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

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 $A \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ as a minimum-length linear combination of pure tensors:

\[ (CPD) \quad A = \sum_{i=1}^{r} a_{1}^i \otimes a_{2}^i \otimes \cdots \otimes a_{d}^i, \quad a_{k}^i \in \mathbb{R}^{n_k}, \]

where $\otimes$ is the tensor product:

\[ (1.1) \quad a_{1}^i \otimes a_{2}^i \otimes \cdots \otimes a_{d}^i = [a_{1}^{(1)}(i_1) a_{2}^{(2)}(i_2) \cdots a_{d}^{(d)}(i_d)]_{i_1, i_2, \ldots, i_d=1}^{n_1, n_2, \ldots, n_d} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \]

with $a_{k}^i = [a_{k}^{(k)}(i_k)]_{i_k=1}^{n_k}$.

The smallest $r$, for which the expression (CPD) is possible is called the rank of $A$.

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] 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 $\mathcal{S} := S_{n_1 \cdots n_d} := \{ a^1 \otimes a^2 \otimes \cdots \otimes a^d \mid a^i \in \mathbb{R}^{n_i} \setminus \{0\} \}$, assuming that $n_1, \ldots, n_k$ were fixed.

According to [6, Theorem 1.3], the condition number of the CPD at a decomposition $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r) \in S^{\times r}$ can then be expressed as the inverse distance of the tuple of tangent spaces $(\mathcal{T}_{\mathfrak{A}_1}, \mathcal{S}, \ldots, \mathcal{T}_{\mathfrak{A}_r}, \mathcal{S})$ to ill-posedness:

$$
\kappa(\mathfrak{A}_1, \ldots, \mathfrak{A}_r) = \frac{1}{\text{dist}_P((\mathcal{T}_{\mathfrak{A}_1}, \mathcal{S}, \ldots, \mathcal{T}_{\mathfrak{A}_r}, \mathcal{S}), \Sigma_{Gr})},
$$

where $\Sigma_{Gr}$ and the distance $\text{dist}_P$ 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 $S^{\times r}$. Our main theorem is Theorem 1.1 below. Before we state it, though, we recall the definitions of $\Sigma_{Gr}$ and $\text{dist}_P$ from [6].

Let $n := \dim \mathcal{S}$ and $\Pi := n_1 \cdots n_d$. Denote by $\text{Gr}(\Pi, n)$ 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 $(\mathcal{T}_{\mathfrak{A}_1}, \mathcal{S})|_{\mathfrak{A}_i} \in \text{Gr}(\Pi, n)^{\times r}$. The projection distance on $\text{Gr}(\Pi, n)$ is given by $\| \pi_V - \pi_W \|_2$, where $\pi_V$ and $\pi_W$ are the orthogonal projections on the spaces $V$ and $W$ respectively, and $\| \cdot \|$ is the spectral norm. This distance measure is extended to $\text{Gr}(\Pi, n)^{\times r}$ in the usual way:

$$
\text{dist}_P((V_i)_{i=1}^r, (W_i)_{i=1}^r) := \left( \sum_{i=1}^r \| \pi_{V_i} - \pi_{W_i} \|^2 \right)^{1/2}.
$$

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

(1.4)

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

The decomposition $(\mathfrak{A}_1, \ldots, \mathfrak{A}_r)$ whose corresponding tangent space lies in $\Sigma_{Gr}$ 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
all of the decompositions \((\mathcal{A}_1(t), \ldots, \mathcal{A}_r(t))\) of \(\mathcal{A}\) are ill-posed decompositions. 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 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 \(\mathcal{A}_i\). Hence, to describe the condition number as an inverse distance to ill-posedness on \(S^{\times r}\) 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 \(\pi : \mathbb{R}\Pi \setminus \{0\} \to \mathbb{P}(\mathbb{R}\Pi)\). We put \(\mathbb{P}\mathcal{S} := \pi(S)\) and for points \(\mathcal{A} \in \mathbb{R}\Pi\) we denote \(|\mathcal{A}| := \pi(|\mathcal{A}|)\). Let \((\mathcal{A}_1, \ldots, \mathcal{A}_r) \in \mathcal{S}^{\times r}\). Then,

\[
\frac{1}{\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)} = \text{dist}_p((T\mathcal{A}_1, S, \ldots, T\mathcal{A}_r, S), \Sigma_{Gr}) \leq \text{dist}_w([|\mathcal{A}_1|], \ldots, [|\mathcal{A}_r|], \Sigma_p),
\]

where

\[\Sigma_p = \{([|\mathcal{A}_1|], \ldots, [|\mathcal{A}_r|]) \in (\mathbb{P}\mathcal{S})^{\times r} | \kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) = \infty\}\]

and the distance \(\text{dist}_w\) 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 \((\mathcal{A}_1, \ldots, \mathcal{A}_r)\) close to \(\Sigma_p\) that there is a constant \(c > 0\) such that \(\text{dist}_w([|\mathcal{A}_1|], \ldots, [|\mathcal{A}_r|], \Sigma_p) \leq c/n(\mathcal{A}_1, \ldots, \mathcal{A}_r)\).

