The condition number of join decompositions — **Source link** ⬈

Paul Breiding, Nick Vannieuwenhoven

Institutions: Max Planck Society

Related papers:

- Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem

- Tensors: Geometry and Applications

- The condition number of join decompositions

- Tensor Decompositions and Applications

- Tensor decompositions for learning latent variable models

THE CONDITION NUMBER OF JOIN DECOMPOSITIONS[*]

PAUL BREIDING[†] AND NICK VANNIEUWENHOVEN[‡]

Abstract. The join set of a finite collection of smooth embedded submanifolds of a mutual vector space is defined as their Minkowski sum. Join decompositions generalize some ubiquitous decompositions in multilinear algebra, namely tensor rank, Waring, partially symmetric rank and block term decompositions. This paper examines the numerical sensitivity of join decompositions to perturbations; specifically, we consider the condition number for general join decompositions. It is characterized as a distance to a set of ill-posed points in a supplementary product of Grassmannians. We prove that this condition number can be computed efficiently as the smallest singular value of an auxiliary matrix. For some special join sets, we characterized the behavior of sequences in the join set converging to the latter's boundary points. Finally, we specialize our discussion to the tensor rank and Waring decompositions and provide several numerical experiments confirming the key results.

Key words. join set, join decomposition problem, condition number, tensor rank decomposition, CP decomposition, Waring decomposition, block term decomposition

AMS subject classifications. 49Q12, 65F35, 53B20, 53B21, 14P10, 65Y20, 15A69

1. Introduction. Many subsets of $\mathbb{R}^N$ admit the structure of a *join set*: the join $\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$ of a collection of smooth embedded submanifolds $\mathcal{M}_i \subset \mathbb{R}^N$ is defined as the image of the addition map

$$(1.1) \qquad \Phi : \mathcal{M}_1 \times \ldots \times \mathcal{M}_r \longrightarrow \mathbb{R}^N, \quad (\mathbf{p}_1, \ldots, \mathbf{p}_r) \longmapsto \mathbf{p}_1 + \cdots + \mathbf{p}_r.$$

The class of join sets comprises, among others, tensor canonical rank decompositions, Waring decompositions, partially symmetric rank decompositions, and block term decompositions. For example, the set of $m \times n$ matrices of rank equal to 1 is a manifold that we refer to as the Segre manifold $\mathcal{S}_{m,n}$. Constructing the join set $\sigma_r(\mathcal{S}_{m,n}) := \mathrm{Join}(\mathcal{S}_{m,n}, \ldots, \mathcal{S}_{m,n})$, we obtain exactly the algebraic variety of matrices of rank at most r [33]. This example generalizes to higher-order tensors, as the set of rank-1 tensors in $\mathbb{R}^{m_1} \otimes \cdots \otimes \mathbb{R}^{m_d} \simeq \mathbb{R}^{m_1 \cdots m_d}$ forms the Segre manifold $\mathcal{S} := \mathcal{S}_{m_1, \ldots, m_d}$ [33]. The join set $\sigma_r(\mathcal{S}) := \mathrm{Join}(\mathcal{S}, \ldots, \mathcal{S})$ is then the semi-algebraic set consisting of the tensors of rank at most r [20]. Note that these examples are special because $\mathcal{M}_1 = \cdots = \mathcal{M}_r$, in which case we refer to the join set as an *r-secant* set. Block term decompositions [19] where the summands have different multilinear ranks are an example of a join set that is not a secant set. An example of secant sets not involving tensors is given in algebraic vision; if the focal locus of a certain camera model, called *congruence*, is a curve Y, then the congruence is given by the set of secants $\sigma_2(Y)$ [7, 36].

