embedding is an isometry; see, e.g., [23, Section 2] or [20, Chapter 3, Section 1.3].

Since σ and ι are isometries, it follows from Lemma 2.4 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 2.3, 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, by applying Lemma 2.6 to both sides of the equality it suffices to prove that

$$\langle x, x\rangle_{\mathrm{w}} = \langle (\mathrm{d}_p\psi)(x), (\mathrm{d}_p\psi)(x)\rangle. \tag{3.3}$$

Whenever $\gamma : (-1, 1) \to \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}$ is a smooth curve with $\gamma(0) = p$ and $\gamma'(0) = x$, the action of the differential is computed as follows according to [32, Corollary 3.25]:

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

We start by constructing γ explicitly. If we let $p = (p_1, \ldots, p_d) \in \mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}$, then

$$\mathrm{T}_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}) = \mathrm{T}_{p_1}\mathbb{P}^{n_1-1} \times \cdots \times \mathrm{T}_{p_d}\mathbb{P}^{n_d-1},$$

and 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., $[\mathbf{a}_i] = p_i$ with $\|\mathbf{a}_i\| = 1$ in the Euclidean norm. Letting $\mathbf{a}_i^\perp = \{\mathbf{u} \in \mathbb{R}^{n_i} \mid \langle \mathbf{u}, \mathbf{a}_i\rangle = 0\}$ denote the orthogonal complement of $\mathbf{a}_i$ in $\mathbb{R}^{n_i}$, by [9, Section 14.2] we can then identify $\mathbf{a}_i^\perp = \mathrm{T}_{p_i}\mathbb{P}^{n_i-1}$. 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. Since the unit sphere $\mathbb{S}(\mathbb{R}^{n_i})$ is a smooth manifold, we can 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 can assume that γ_i is the exponential map [32, Chapter 20]. We claim that we can write $\gamma(t) = (\pi_1 \circ \gamma_1(t), \ldots, \pi_d \circ \gamma_d(t))$, where $\pi_i : \mathbb{S}(\mathbb{R}^{n_i}) \to \mathbb{P}^{n_i - 1}$ is the canonical projection. Indeed, $\gamma(0) = ([\mathbf{a}_1], \ldots, [\mathbf{a}_d]) = p$ and

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

where P_A denotes the orthogonal projection onto the linear subspace A, where the second equality is due to [9, Lemma 14.8], and where the last step is due to the identification $\mathbf{a}_i^\perp \simeq \mathrm{T}_{p_i}\mathbb{P}^{n_i-1}$.

Next, we compute $\psi \circ \gamma$. First, we have

$$(\sigma \circ \gamma)(t) = [\gamma_1(t) \otimes \cdots \otimes \gamma_d(t)].$$

Now note that by applying Proposition 2.7 to γ_i we find a smooth curve

$$\Gamma_i(t) = \big(U_i\gamma_i(t), \mathbf{u}_2^i, \ldots, \mathbf{u}_{n_i-1}^i, \gamma_i(t)\big) := \big(\mathbf{u}_1^i(t), \mathbf{u}_2^i(t), \ldots, \mathbf{u}_{n_i-1}^i(t), \gamma_i(t)\big) \tag{3.4}$$

in the set of orthonormal frames on $\mathbb{R}^{n_i}$, where $U_i \in \mathbb{R}^{n_i \times n_i}$ and $\mathbf{u}_j^i \in \mathbb{R}^{n_i}$.

By [31, Section 4.6.2] and the definition of the orthonormal frames $\Gamma_i(t)$, it follows that a basis for $\mathrm{T}_{\mathfrak{A}(t)}\mathcal{S}$ is given by

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

where

$$\mathfrak{A}(t) := \gamma_1(t) \otimes \cdots \otimes \gamma_d(t).$$

and

$$\mathfrak{A}_{(i,j)}(t) = \gamma_1(t) \otimes \cdots \otimes \gamma_{i-1}(t) \otimes \mathbf{u}_j^i(t) \otimes \gamma_{i+1}(t) \otimes \cdots \otimes \gamma_d(t). \tag{3.5}$$

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

$$(\psi \circ \gamma)(t) = (\iota \circ \phi)([\mathfrak{A}(t)]) = \pi\left(\mathfrak{A}(t) \wedge \left(\bigwedge_{i=1}^{d} \bigwedge_{j=1}^{n_i-1} \mathfrak{A}_{(i,j)}(t)\right)\right) =: \pi\big(\mathbf{g}(t)\big); \tag{3.6}$$

see [21, Chapter 3.1.C]. Note in particular that the right-hand side of (3.6) is independent of the specific choice of the orthonormal frames $\Gamma_i(t)$ that were constructed via Proposition 2.7, because the exterior product of another basis is just a scalar multiple of the basis we chose.