It is important to note that the lower bound in the theorem can be computed efficiently via the spectral characterization of the condition number \(\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)\) 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 \((\cdot, \cdot)\) denote the Fubini-Study metric on \(\mathbb{P}(\mathbb{R}^n)\) and let \(d_p\) be the corresponding distance on \(\mathbb{P}(\mathbb{R}^n)\); 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

\[
d_w(p, q) := \left(\sum_{i=1}^d (n - n_i) d_p(q_i, q_i)^2\right)^{\frac{1}{2}},
\]

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

\[
\text{dist}_w([\mathcal{A}_1], \ldots, [\mathcal{A}_r], ([\mathcal{B}_1], \ldots, [\mathcal{B}_r])) := \left(\sum_{i=1}^r d_w(\sigma^{-1}(\mathcal{A}_i), \sigma^{-1}(\mathcal{B}_i))^2\right)^{\frac{1}{2}},
\]

where \(\sigma^{-1}\) is the inverse of the projective Segre map

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

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.
Figure 1.1. The picture depicts relative errors in the weighted distance, where $x^1 \in P(\mathbb{R}^{n_1})$ and $x^2 \in P(\mathbb{R}^{n_2})$ with $n_1 > n_2$. The relative errors of the tangent directions $\Delta x^1$ and $\Delta x^2$ 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 $P(\mathbb{R}^{n_2})$.

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 $P(\mathbb{R}^n)$ and the unit sphere of dimension $n - 1$ is denoted $S(\mathbb{R}^n)$.

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

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

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

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

Lemma 2.1. For $1 \leq k \leq d$, let $x_k, y_k \in \mathbb{R}^{n_k}$, and let $\langle \cdot, \cdot \rangle$ denote the standard Euclidean inner product. Then, $(x_1 \otimes \cdots \otimes x_d, y_1 \otimes \cdots \otimes y_d) = \prod_{j=1}^{d} \langle x_j, y_j \rangle$.

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

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

$$\wedge : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \wedge^d \mathbb{R}^n, (x_1,\ldots,x_d) \mapsto \frac{1}{d!} \sum_{\pi \in S_d} \text{sgn}(\pi)x_{\pi_1} \otimes x_{\pi_2} \otimes \cdots \otimes x_{\pi_d};$$

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 $x_1 \wedge \cdots \wedge x_d$ for $\wedge(x_1,\ldots,x_d)$. The next result is well known.

Lemma 2.2. Let $x_1,\ldots,x_d,y_1,\ldots,y_d \in \mathbb{R}^m$. Let $\langle \cdot, \cdot \rangle$ be the standard Euclidean inner product. Then the induced inner product on $\wedge^d \mathbb{R}^m$ satisfies $(x_1 \wedge \cdots \wedge x_d, y_1 \wedge \cdots \wedge y_d) = \det(\langle x_i, y_j \rangle)_{i,j=1}^{d}$.

Proof. See, e.g., [22, Section 4.8] or [32, Proposition 14.11].
Proof. Let differentiable maps. We have\(\langle M(2.2)\text{ dist}\) between two points \(p, q \in M\) is defined as
\[
\text{dist}_M(p, q) = \inf \{l(\gamma) \mid \gamma(0) = p, \gamma(1) = q\},
\]
where the infimum is over all piecewise differentiable curves \(\gamma : [0, 1] \to M\) and the length of a curve is defined as \(l(\gamma) = \int_0^1 \sqrt{\gamma'(t) \cdot \gamma'(t)} \, dt\). The distance \(\text{dist}_M\) 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 \(d_pf\) 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 \(u, v \in T_pM\) it holds that \(g_p(u, v) = h_{f(p)}(d_p f(u), d_p f(v))\). 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 \(d_p(g \circ f) = d_{f(p)} g \circ d_pf\). Hence, for all \(u, v \in T_pM\) we have \(\langle d_p(g \circ f) u, d_p(g \circ f) v\rangle = \langle d_{f(p)} g d_pf u, d_{f(p)} g d_pf v\rangle\). We prove (1): If \(g\) is isometric, then we have \(\langle d_p(g \circ f) u, d_p(g \circ f) v\rangle = \langle d_pf u, d_pf v\rangle = \langle u, v\rangle\) 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 the length \(l(\gamma) = l(f \circ \gamma)\). In particular, for all \(p, q \in M\) we have \(\text{dist}_M(p, q) \geq \text{dist}_N(f(p), f(q))\).
2. If \(f\) is an isometry, for all \(p, q \in M\) we have \(\text{dist}_M(p, q) = \text{dist}_N(f(p), f(q))\).