The *join decomposition problem* (JDP) is a natural computational problem associated with a join set, comprising several well-known tensor decomposition problems as particular instances. Given smooth embedded manifolds $\mathcal{M}_1, \ldots, \mathcal{M}_r \subset \mathbb{R}^N$ and a $\mathbf{p} \in \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$ the JDP is this:

$$(\mathrm{JDP}) \qquad \text{Compute } \mathfrak{p} \in \mathcal{M}_1 \times \ldots \times \mathcal{M}_r \text{ such that } \Phi(\mathfrak{p}) = \mathbf{p}.$$

From our applications in tensor decompositions, we are particularly interested in the case where the JDP has only a finite number of solutions for a given $\mathbf{p} \in \mathcal{J}$. In fact, most of the aforementioned join sets even offer some guarantees of a *unique* solution of

---

[†]Max-Planck Institute for Mathematics in the Sciences Leipzig. (breiding@mis.mpg.de)
 [‡]KU Leuven, Department of Computer Science. (nick.vannieuwenhoven@cs.kuleuven.be)

the JDP, usually up to a permutation of the summands; this is the case for tensor rank decompositions [15–17,22,28,31], Waring decompositions [17,18], partially symmetric decompositions [31] and specific types of block term decompositions [19,43]. These uniqueness properties are often of significant practical value, e.g., for identifying the parameters of certain latent variable models [4,6].

While it happens for theoretical reasons that $\mathbf{p} \in \mathbb{R}^N$ belongs exactly to $\mathcal{J}$, this is an uncommon situation in applications. The reason is that $\mathbf{p}$ is often corrupted by different sources of noise; for example, representation errors due to roundoff are typically incurred when representing $\mathbf{p}$ on a computer. Further sources of error could be measurement errors, numerical approximation errors, accumulation of roundoff errors, and modeling errors. For these reasons, a more common computational problem is the *join approximation problem* (JAP), which consists of finding a point on a join set that is a closest approximation to a given point $\widetilde{\mathbf{p}} \in \mathbb{R}^N$:

(JAP)   Compute $\widetilde{\mathfrak{p}} \in \mathcal{M}_1 \times \ldots \times \mathcal{M}_r$ such that $\widetilde{\mathfrak{p}} \in \underset{\mathfrak{q} \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_r}{\arg\min} \|\Phi(\mathfrak{q}) - \widetilde{\mathbf{p}}\|.$

For now we ignore the question whether the above problem is actually well-posed; see section 4 for our treatment of this issue.

The key question addressed in this paper is the following. If $\mathfrak{p}$ is the decomposition of the true but unobservable model $\mathbf{p} = \Phi(\mathfrak{p})$ and $\widetilde{\mathfrak{p}}$ is the decomposition obtained from solving (approximately) the JAP with an observed $\widetilde{\mathbf{p}} \approx \mathbf{p}$ as input, then what is the relationship between the decompositions $\mathfrak{p}$ and $\widetilde{\mathfrak{p}}$? We focus on the first-order behavior of this relationship which is captured by the *condition number* of the JDP.

Recall that an *absolute condition number* of a function $f : D \to I$ at $x \in D$ is classically defined in numerical analysis [27,42] as the maximum magnification of an infinitesimal input perturbation by the function:

(1.2) $$\kappa[f](x) := \lim_{\epsilon \to 0} \max_{x' \in (B_\epsilon(x) \cap D)} \frac{\|f(x) - f(x')\|_I}{\|x - x'\|_D},$$

where $\|\cdot\|_D$ and $\|\cdot\|_I$ are norms on the domain D and image I respectively, and $B_\epsilon(x)$ is the ϵ-ball about x in the norm $\|\cdot\|_D$. In this paper, we thus seek the condition number of Φ^{-1}. A complication immediately arises because a join decomposition is usually not strictly unique; that is, Φ is not injective. This entails that the condition number would always be ∞ by the above definition. Nevertheless, Φ can still be *locally invertible*; that is, there exists an open neighborhood $\mathcal{N} \subset \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ of $\mathfrak{p}$ and a *local inverse function* $\Phi_\mathfrak{p}^{-1}$ at $\mathfrak{p}$ so that $\Phi_\mathfrak{p}^{-1} \circ \Phi|_\mathcal{N} = \mathrm{id}_\mathcal{N}$. In this case it is still possible to investigate the *local condition number* of the JDP at $\mathfrak{p}$:

(1.3) $$\kappa(\mathfrak{p}, \mathbf{p}) := \begin{cases} \kappa[\Phi_\mathfrak{p}^{-1}](\Phi(\mathfrak{p})) & \text{if there is a local inverse function } \Phi_\mathfrak{p}^{-1}, \\ \infty & \text{otherwise.} \end{cases}$$

where $\mathbf{p} = \Phi(\mathfrak{p})$ and $\Phi_\mathfrak{p}^{-1}(\mathbf{p}) = \mathfrak{p}$. Because $\mathbf{p}$ is implicitly given by $\mathfrak{p}$, we omit $\mathbf{p}$ in the notation (1.3) and in the following we simply write $\kappa(\mathfrak{p})$ instead.