We are now prepared to compute the derivative of $(\psi \circ \gamma)(t) = (\pi \circ \mathbf{g})(t) = [\mathbf{g}(t)]$. According to [9, 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 $\wedge^n \mathbb{R}^\Pi$. Using the computation rules for inner products from Lemma 2.1 (1) and the definitions of the orthonormal frames $\Gamma_i(t)$ in (3.4), we find

$$\langle \mathfrak{A}(t), \mathfrak{A}(t)\rangle = \prod_{i=1}^{d} \langle \gamma_i(t), \gamma_i(t)\rangle = 1; \tag{3.7}$$

$$\langle \mathfrak{A}(t), \mathfrak{A}_{(i,j)}(t)\rangle = \langle \gamma_i(t), \mathbf{u}_j^i(t)\rangle \prod_{k\neq i} \langle \gamma_k(t), \gamma_k(t)\rangle = 0; \tag{3.8}$$

$$\langle \mathfrak{A}_{(i,j)}(t), \mathfrak{A}_{(k,\ell)}(t)\rangle = \begin{cases} 1, & \text{if } (i,j) = (k,\ell) \\ 0, & \text{else.} \end{cases} \tag{3.9}$$

In other words, $\mathcal{B}(t)$ is an *orthonormal basis* for $\mathrm{T}_{\mathfrak{A}(t)}\mathcal{S}$. By Lemma 2.1 (2), we therefore have

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

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 3.4. *Let* $\mathfrak{A} := \mathfrak{A}(0)$ *and* $\mathfrak{A}_{(i,j)} := \mathfrak{A}_{(i,j)}(0)$ *and write*

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

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

We prove this lemma at the end of this section. We can now prove (3.3). From Lemma 3.4, we find

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

Reordering the terms, one finds

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

where the penultimate equality follows from the formula $n = 1 + \sum_{i=1}^{d}(n_i - 1)$ in (2.1). This proves (3.3) so that ϕ is an isometric map.

Finally, (3.3) 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. □

The final step consists of proving Lemma 3.4. This is performed next.

Proof of Lemma 3.4. By the definition of $\mathbf{g}(t)$ and the product rule of differentiation, the first term of $\mathrm{d}_0\mathbf{g}$ is $\mathfrak{A}'(0) \wedge \bigwedge_{i=1}^{d}\bigwedge_{j=1}^{n_i-1} \mathfrak{A}_{(i,j)}$. We have

$$\mathfrak{A}'(0) = \sum_{\lambda=1}^{d} \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{\lambda-1} \otimes \mathbf{x}_\lambda \otimes \mathbf{a}_{\lambda+1} \otimes \cdots \otimes \mathbf{a}_d = \sum_{\lambda=1}^{d} \|\mathbf{x}_\lambda\| \mathfrak{A}_{(\lambda,1)}. \tag{3.10}$$

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

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

From the above it follows that all of the terms of $\mathrm{d}_0\mathbf{g}$ involve $\mathfrak{A}'_{(i,j)}(0)$. From (3.5), we find

$$\mathfrak{A}'_{(i,j)}(0) = \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes \mathrm{d}_0\mathbf{u}^i_j(t) \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d + \sum_{1\le\lambda\ne i\le d} \mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_{\lambda-1} \otimes \mathbf{x}_\lambda \otimes \mathbf{a}_{\lambda+1} \otimes \cdots \otimes \mathbf{a}_{i-1} \otimes \mathbf{u}^i_j \otimes \mathbf{a}_{i+1} \otimes \cdots \otimes \mathbf{a}_d,$$

where the shorthand notation $\mathbf{u}^i_j = \mathbf{u}^i_j(0)$ was used. We introduce now the notation