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: \(\forall x, y \in \mathbb{R}^n : \langle Ax, Ay \rangle = \langle x, y \rangle\) if and only if \(\forall x \in \mathbb{R}^n : \langle Ax, Ax \rangle = \langle x, x \rangle\).

Proof. The claim follows from \(\langle x, y \rangle = \frac{1}{2}(\langle x - y, x - y \rangle - \langle x, x \rangle - \langle y, y \rangle)\).
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 \((u_1, u_2, \ldots, u_n)\). The following proposition will be useful.

**Proposition 2.7.** Let \(\gamma : (-1, 1) \to S(\mathbb{R}^n)\) be a curve with \(a := \gamma(0)\) and \(x := \gamma'(0) \neq 0\). Then, there exists a curve \(\Gamma(t) = (u_1(t), u_2, \ldots, u_{n-1}, a(t))\) in the set of orthonormal frames satisfying the following properties:

1. \(u_1(0) = \frac{x}{\|x\|} a\);
2. \(u_1'(0) = -\|x\| a\);
3. \(\langle u_1'(0), u_1(0) \rangle = 0\);
4. For all \(2 \leq j \leq n-1\): \(\langle u_j'(0), u_j \rangle = 0\);
5. For all \(2 \leq j \leq n-1\): \(\langle x, u_j \rangle = 0\);

**Proof.** We construct \(\Gamma(t)\) explicitly. Let \(u_1 := \frac{x}{\|x\|} a\). Since \(T_a S(\mathbb{R}^n) = \{w \in \mathbb{R}^n \mid \langle w, a \rangle = 0\}\), we have \(\langle a, u_1 \rangle = 0\). We can thus complete \(\{a, u_1\}\) to an orthonormal basis \(\{a, u_1, u_2, \ldots, u_{n-1}\}\) of \(\mathbb{R}^n\). Consider the orthogonal transformation \(U = u_1 a^T - a u_1^T + \sum_{j=2}^{n-1} u_j u_j^T\) that rotates \(a\) to \(u_1\) and \(u_1\) to \(-a\) and leaves \(\{u_2, \ldots, u_{n-1}\}\) fixed. Let \(a(t) := \gamma(t)\) and \(u_1(t) := U a(t)\); see Figure 2.1 for a sketch of this construction. Now take \(\Gamma(t) = (u_1(t), u_2, \ldots, u_{n-1}, a(t))\). By construction, conditions (1) and (5) hold. Moreover, \(u_1'(0) = U a'(0) = U x = \|x\| U u_1 = -\|x\| a\), which implies (2), (3) and (4). This finishes the proof. \(\square\)

3. **Proof of the main theorem**

Recall from the introduction the projection distance that was defined on \(\text{Gr}(\Pi, n)\): 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 \(\text{Gr}(\Pi, n)\). In fact, there is a unique orthogonally invariant Riemannian metric on \(\text{Gr}(\Pi, n)\) 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 \(\text{Gr}(\Pi, n)^r\):

\[
(3.1) \quad \text{dist}_R((V_i)_{i=1}^r, (W_i)_{i=1}^r) := \sqrt{\sum_{i=1}^r d(V_i, W_i)^2}.
\]

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) | \leq |\theta|, \text{ this shows that} \\
(3.2) \quad \text{dist}_P((V_1')_{i=1}^r, (W_i')_{i=1}^r) \leq \text{dist}_R((V_1')_{i=1}^r, (W_i')_{i=1}^r) \\
This inequality is important as it allows us to prove the inequality from Theorem 1.1 by replacing \text{dist}_R by \text{dist}_P. The advantage of using \text{dist}_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) and let the weighted inner product \langle \cdot, \cdot \rangle_w on the tangent space to a point \( p \in \mathbb{P}(\mathbb{R}^n) \times \cdots \times \mathbb{P}(\mathbb{R}^n) \) be defined as follows: For all \( u, v \in T_p(\mathbb{P}(\mathbb{R}^n) \times \cdots \times \mathbb{P}(\mathbb{R}^n)) \), where \( u = (u^1, \ldots, u^d) \) and \( v = (v^1, \ldots, v^d) \), we define \( \langle u, v \rangle_w := \sum_{i=1}^d (n - n_i)\langle u^i, v^i \rangle \). Then, the distance on \( \mathbb{P}(\mathbb{R}^n) \times \cdots \times \mathbb{P}(\mathbb{R}^n) \) corresponding to \langle \cdot, \cdot \rangle_w is \( d_w \).