This local condition number captures $\mathfrak{p}$'s sensitivity to structured perturbations[1] of $\mathbf{p} = \Phi(\mathfrak{p})$. In applications where the components of $\mathfrak{p} = (p_1, \ldots, p_r)$ have an interpretation, this condition number is naturally of interest because it bounds the perturbations to the p_i's in the given join decomposition $\mathfrak{p}$. Moreover, (1.3) obeys the infamous rule of thumb of classic numerical analysis:

(1.4) $$\underbrace{\|\mathfrak{p} - \Phi_\mathfrak{p}^{-1}(\mathbf{w})\|}_{\text{forward error}} \lesssim \underbrace{\kappa(\mathfrak{p})}_{\text{condition number}} \cdot \underbrace{\|\Phi(\mathfrak{p}) - \mathbf{w}\|}_{\text{backward error}}$$

---

[1]By a structured perturbation of $\mathbf{p}$ we mean that the perturbation is such that $\widetilde{\mathbf{p}}$ is still a point in the join.

for all $\mathbf{w}$ in a small neighborhood of $\Phi(\mathfrak{p})$. Recall that this rule is one of the main applications of condition numbers in data analysis applications where one usually seeks to estimate the forward error based only on computable information [27, 42].

Following [14, Remark 14.14], the *maximum local condition number* of the JPD at $\mathbf{p}$ is naturally defined as the maximum of the local condition numbers over the fiber of Φ at $\mathbf{p}$: $K(\mathbf{p}) := \max_{\mathfrak{p} \in \Phi^{-1}(\mathbf{p})} \kappa(\mathfrak{p})$. Obtaining meaningful results for this condition number is vastly more complicated as we believe it requires deep understanding of the global geometry of join sets. This lies beyond the scope of this work. Note that in the case of the aforementioned tensor decompositions, which are often unique up to the order of the summands, the condition number (1.3) is the same for all decompositions and hence coincides with the maximum condition number.

Despite the importance of condition numbers in numerical analysis and the interest in specific types of join decompositions, heretofore few results exist on their conditioning. Only recently a local condition number was proposed for the tensor rank decomposition [40]. In this paper, we aspire towards treating JDPs in significant generality using a more elegant approach. Another notion of the stability of the parameters of a tensor rank decomposition is the Cramér–Rao bound; its connection to the condition number of the factor matrices is discussed in [40].

1.1. Contributions. For brevity, we call $\kappa(\mathfrak{p})$ in (1.3) the condition number in the remainder of the paper, its locality being implicit. We derived the condition number of the JDP via a local application of the framework of Bürgisser and Cucker [14]. In the rest of this subsection, let $\mathcal{M}_i$ be an n_i-dimensional embedded submanifold of $\mathbb{R}^N$, and $n := n_1 + \cdots + n_r$. Moreover, let $\mathrm{T}_x \mathcal{M}$ denote the tangent space to a manifold $\mathcal{M}$ at $x \in \mathcal{M}$.

THEOREM 1.1 (Spectral characterization). *Let $\varsigma_n(A)$ denote the nth largest singular value of the linear operator A. The condition number of the JDP at the decomposition $\mathfrak{p} = (p_1, \ldots, p_r) \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ is*

$$(1.5) \qquad \kappa(\mathfrak{p}) = \frac{1}{\varsigma_n(\mathrm{d}_{\mathfrak{p}} \Phi)} = \frac{1}{\varsigma_n(U)},$$

where $\mathrm{d}_{\mathfrak{p}} \Phi : \mathrm{T}_{\mathfrak{p}}(\mathcal{M}_1 \times \cdots \times \mathcal{M}_r) \to \mathbb{R}^N$ is the derivative of the addition map Φ at $\mathfrak{p}$, and $U := \begin{bmatrix} U_1 & \cdots & U_r \end{bmatrix}$ wherein $U_i \in \mathbb{R}^{N \times n_i}$ is an orthonormal basis of $\mathrm{T}_{p_i} \mathcal{M}_i$.