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

The rationale behind this is that for $j = 1$ we have $\mathrm{d}_0\mathbf{u}^i_1(t) = \mathrm{d}_0 U_i \gamma_i(t) = U_i \mathbf{x}_i = -\|\mathbf{x}_i\|\mathbf{a}_i$, while for $j > 1$ we have $\mathrm{d}_0\mathbf{u}^i_j(t) = 0$. Hence, we can write compactly

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

Then,

$$\mathbf{f}_{(i,j)} = s_{(i,j)}\, \mathfrak{A} \wedge \left(\sum_{\lambda=1}^{d} \mathfrak{A}^\lambda_{(i,j)}\right) \wedge \bigwedge_{i=1}^{d} \bigwedge_{1\le j\ne i<n_i} \mathfrak{A}_{(i,j)} = s_{(i,j)} \sum_{1\le\lambda\ne i\le d} \mathfrak{A} \wedge \mathfrak{A}^\lambda_{(i,j)} \wedge \bigwedge_{i=1}^{d} \bigwedge_{1\le j\ne i<n_i} \mathfrak{A}_{(i,j)} =: s_{(i,j)} \sum_{1\le\lambda\ne i\le d} \mathbf{f}^\lambda_{(i,j)}, \tag{3.11}$$

where $s_{(i,j)} \in \{-1, 1\}$ is the sign of the permutation for moving $\mathfrak{A}'_{(i,j)}(0)$ to the second position in the exterior product. We continue by computing for $\lambda \ne i$ and $\mu \ne k$ the value

$$\langle \mathbf{f}^\lambda_{(i,j)}, \mathbf{f}^\mu_{(k,\ell)} \rangle = \det\left(B^T_{(i,j),\lambda} B_{(k,\ell),\mu}\right),$$

where

$$B_{(i,j),\lambda} := \begin{bmatrix} \mathfrak{A} & \mathfrak{A}^\lambda_{(i,j)} & [[\mathfrak{A}_{(i,j)}]_{j\ne i}]_{i=1}^d \end{bmatrix};$$

herein, the column vectors should be interpreted as vectorized tensors. Recall that we have $\langle \mathbf{a}_i, \mathbf{x}_i \rangle = \langle \mathbf{a}_i, \mathbf{u}^i_j \rangle = 0$. Then, it follows from Lemma 2.1 and direct computation that for $\lambda \ne i$ and $\mu \ne k$, we have

$$\begin{aligned} \langle \mathfrak{A}, \mathfrak{A}^\mu_{(k,\ell)} \rangle &= \langle \mathfrak{A}, \mathfrak{A}_{(k,\ell)} \rangle = 0, \\ \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A}^\mu_{(k,\ell)} \rangle &= \delta_{ik}\delta_{j\ell}\delta_{\lambda\mu}\|\mathbf{x}_\lambda\|^2, \\ \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A}_{(k,\ell)} \rangle &= 0. \end{aligned}$$

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

$$\begin{bmatrix} \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A} \rangle & \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A}^\mu_{(k,\ell)} \rangle & [[\langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A}_{(k,\ell)} \rangle]_{\ell\ne k}]_k \end{bmatrix}$$

is a zero row, which implies that $\langle \mathbf{f}_{(i,j),\lambda}, \mathbf{f}_{(k,\ell),\mu} \rangle = 0$. 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

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

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

$$\langle \mathbf{f}_{(i,j)}^\lambda, \mathbf{f}_{(k,\ell)}^\mu \rangle = \delta_{ik}\delta_{j\ell}\delta_{\lambda\mu}\|\mathbf{x}_\lambda\|^2. \tag{3.12}$$

Finally, we can compute $\langle \mathbf{f}_{(i,j)}, \mathbf{f}_{(k,\ell)} \rangle$. From (3.11),