**Proof.** Let \( \gamma(t) = (\gamma_1(t), \ldots, \gamma_d(t)) \) be a piecewise continuous curve in \( \mathbb{P}(\mathbb{R}^n) \times \cdots \times \mathbb{P}(\mathbb{R}^n) \) connecting \( p, q \in \mathbb{P}(\mathbb{R}^n) \times \cdots \times \mathbb{P}(\mathbb{R}^n) \), such that the distance between \( p, q \) given by \langle \cdot, \cdot \rangle_w is

\[
\int_0^1 \left( \sum_{i=1}^d (n - n_i)(\gamma_i'(t), \gamma_i'(t)) \right)^{\frac{1}{2}} dt.
\]

Because \( (n - n_i)(\gamma_i'(t), \gamma_i'(t)) = (\sqrt{n - n_i} \gamma_i'(t), \sqrt{n - n_i} \gamma_i'(t)) \) and because we have the identity of tangent spaces \( T_{\gamma(t)}(\mathbb{P}(\mathbb{R}^n)) = T_{\gamma(t)}S(\mathbb{R}^n) \) for all \( t \), we may view the curve \( \gamma \) as the shortest path between two points on a product of \( d \) spheres with radii \( \sqrt{n - n_1}, \ldots, \sqrt{n - n_d} \). The length of this shortest path is \( d_w(p, q) \).

Let \( \sigma \) be the projective Segre map from (1.5). By [31, Section 4.3.1], \( \sigma \) is a diffeomorphism and we define a Riemannian metric \( g \) on \( \mathbb{P}S \) to be the pull-back metric of \langle \cdot, \cdot \rangle_w under \( \sigma^{-1} \); see [32, Proposition 13.9]. Then, by construction, we have the following result.

**Corollary 3.2.** The weighted distance \( \text{dist}_w \) on \( \mathbb{P}S^{\times r} \) 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 \( S^{\times r} \) and \( \text{Gr}(\Pi, n) \) using Lemma 2.5.

**Proposition 3.3.** We consider \( \mathbb{P}S \) to be endowed with the weighted metric from Definition 1.2 and \( \text{Gr}(\Pi, n) \) to be endowed with the unique orthogonal invariant metric on the Grassmannian. Then, the map \( \phi : \mathbb{P}S \to \text{Gr}(\Pi, n) \) is an isometric immersion.

**Remark.** Note that \( \phi \) is not the Gauss map \( \mathbb{P}S \to \text{Gr}(n-1, \mathbb{P}H), [A] \mapsto [T_A S] \), which maps a tensor to a projective subspace of \( \mathbb{P}H \) of dimension \( n-1 = \dim \mathbb{P}S \).

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 \( \text{Gr}(\Pi, n)^{\times r} \) is endowed with the product metric of the unique orthogonally invariant metric on \( \text{Gr}(\Pi, n) \). Since \( \phi \) is an isometric immersion, it follows from the definitions of the product metrics on the r-fold products of the smooth manifolds \( \mathbb{P}S \) and \( \text{Gr}(\Pi, n) \), respectively, that the r-fold product

\[
\phi^{\times r} : (\mathbb{P}S)^{\times r} \to \text{Gr}(\Pi, n)^{\times r}, ([A_1], \ldots, [A_r]) \mapsto (T_{A_1}S, \ldots, T_{A_r}S)
\]

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

\[
\text{dist}_w\left( ([A_1], \ldots, [A_r]), \Sigma_{P} \right) \geq \text{dist}_R\left( (T_{A_1}S, \ldots, T_{A_r}S), \phi^{\times r}(\Sigma_{P}) \right).
\]
Recall from (1.4) the definition of $\Sigma_{Gr}$ and note that $\phi^{x,y}(\Sigma_{\varphi}) \subset \Sigma_{Gr}$ by construction. Consequently, 
\[ \operatorname{dist}_w((\mathbb{A}_1, \ldots, [\mathbb{A}_r]), \Sigma_{\varphi}) \geq \operatorname{dist}_R((T_{\mathbb{A}_1} S, \ldots, T_{\mathbb{A}_r} S), \Sigma_{Gr}), \]
so that, by (3.2),
\[ \operatorname{dist}_w((\mathbb{A}_1, \ldots, [\mathbb{A}_r]), \Sigma_{\varphi}) \geq \operatorname{dist}_P((T_{\mathbb{A}_1} S, \ldots, T_{\mathbb{A}_r} S), \Sigma_{Gr}). \]
By (1.2), the latter equals $1/\kappa(\mathbb{A}_1, \ldots, \mathbb{A}_r)$, which proves the assertion. \hfill $\Box$