Note that contrary to definition (1.3), the expression derived above is efficiently computable, as it suffices computing the least singular value of a matrix.

An important consequence of Theorem 1.1 is that the condition number $\kappa(\mathfrak{p})$ is unbounded if the join decomposition of $\Phi(\mathfrak{p})$ is not locally unique on a smooth submanifold. Under these conditions, the derivative $\mathrm{d}_{\mathfrak{p}} \Phi$ is not injective, by the contrapositive of [34, Proposition 4.1], so that $\kappa(\mathfrak{p}) = \infty$ by Theorem 1.1.

COROLLARY 1.2. *If there exists a smooth submanifold $\mathcal{E} \subset \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ for which $\Phi(\mathfrak{p}) = \Phi(\mathfrak{q})$ for all $\mathfrak{p}, \mathfrak{q} \in \mathcal{E}$, then every $\mathfrak{p}$ in the interior of $\mathcal{E}$ has $\kappa(\mathfrak{p}) = \infty$.*

Another key contribution of this paper is the characterization of the condition number $\kappa(\mathfrak{p})$ of JDPs at a decomposition $\mathfrak{p}$ as an inverse distance to a set of *ill-posed* inputs in the spirit of Demmel [21].

THEOREM 1.3 (Characterization as inverse distance). *The condition number of the JDP at $\mathfrak{p} = (p_1, \ldots, p_r) \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ is*

$$\kappa(\mathfrak{p}) = \frac{1}{\mathrm{dist}\left((\mathrm{T}_{p_1} \mathcal{M}_1, \ldots, \mathrm{T}_{p_r} \mathcal{M}_r), \Sigma_{\mathrm{Gr}}\right)}$$

where the distance measure is defined as in (3.1), and the ill-posed locus Σ_{Gr} is

$$\Sigma_{\mathrm{Gr}} = \{(W_1, \ldots, W_r) \in \mathrm{Gr}(N, n_1) \times \cdots \times \mathrm{Gr}(N, n_r) \mid \dim(W_1 + \cdots + W_r) < n\}$$

3

with $\mathrm{Gr}(N, k)$ *the Grassmannian of k-dimensional subspaces of $\mathbb{R}^N$.*

The last main contribution concerns the behavior of the condition number near the boundary of a join set. Indeed, many join sets $\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$ are not closed in the Euclidean topology; see Remark 4.1. Consequently, there may exist *convergent* sequences $\mathbf{p}_0, \mathbf{p}_1, \ldots$ with every $\mathbf{p}_i \in \mathcal{J}$ but for which nevertheless $\lim_{i \to \infty} \mathbf{p}_i \to \mathbf{p}_\star \notin \mathcal{J}$. In the context of tensor rank decompositions, de Silva and Lim [20] explained what may be expected of such sequences: for $\mathfrak{p}_i = (p_1^{(i)}, \ldots, p_r^{(i)}) \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ with $\mathbf{p}_i = \Phi(\mathfrak{p}_i)$ some of the summands $p_j^{(i)}$ become of unbounded norm when $\mathbf{p}_i \to \mathbf{p}_\star$. In fact, this result generalizes to join decompositions. Moreover, we show that also *the condition number of the $\mathfrak{p}_i$'s becomes unbounded* in this case. This provides a more rigorous criterion for determining if a sequence $\mathbf{p}_i \in \mathcal{J}$ is converging to a $\mathbf{p}_\star \notin \mathcal{J}$.

THEOREM 1.4. *Assume that* $\mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r) = \mathrm{Join}(\overline{\mathcal{M}_1}, \ldots, \overline{\mathcal{M}_r})$, *where the overline denotes Euclidean closure, and that $\mathcal{M}_i$ is a cone.[2] Let $\mathbf{p}_i(t) \subset \mathcal{M}_i$, $1 \le i \le r$, be a set of r smooth curves simultaneously defined for all $t \in (0, 1)$. If*

$$\lim_{t \to 0}(\mathbf{p}_1(t) + \ldots + \mathbf{p}_r(t)) = \mathbf{p}_\star \notin \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$$

then the condition number grows without bound:

$$\lim_{t \to 0} \kappa(\mathbf{p}_1(t), \ldots, \mathbf{p}_r(t)) \to \infty;$$

moreover, at least two summands diverge:

$$\exists 1 \le i_\star < j_\star \le r : \lim_{t \to 0} \|\mathbf{p}_{i_\star}(t)\| \to \infty \ \textit{and} \ \lim_{t \to 0} \|\mathbf{p}_{j_\star}(t)\| \to \infty.$$