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

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

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

proving the result. □

4. Numerical experiments

To illustrate Theorem 1.1 we performed the following experiment in Matlab R2017b [34] with tensors in $\mathbb{R}^{11} \otimes \mathbb{R}^{10} \otimes \mathbb{R}^5$. Note that the generic rank in that space is 23. For each $2 \le r \le 5$ we first select an ill-posed tensor decomposition $A := (\mathfrak{A}_1, \ldots, \mathfrak{A}_r) \in \mathcal{S}^{\times r}$ as explained next. We can randomly generate a rank-1 tensor $\mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_d$ by sampling the elements of $\mathbf{a}_i$ from a standard normal distribution. Then, A is generated by randomly sampling the first $r-1$ rank-one tensors $\mathfrak{A}_1, \ldots, \mathfrak{A}_{r-1} \in \mathbb{R}^{11\times 10\times 5}$, and then putting $\mathfrak{A}_r := \mathbf{a}_1^1 \otimes \mathbf{x}_2 \otimes \mathbf{x}_3$, where $\mathfrak{A}_1 = \mathbf{a}_1^1 \otimes \mathbf{a}_1^2 \otimes \mathbf{a}_1^3$ and the components of $\mathbf{x}_i$ are sampled from a standard normal distribution. Now,

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

and since a rank-2 matrix decomposition is never unique, it follows that $\mathfrak{A}_1 + \mathfrak{A}_r$ has at least a 2-dimensional family[2] of decompositions, and, hence, so does $\mathfrak{A}_1 + \cdots + \mathfrak{A}_r$. Then, it follows from [6, Corollary 1.2] that $\kappa(A) = \infty$ and hence $A \in \Sigma_{\mathbb{P}}$. Finally, we generate a neighboring tensor decomposition $B := (\mathfrak{B}_1, \ldots, \mathfrak{B}_r) \in \mathcal{S}^{\times r}$ by perturbing A as follows. Let $\mathfrak{A}_i = \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \mathbf{a}_i^3$, and then we set $\mathfrak{B}_i = (\mathbf{a}_i^1 + 10^{-2} \cdot \mathbf{x}_i^1) \otimes (\mathbf{a}_i^2 + 10^{-2} \cdot \mathbf{x}_i^2) \otimes (\mathbf{a}_i^3 + 10^{-2} \cdot \mathbf{x}_i^3)$, where the elements of $\mathbf{x}_i^k$ are randomly drawn from a standard normal distribution.

Denote by $(0,1) \to \mathcal{S}^{\times r}, t \mapsto B_t$ a curve between A and B whose length is $\operatorname{dist}_{\mathrm{w}}(A,B)$. Then, for all t, we have $\operatorname{dist}_{\mathrm{w}}(B_t, \Sigma_{\mathbb{P}}) \le \operatorname{dist}_{\mathrm{w}}(A, B_t)$ and hence, by Theorem 1.1,

$$\frac{1}{\kappa(B_t)} \le \operatorname{dist}_{\mathrm{w}}(A, B_t). \tag{4.1}$$

We expect for small t that $\operatorname{dist}_{\mathrm{w}}(A, B_t) \approx \operatorname{dist}_{\mathrm{w}}(A, B_t)$ and so (4.1) is a good substitute for the true inequality from Theorem 1.1.

The data points in the plots in Figure 4.1 show, for each experiment, $\operatorname{dist}_{\mathrm{w}}(A, B_t)$ 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 (4.1) holds. Moreover, since the data points (approximately) lie on a line parallel to the red line, the plots strongly suggest, at least in the cases covered by the experiments, that for decompositions $A = (\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$ close to $\Sigma_{\mathbb{P}}$ the reverse of Theorem 1.1 could hold as well, i.e., $\operatorname{dist}_{\mathrm{w}}(([\mathfrak{A}_1], \ldots, [\mathfrak{A}_r]), \Sigma_{\mathbb{P}}) \le c \frac{1}{\kappa(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)}$, for some constant $c > 0$ that might dependent on A.

[2]The fact that the family is at least two-dimensional follows from the fact that defect of the 2-secant variety of the Segre embedding of $\mathbb{R}^m \times \mathbb{R}^n$ is exactly 2; see, e.g., [31, Proposition 5.3.1.4].