Having shown that proving Proposition 3.3 suffices for concluding the proof of the main theorem, we now focus on proving that $\phi$ 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:
\[
\begin{array}{ccc}
\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1} & \xrightarrow{\sigma} & \mathbb{P}S \\
\psi := \circ \circ \circ \circ & \downarrow & \circ \circ \circ \\
\mathbb{P}(\Lambda^n \mathbb{R}^\Pi) & \xleftarrow{\iota} & \text{Gr}(\Pi, n)
\end{array}
\]
Herein, $\sigma$ as defined in (1.5) is an isometry by the definition, $\phi$ is defined as in the statement of the proposition, and $\iota$ is the Plücker embedding [21, Chapter 3.1]. The image of the Plücker embedding $\mathcal{P} := \iota(\text{Gr}(\Pi, n)) \subset \mathbb{P}(\Lambda^n \mathbb{R}^\Pi)$ is a smooth variety called the Plücker variety. The Fubini-Study metric on $\mathbb{P}(\Lambda^n \mathbb{R}^\Pi)$ makes $\mathcal{P}$ 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 $\sigma$ and $\iota$ are isometries, it follows from Lemma 2.4 that $\phi$ 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 $x, y \in T_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$ we have
\[ \langle x, y \rangle_w = \langle (d_p \psi)(x), (d_p \psi)(y) \rangle. \]
However, by applying Lemma 2.6 to both sides of the equality it suffices to prove that
\[ \langle x, x \rangle_w = \langle (d_p \psi)(x), (d_p \psi)(x) \rangle. \]
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]:
\[ (d_p \psi)(x) = d_0(\psi \circ \gamma). \]
We start by constructing $\gamma$ 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
\[ T_p(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}) = T_{p_1} \mathbb{P}^{n_1-1} \times \cdots \times T_{p_d} \mathbb{P}^{n_d-1}, \]
and we can write $x = (x_1, \ldots, x_d)$ with $x_i \in T_{p_i} \mathbb{P}^{n_i-1}$. For each $i$, we denote by $a_i \in S(\mathbb{R}^{n_i})$ a unit-norm representative for $p_i$, i.e., $[a_i] = p_i$ with $\|a_i\| = 1$ in the Euclidean norm. Letting $a_i^+ = \{ u \in \mathbb{R}^{n_i} \mid \langle u, a_i \rangle = 0 \}$ denote the orthogonal complement of $a_i$ in $\mathbb{R}^{n_i}$, by [9, Section 14.2] we can then identify $a_i^+ = T_{p_i} \mathbb{P}^{n_i-1}$. Moreover, because $a_i$ is of unit norm, the Fubini–Study metric on $T_{p_i} \mathbb{P}^{n_i-1}$ is given by the Euclidean inner product on the linear subspace $a_i^+$. Now, let $x_i$ denote the unique vector in $a_i^+$ corresponding to $x_i$. Since the unit sphere $S(\mathbb{R}^{n_i})$ is a smooth manifold, we can find a curve $\gamma_i : (-1, 1) \to S(\mathbb{R}^{n_i})$ with $\gamma_i(0) = a_i$ and $\gamma_i'(0) = x_i$. Without loss of generality we can assume that $\gamma_i$ 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, \( \gamma(0) = ([a_1], \ldots, [a_d]) = p \) and

\[
\gamma'(0) = ((\pi_1 \circ \gamma_1)'(0), \ldots, (\pi_d \circ \gamma_d)'(0)) = (P_{a_1^+} \gamma_1'(0), \ldots, P_{a_d^+} \gamma_d'(0))
\]

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 \( a_i^+ \simeq T_{n_i} \mathbb{P}^{n_i - 1} \).

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 \( \gamma_1 \) we find a smooth curve

\[
\Gamma_1(t) = \left( U_1(t), u_{21}, \ldots, u_{(n_i - 1)i}, \gamma_1(t) \right) = (u_1(t), u_2(t), \ldots, u_{(n_i - 1)}(t), \gamma_1(t))
\]

in the set of orthonormal frames on \( \mathbb{R}^{n_i} \), where \( U_i \in \mathbb{R}^{n_i \times n_i} \) and \( u_j^i \in \mathbb{R}^{n_i} \).