1.2. Outline. The rest of this paper is structured as follows. In section 2 we derive the condition number of the JDP and prove Theorem 1.1. We characterize it as an inverse distance in section 3, establishing Theorem 1.3. The condition number of boundary points of join sets is investigated in section 4, and we prove Theorem 1.4. In section 5 the discussion is specialized to two important join decompositions, namely the Waring and tensor rank decompositions. Some numerical experiments involving tensor rank decomposition problems are presented in section 6. The final section summarizes our main conclusions.

1.3. Notation. We fix some notation for the rest of this article. Vectors are typeset in lowercase boldface letters ($\mathbf{p}$), matrices in uppercase (U), tensors in an uppercase Fraktur font ($\mathfrak{A}$), tuples in a lowercase Fraktur font ($\mathfrak{p}$), and manifolds, varieties and join sets in an uppercase calligraphic font ($\mathcal{M}$).

Let $A \in \mathbb{R}^{m \times n}$. The transpose of A is denoted by A^T, and A's pseudo-inverse is denoted by $A^\dagger$. The ith column of A is denoted by $\mathbf{a}_i$. We denote by $\varsigma_k(A)$ the kth largest singular value of A. Its spectral norm is $\|A\| := \varsigma_1(A)$. The rank of A is $\mathrm{rk}(A)$. The $m \times m$ identity matrix is denoted by I_m.

We define $\mathbb{R}_0^m := \mathbb{R}^m \setminus \{0\}$. The standard inner product on $\mathbb{R}^n$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$, which induces the norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}_0^n$ the angle between $\mathbf{x}$ and $\mathbf{y}$ is $\sphericalangle(\mathbf{x}, \mathbf{y}) := \arccos \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$. The unit sphere in $\mathbb{R}^n$ is denoted by $\mathbb{S}(\mathbb{R}^n) := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$. The ball of radius r around $x \in X$ is $B_r(x) = \{y \mid \|x - y\| \le r\}$, where X is some space and the norm is understood from the context.

Let $\mathcal{X}$ be a smooth manifold. If $F : \mathcal{X} \to \mathbb{R}^N$ is a differentiable function on $\mathcal{X}$, then we denote its derivative at $x \in \mathcal{X}$ by $\mathrm{d}_x F$.

Throughout this paper, $\mathcal{M}_i \subset \mathbb{R}^N$ denotes an n_i-dimensional smooth embedded submanifold of $\mathbb{R}^N$ for $i = 1, \ldots, r$, and $\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r) \subset \mathbb{R}^N$ denotes their

---

[2]We call a set $C \subseteq \mathbb{R}^N$ a cone, if $p \in C$ implies that $\alpha p \in C$ for all $\alpha \in \mathbb{R} \setminus \{0\}$.

join set. The product manifold $\mathcal{M}_1 \times \cdots \times \mathcal{M}_r$ is denoted by $\mathcal{P}$. Its dimension is denoted by $n = \sum_{i=1}^r n_i$. The map $\Phi : \mathcal{P} \to \mathcal{J}$ is defined as in (1.1).

2. The condition number of JDPs. This section derives an efficiently computable expression for the condition number of the JDP at a particular decomposition. As we are dealing with the inverse of a smooth function on a product manifold, the natural starting point of our analysis is the differential-geometric framework of condition [9, 14] that applies to smooth maps between manifolds. However, Φ is not a map between smooth manifolds since $\mathcal{J}$ is not necessarily a manifold. Indeed, in the specific case that the $\mathcal{M}_i$ are smooth *algebraic varieties*, i.e., $\mathcal{M}_i$ is the solution set of a collection of polynomial equations, then the join set is a *semi-algebraic set* by the Tarski–Seidenberg principle [10] as it corresponds to the projection onto the last factor of the graph of Φ, which is an algebraic set. Such sets generally have *singular points* where the geometric tangent space [34, Chapter 3] is not defined. For overcoming this obstacle, the key observation is that we can localize the analysis of [14, Section 14.3], *hereby generalizing the differential geometric framework to local inverses of smooth functions on manifolds.* This is performed in the next subsection. In subsection 2.2, we propose relative condition numbers for the JDP.