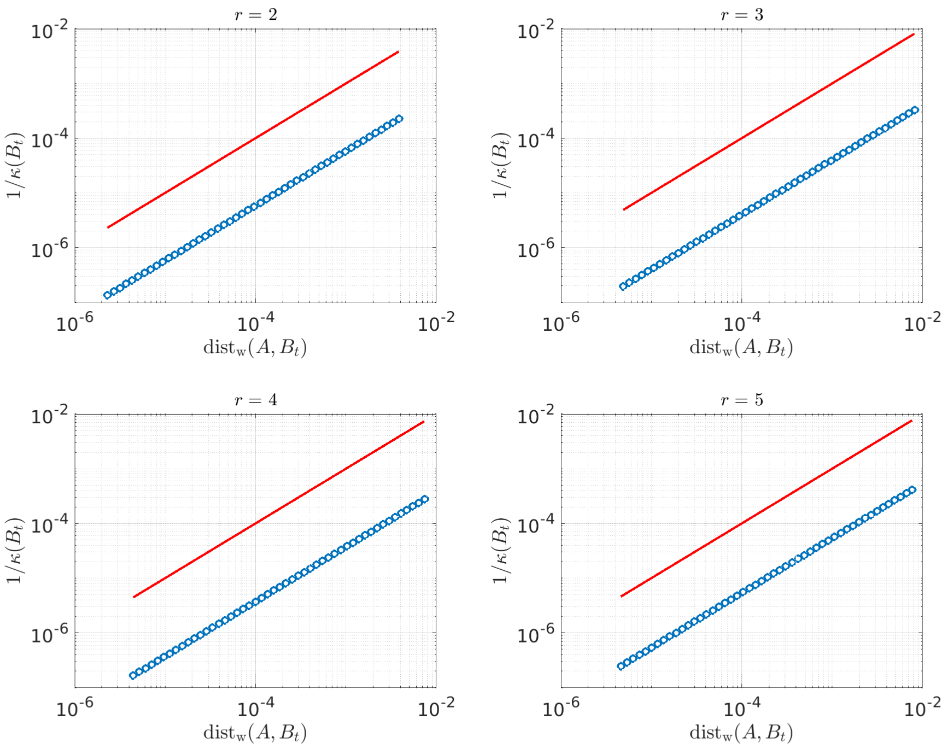


FIGURE 4.1. The blue data points compare the inverse condition number and the estimate of the weighted distance to the locus of ill-posed CPDs for the tensors described in Section 4. The red line illustrates where the data points would lie if the inequality in Theorem 1.1 were an equality. The gap between the red line and the blue data points thus illustrates the sharpness of the bound in Theorem 1.1.

For completeness, in the experiments shown in Figure 4.1, such a bound seems to hold for $c = 17$, 25, 27, and 19 respectively in the cases $r = 2$, 3, 4, and 5.

REFERENCES

1. H. Abo, G. Ottaviani, and C. Peterson, *Induction for secant varieties of Segre varieties*, Trans. Amer. Math. Soc. **361** (2009), 767–792.
2. E. S. Allman, C. Matias, and J. A. Rhodes, *Identifiability of parameters in latent structure models with many observed variables*, Ann. Statist. **37** (2009), no. 6A, 3099–3132.
3. A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, *Tensor decompositions for learning latent variable models*, J. Mach. Learn. Res. **15** (2014), 2773–2832.
4. Å. Björck and Gene H. Golub, *Numerical methods for computing angles between linear subspaces*, Math. Comp. **27** (1973), no. 123, 579–594.
5. C. Bocci, L. Chiantini, and G. Ottaviani, *Refined methods for the identifiability of tensors*, Ann. Mat. Pura Appl. **193** (2014), 1691–1702.
6. P. Breiding and N. Vannieuwenhoven, *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl. (2017), Accepted.
7. Paul Breiding, *Numerical and statistical aspects of tensor decompositions*, PhD Thesis, Technische Universität Berlin, 2017, http://dx.doi.org/10.14279/depositonce-6148.
8. P. Bürgisser, M. Clausen, and M. A. Shokrollahi, *Algebraic Complexity Theory*, Grundlehren der mathematischen Wissenshaften, vol. 315, Springer, Berlin, Germany, 1997.