By [31, Section 4.6.2] and the definition of the orthonormal frames \( \Gamma_1(t) \), it follows that a basis for \( T_{\mathbb{A}(t)} \mathbb{S} \) is given by

\[
B(t) = \{\mathbb{A}(t)\} \cup \{\mathbb{A}_{(i,j)}(t) \mid 1 \leq i \leq d, 1 \leq j \leq n_i - 1\},
\]

where

\[
\mathbb{A}(t) := \gamma_1(t) \otimes \cdots \otimes \gamma_d(t).
\]

and

\[
\mathbb{A}_{(i,j)}(t) = \gamma_1(t) \otimes \cdots \otimes \gamma_{i-1}(t) \otimes u_j^i(t) \otimes \gamma_{i+1}(t) \otimes \cdots \otimes \gamma_d(t).
\]

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

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

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

\[
d_0(\psi \circ \gamma) = P_{(\mathbf{g}(0))} \frac{\mathbf{g}'(0)}{||\mathbf{g}'(0)||}.
\]

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

\[
d_0(\psi \circ \gamma) = P_{(\mathbf{g}(0))} \mathbf{g}'(0) = \mathbf{g}'(0) = d_0 \mathbf{g},
\]

as \( \mathbf{g}'(t) \) would in this case be contained in the tangent space to the sphere over \( \wedge^n \mathbb{R}^n \). Using the computation rules for inner products from Lemma 2.1 (1) and the definitions of the orthonormal
In other words, \( B(t) \) is an orthonormal basis for \( T_{X(t)}S \). By Lemma 2.1 (2), we therefore have

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

which equals \( \det I_n = 1 \).

It now only remains to compute \( d_0g \). For this we have the following result.

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

\[
f_{(i,j)} := \mathcal{A} \wedge \mathcal{A}_{(i,j)} \wedge \cdots \wedge \mathcal{A}_{(j-1,i)} \wedge \mathcal{A}_{(i,j+1)} \wedge \cdots \wedge \mathcal{A}_{(p,n_d-1)}.
\]

The differential satisfies \( d_0g = \sum_{i=1}^d \sum_{j=1}^{n_i-1} f_{(i,j)} \), where \( \langle f_{(i,j)}, f_{(k,l)} \rangle = \delta_{ik} \delta_{jl} \sum_{1 \leq k \neq l \leq d} \langle x_\lambda, x_\lambda \rangle \), where \( \delta_{ij} \) 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

\[
\langle (d_\psi x)(x), (d_\psi x)(x) \rangle = \langle d_0g, d_0g \rangle = \left( \sum_{i=1}^d \sum_{j=1}^{n_i-1} f_{(i,j)} \right)^2 = \sum_{i=1}^d \sum_{j=1}^{n_i-1} \sum_{1 \leq k \neq l \leq d} \langle x_\lambda, x_\lambda \rangle.
\]

Reordering the terms, one finds

\[
\langle (d_\psi x)(x), (d_\psi x)(x) \rangle = \sum_{i=1}^d \langle x_i, x_i \rangle \sum_{1 \leq k \neq l \leq d} \sum_{j=1}^{n_i-1} 1 = \sum_{i=1}^d \langle x_i, x_i \rangle \cdot (n - n_i) = \langle x, x \rangle_w,
\]

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 \( \phi \) is an isometric map.

Finally, (3.3) also entails that \( \phi \) is an immersion. Indeed, for an immersion it is required that \( d_\psi x \) is injective. Suppose that this is false, then there is a nonzero \( x \in T_p(\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}) \) with corresponding nonzero \( x \) such that

\[
0 = \langle 0, 0 \rangle = \langle (d_\psi x)(x), (d_\psi x)(x) \rangle = \langle x, x \rangle_w > 0,
\]

which is a contradiction. Consequently, \( \phi \) is an isometric immersion, concluding the proof. \( \square \)

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

**Proof of Lemma 3.4.** By the definition of \( g(t) \) and the product rule of differentiation, the first term of \( d_0g \) is \( \mathcal{A}’(0) \wedge \bigwedge_{i=1}^d \mathcal{A}_{(i,j)} \). We have

\[
\mathcal{A}’(0) = \sum_{\lambda=1}^d a_1 \otimes \cdots \otimes a_{\lambda-1} \otimes x_\lambda \otimes a_{\lambda+1} \otimes \cdots \otimes a_d = \sum_{\lambda=1}^d \lVert x_\lambda \lVert \mathcal{A}_{(\lambda,1)}.
\]
Hence, from the multilinearity of the exterior product it follows that the first term of $d_0g$ is
\[ \sum_{\lambda=1}^{d} \| x^{\lambda} \| (\mathfrak{A}(\lambda,1) \wedge \mathfrak{A}(1,1) \wedge \cdots \wedge \mathfrak{A}(d,n_d - 1)) = \sum_{\lambda} 0 = 0. \]