2.1. The local differential-geometric approach. The proof of Theorem 1.1 is presented in this subsection. We seek to obtain the condition number of the inverse of a smooth function on a manifold. The first helpful result is stated next.

LEMMA 2.1. *Let $\mathcal{M} \subset \mathbb{R}^N$ be an embedded submanifold. Let $F : \mathcal{M} \to \mathbb{R}^p$ be a smooth function on $\mathcal{M}$. Let $\mathbf{x} \in \mathcal{M}$. There exist constants $r_F > 0$ and $\gamma_F \geq 0$ such that for all $\mathbf{y} \in B_{r_F}(\mathbf{x}) \cap \mathcal{M}$ we have $\Delta = (\mathbf{y} - \mathbf{x}) \in \mathbb{R}^N$ and*

$$F(\mathbf{y}) = F(\mathbf{x}) + (\mathrm{d}_{\mathbf{x}} F) P_{\mathrm{T}_{\mathbf{x}} \mathcal{M}} \Delta + \mathbf{v}_{\mathbf{x}, \mathbf{y}}, \text{ where } \|\mathbf{v}_{\mathbf{x}, \mathbf{y}}\| \leq \gamma_F \|\Delta\|^2,$$

and P_A denotes the orthogonal projection onto the linear subspace $A \subset \mathbb{R}^N$.

Proof. Let $P_{\mathcal{M}} : \mathbb{R}^N \to \mathcal{M}, \mathbf{z} \mapsto \arg\min_{\mathbf{y} \in \mathcal{M}} \|\mathbf{z} - \mathbf{y}\|$ be the projection onto the manifold $\mathcal{M}$, which is a smooth function in the neighborhood of $\mathbf{x} \in \mathcal{M}$ [2, Lemma 4]. By definition we have $\mathbf{x} = P_{\mathcal{M}} \mathbf{x}$ for all $\mathbf{x} \in \mathcal{M}$. By [2, Proposition 5], the map $\widehat{R}_{\mathbf{x}} : \mathbb{R}^N \to \mathcal{M}, \eta \mapsto P_{\mathcal{M}}(\mathbf{x} + \eta)$ restricted to $\mathrm{T}_{\mathbf{x}} \mathcal{M}$ is the *projective retraction*, which is a smooth and well-defined retraction for all $\xi \in \mathrm{T}_{\mathbf{x}} \mathcal{M}$ in a neighborhood of $0_{\mathbf{x}}$. Let $\tau = \|\Delta\|$ and fix $\eta = \Delta \tau^{-1}$. Consider then $G : \mathbb{R} \to \mathbb{R}^N, t \mapsto F(\widehat{R}_{\mathbf{x}}(t\eta))$. By definition it admits a Taylor series approximation

$$G(\tau) = G(0) + \mathrm{d}_0 G \cdot \tau + \mathcal{O}(\tau^2),$$

which is well defined in a neighborhood of 0. We have $G(0) = F(\widehat{R}_{\mathbf{x}}(0)) = F(\mathbf{x})$ and $G(\tau) = F(\widehat{R}_{\mathbf{x}}(\Delta)) = F(P_{\mathcal{M}} \mathbf{y}) = F(\mathbf{y})$. Moreover, it follows from the chain rule that $\mathrm{d}_0 G = \mathrm{d}_{\widehat{R}_{\mathbf{x}}(0)} F \circ \mathrm{d}_0 \widehat{R}_{\mathbf{x}} \circ \eta = \mathrm{d}_{\mathbf{x}} F \circ (\mathrm{d}_{\mathbf{x}} P_{\mathcal{M}}) \circ \eta$. Here η also denotes the map $t \mapsto t\eta$. As $\mathrm{d}_{\mathbf{x}} P_{\mathcal{M}} = P_{\mathrm{T}_{\mathbf{x}} \mathcal{M}}$ by [2, Lemma 4], we have $\mathrm{d}_0 G = \mathrm{d}_{\mathbf{x}} F \circ P_{\mathrm{T}_{\mathbf{x}} \mathcal{M}} \circ \eta$. $\qquad \square$

The next ingredient we need is a standard fact in differential geometry.