9. P. Bürgisser and F. Cucker, *Condition: The Geometry of Numerical Algorithms*, Grundlehren der mathematischen Wissenschaften, vol. 349, Springer, Heidelberg, 2013. MR 3098452
10. G. Young C. Eckart, *A principal axis transformation for non-hermitian matrices*, Bull. Am. Math. Soc. **45** (1939), 118–121.
11. L. Chiantini and G. Ottaviani, *On generic identifiability of* 3*-tensors of small rank*, SIAM J. Matrix Anal. Appl. **33** (2012), no. 3, 1018–1037.
12. L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIAM J. Matrix Anal. Appl. **35** (2014), no. 4, 1265–1287.
13. P. Comon, *Independent component analysis, a new concept?*, Signal Proc. **36** (1994), no. 3, 287–314.
14. P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Elsevier, 2010.
15. F. Cucker D. Cheung, *Solving linear programs with finite precision: I, condition numbers and random programs*, Math. Program. **99** (2004), 175–196.
16. ———, *Solving linear programs with finite precision: II. algorithms*, J. Complexity **22** (2006), 305–335.
17. J. W. Demmel, *The geometry of ill-conditioning*, J. Complexity **3** (1987), no. 2, 201–229.
18. ———, *On condition numbers and the distance to the nearest ill-posed problem*, Numer. Math. **51** (1987), no. 3, 251–289.
19. M. do Carmo, *Riemannian Geometry*, Birhäuser, 1993.
20. D. B. Fuchs, *Topology II*, Encyclopaedia of Mathematical Sciences, vol. 24, ch. Classical Manifolds, pp. 199–251, Springer–Verlag, Berlin, Heidelberg, 2004.
21. I.M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants and Multidimensional Determinants*, Modern Birkhäuser Classics, Birkhäuser, 1994.
22. W. H. Greub, *Multilinear algebra*, Springer-Verlag, 1978.
23. P. Griffiths, *On Cartan's method of Lie groups and moving frames as applied to uniqueness and existence questions in differential geometry*, Duke Math. J. **41** (1974), no. 4, 775–814.
24. W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Series in Computational Mathematics, vol. 42, Springer-Verlag, 2012.
25. J. Harris, *Algebraic Geometry, A First Course*, Graduate Text in Mathematics, vol. 133, Springer-Verlag, 1992.
26. N. J. Higham, *Accuracy and stability of numerical algorithms*, second ed., Society for Industrial and Applied Mathematics, 1996.
27. F. L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, Journal of Mathematics and Physics **6** (1927), 164–189.
28. D. Hough, *Explaining and ameliorating the condition of zeros of polynomials*, PhD Thesis, Mathematics Department, University of California, Berkeley, 1977.
29. W. Kahan, *Numerical linear algebra*, Can. Math. Bull. **9** (1966), 757–801.
30. P. M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley series in probability and statistics, John Wiley & Sons, Hoboken, New Jersey, 2008.
31. J. M. Landsberg, *Tensors: Geometry and Applications*, Graduate Studies in Mathematics, vol. 128, AMS, Providence, Rhode Island, 2012.
32. J. M. Lee, *Introduction to Smooth Manifolds*, second ed., Graduate Texts in Mathematics, vol. 218, Springer, New York, USA, 2013.
33. Kurt Leichtweiss, *Zur riemannschen geometrie in grassmannschen mannigfaltigkeiten*, Mathematische Zeitschrift **76** (1961), 334–366.
34. MATLAB, *R2016b*, Natick, Massachusetts, 2016.
35. P. McCullagh, *Tensor Methods in Statistics*, Monographs on statistics and applied probability, Chapman and Hall, New York, 1987.
36. N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and Ch. Faloutsos, *Tensor decomposition for signal processing and machine learning*, arXiv:1607.01668 (2016).
37. A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*, John Wiley & Sons, Hoboken, New Jersey, 2004.
38. N. Vannieuwenhoven, *A condition number for the tensor rank decomposition*, Linear Algebra Appl. **535** (2017), 35–86.
39. J. H. Wilkinson, *Note on matrices with a very ill-conditioned eigenproblem*, Numer. Math. **19** (1972), 176–178.
40. K. Ye and L.-H. Lim, *Schubert varieties and distances between subspaces of different dimensions*, arXiv:1407.0900 (2014).