From the above it follows that all of the terms of $d_0g$ involve $\mathfrak{A}'_{(i,j)}(0)$. From (3.5), we find
\[
\mathfrak{A}'_{(i,j)}(0) = a_1 \otimes \cdots \otimes a_{i-1} \otimes d_0 u_j^i(t) \otimes a_{i+1} \otimes \cdots \otimes a_d + \sum_{1 \leq \lambda \neq i \leq d} a_1 \otimes \cdots \otimes a_{\lambda-1} \otimes x_{\lambda} \otimes a_{\lambda+1} \otimes \cdots \otimes a_{i-1} \otimes u_j^i \otimes a_{i+1} \otimes \cdots \otimes a_d,
\]
where the shorthand notation $u_j^i(0)$ was used. We introduce now the notation
\[
\mathfrak{A}^\lambda_{(i,j)} := \begin{cases} a_1 \otimes \cdots \otimes a_{\lambda-1} \otimes x_{\lambda} \otimes a_{\lambda+1} \otimes \cdots \otimes a_{i-1} \otimes u_j^i \otimes a_{i+1} \otimes \cdots \otimes a_d & \text{if } \lambda \neq i, \\ 0 & \text{if } (\lambda, j) = (i, 1), \end{cases}
\]
otherwise.

The rationale behind this is that for $j = 1$ we have $d_0 u_j^i(t) = d_0 U_i g_i(t) = U_i x_i = -\|x_i\|a_i$, while for $j > 1$ we have $d_0 u_j^i(t) = 0$. Hence, we can write compactly
\[
\mathfrak{A}'_{(i,j)}(0) = \sum_{\lambda=1}^{d} \mathfrak{A}^\lambda_{(i,j)}.
\]

Then,
\[
(3.11) \quad f_{(i,j)} = s_{(i,j)} \mathfrak{A} \wedge \left( \sum_{\lambda=1}^{d} \mathfrak{A}^\lambda_{(i,j)} \right) \wedge \bigwedge_{i=1}^{d} \bigwedge_{i \neq i} \mathfrak{A}_{(i,j)} = \sum_{1 \leq \lambda \neq i \leq d} \mathfrak{A} \wedge \mathfrak{A}^\lambda_{(i,j)} \wedge \bigwedge_{i=1}^{d} \bigwedge_{i \neq i} \mathfrak{A}_{(i,j)} + s_{(i,j)} \sum_{1 \leq \lambda \neq i \leq d} f_{\lambda_{(i,j)}},
\]
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 $\lambda \neq i$ and $\mu \neq k$ the value
\[
(f_{\lambda_{(i,j)}}, f_{\mu_{(k,l)}}) = \det( B_{\lambda_{(i,j)}, \mu_{(k,l)}}),
\]
where
\[
B_{\lambda_{(i,j)}, \mu_{(k,l)}} := \left[ \mathfrak{A} \mathfrak{A}^\mu_{(k,l)} \right]_{\lambda_{(i,j)}, \mu_{(k,l)}}[1, d]^{d}. \]

Herein, the column vectors should be interpreted as vectorized tensors. Recall that we have $\langle a_i, x_i \rangle = \langle a_i, u_j^i \rangle = 0$. Then, it follows from Lemma 2.1 and direct computation that for $\lambda \neq i$ and $\mu \neq k$, we have
\[
\langle \mathfrak{A}, \mathfrak{A}^\mu_{(k,l)} \rangle = \langle \mathfrak{A}, \mathfrak{A}_{(k,l)} \rangle = 0,
\]
\[
\langle \mathfrak{A}_{(i,j)}, \mathfrak{A}^\mu_{(k,l)} \rangle = \delta_{ik} \delta_{jl} \delta_{\mu \lambda} \|x_i\|^2,
\]
\[
\langle \mathfrak{A}_{(i,j)}, \mathfrak{A}_{(k,l)} \rangle = 0.
\]

We distinguish between two cases. If $(i, j) \neq (k, l)$, $\lambda \neq i$ and $\mu \neq k$, it follows from the above equations that the row of $(B_{\lambda_{(i,j)}, \mu_{(k,l)}})_{i \neq k}$ consisting of
\[
\left[ \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A} \rangle \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A}^\mu_{(k,l)} \rangle \langle \mathfrak{A}^\lambda_{(i,j)}, \mathfrak{A}_{(k,l)} \rangle \right]_{i \neq k}
\]
is a zero row, which implies that \( \langle f_{(i,j),\lambda} \rangle = \langle f_{(k,\ell),\mu} \rangle = 0 \). On the other hand, if \((i, j) = (k, \ell)\), \( \lambda \neq i \) and \( \mu \neq k \), then it follows from the above equations that \( B^T_{(i,j),\lambda} B_{(i,j),\mu} \) is a diagonal matrix, namely

\[
B^T_{(i,j),\lambda} B_{(i,j),\mu} = \text{diag}(1, (\lambda_{(i,j)}, \lambda_{(i,j)}^\mu), 1, \ldots, 1).
\]

Its determinant is then \( \langle \lambda_{(i,j)}, \lambda_{(i,j)}^\mu \rangle = \delta_{\lambda\mu} \|x_1\|^2 \). Therefore,

\[
\langle f_{(i,j),\lambda}^\mu \rangle = \delta_{\lambda\mu} \|x_1\|^2.
\]

Finally, we can compute \( \langle f_{(i,j),\lambda} \langle f_{(k,\ell),\mu} \rangle \rangle \). From (3.11),

\[
\langle f_{(i,j),\lambda} \rangle = s_{(i,j)} s_{(k,\ell)} \left( \sum_{1 \leq \lambda \neq \mu \leq d} f_{(i,j),\lambda}^\mu \sum_{1 \leq \mu \neq k \leq d} f_{(k,\ell),\mu}^\lambda \right) = s_{(i,j)} s_{(k,\ell)} \sum_{1 \leq \lambda \neq \mu \leq d} \delta_{\lambda\mu} \|x_1\|^2,
\]

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

\[
\|f_{(i,j)}\|^2 = s_{(i,j)}^2 \sum_{1 \leq \lambda \neq \mu \leq d} \|x_1\|^2 = \sum_{1 \leq \lambda \neq \mu \leq d} \|x_1\|^2,
\]

proving the result.

\[ \square \]

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 \( 2 \leq r \leq 5 \) we first select an ill-posed tensor decomposition \( A := (\mathcal{A}_1, \ldots, \mathcal{A}_r) \in S^{x^r} \) as explained next. We can randomly generate a rank-1 tensor \( a_1 \otimes \cdots \otimes a_d \) by sampling the elements of \( a_1 \) from a standard normal distribution. Then, \( A \) is generated by randomly sampling the first \( r - 1 \) rank-one tensors \( \mathcal{A}_1, \ldots, \mathcal{A}_{r-1} \in \mathbb{R}^{11 \times 10^{r-5}} \) and then putting \( \mathcal{A}_r := a_1^1 \otimes x_2 \otimes x_3 \), where \( a_1^1 = a_1^2 \otimes a_1^3 \otimes a_1^4 \) and the components of \( x_i \) are sampled from a standard normal distribution. Now,

\[
\mathcal{A}_1 + \mathcal{A}_r = a_1^1 \otimes (a_1^2 \otimes a_1^3 + x_2 \otimes x_3),
\]

and since a rank-2 matrix decomposition is never unique, it follows that \( \mathcal{A}_1 + \mathcal{A}_r \) has at least a 2-dimensional family\(^2\) of decompositions, and, hence, so does \( \mathcal{A}_1 + \cdots + \mathcal{A}_r \). Then, it follows from [6, Corollary 1.2] that \( \kappa(A) = \infty \) and hence \( A \in \Sigma_p \). Finally, we generate a neighboring tensor decomposition \( B := (\mathcal{B}_1, \ldots, \mathcal{B}_r) \in S^{x^r} \) by perturbing \( A \) as follows. Let \( \mathcal{A}_r = a_1^1 \otimes a_2^2 \otimes a_3^3 \), and then set \( \mathcal{B}_r = (a_1^1 + 10^{-2} \cdot x_1^1) \otimes (a_2^2 + 10^{-2} \cdot x_2^2) \otimes (a_3^3 + 10^{-2} \cdot x_3^3) \), where the elements of \( x_i^j \) are randomly drawn from a standard normal distribution.

Denote by \((0, 1) \to S^{x^r}, t \mapsto B_t \) a curve between \( A \) and \( B \) whose length is \( \text{dist}_w(A, B) \). Then, for all \( t \), we have \( \text{dist}_w(B_t, \Sigma_p) \leq \text{dist}_w(A, B_t) \) and hence, by Theorem 1.1,

\[
\frac{1}{\kappa(B_t)} \leq \text{dist}_w(A, B_t).
\]

We expect for small \( t \) that \( \text{dist}_w(A, B_t) \approx \text{dist}_w(A, B) \) 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, \( \text{dist}_w(A, B_t) \) 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 \( A = (\mathcal{A}_1, \ldots, \mathcal{A}_r) \) close to \( \Sigma_p \) the reverse of Theorem 1.1 could hold as well, i.e., \( \text{dist}_w((\mathcal{A}_1, \ldots, \mathcal{A}_r), \Sigma_p) \leq c \frac{1}{\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)} \) for some constant \( c > 0 \) that might depend on \( A \).

\(^{2}\)The fact that the family is at least two-dimensional follows from the fact that the 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].
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