LEMMA 2.2. *Let $F : \mathcal{M} \to \mathcal{N}$ be a smooth map between manifolds $\mathcal{M}$ and $\mathcal{N}$. If the derivative $\mathrm{d}_x F : \mathrm{T}_x \mathcal{M} \to \mathrm{T}_{F(x)} \mathcal{N}$ is injective at $x \in \mathcal{M}$, then F is a local diffeomorphism onto its image.*

Proof. Combine Proposition 4.1, Theorem 4.25 and Proposition 5.2 of [34]. $\qquad \square$

This lemma essentially proves the case in Theorem 1.1 where Φ has no unique local inverse, i.e., where the JDP is ill-posed.

COROLLARY 2.3. *If Φ is not locally invertible at $\mathfrak{p} \in \mathcal{P}$, then $\kappa(\mathfrak{p}) = \infty = \frac{1}{\varsigma_n(\mathrm{d}_{\mathfrak{p}} \Phi)}$.*

Proof. If $n = \dim \mathcal{P} > N$, then Φ is not a local homeomorphism and hence $\varsigma_n(\mathrm{d}_{\mathfrak{p}}\Phi) = 0$. Otherwise, the contrapositive of Lemma 2.2 entails that $\mathrm{d}_{\mathfrak{p}}\Phi$ cannot be injective. Since $n \leq N$ in this case, $\mathrm{d}_{\mathfrak{p}}\Phi$ is not of full rank n. $\qquad\square$

We are now ready to prove the general case. The proof strategy is quite standard.

Proof of Theorem 1.1. Let $\mathfrak{p} \in \mathcal{P}$. If Φ is not locally invertible at $\mathfrak{p}$, then by Corollary 2.3, $\kappa(\mathfrak{p}) = \infty$. Note that this includes the case $n > N$, hence we can assume $n \leq N$ in the remainder. It remains to prove Theorem 1.1 in the case when there exists a local inverse function $\Phi_{\mathfrak{p}}^{-1}$ of Φ at $\mathbf{p} = \Phi(\mathfrak{p})$. In this case, there exists some open neighborhood $\mathcal{N}$ of $\mathfrak{p} \in \mathcal{P}$ such that $\Phi_{\mathfrak{p}}^{-1} \circ \Phi|_{\mathcal{N}} = \mathrm{id}_{\mathcal{N}}$. Assume first that $\mathrm{d}_{\mathfrak{p}}\Phi$ is injective. It follows from Lemma 2.2 that Φ is a local diffeomorphism onto its image, i.e., there exists some open neighborhood $\mathcal{N}' \subset \mathcal{N}$ of $\mathfrak{p}$ such that $\mathcal{J}' := \Phi(\mathcal{N}') \subset \mathcal{J}$ is an embedded submanifold of $\mathcal{J}$ in the subspace topology. We may assume without loss of generality that the domain of $\Phi_{\mathfrak{p}}^{-1}$ is this manifold $\mathcal{J}'$. Since $\Phi|_{\mathcal{N}'}$ is diffeomorphic to $\mathcal{J}'$, it follows that $\dim \mathcal{N}' = \dim \mathcal{J}'$ and that $\mathrm{d}_{\mathfrak{p}}\Phi$ is invertible. By applying Lemma 2.1 to $\Phi_{\mathfrak{p}}^{-1}$, we find

$$(2.1) \quad \kappa(\mathfrak{p}) = \kappa[\Phi_{\mathfrak{p}}^{-1}](\mathbf{p}) = \lim_{\epsilon \to 0} \max_{\mathbf{y} \in (B_\epsilon(\mathbf{p}) \cap \mathcal{J}')} \frac{\|\mathrm{d}_{\mathbf{p}}\Phi_{\mathfrak{p}}^{-1} P_{\mathrm{T}_{\mathbf{p}}\mathcal{J}'}(\mathbf{y} - \mathbf{p})\|}{\|\mathbf{y} - \mathbf{p}\|} + \mathcal{O}(\|\mathbf{y} - \mathbf{p}\|).$$

For sufficiently small ϵ, $B_\epsilon(\mathbf{p}) \cap \mathcal{J}'$ is a path-connected submanifold. Then, for every $\mathbf{y} \in B_\epsilon(\mathbf{p}) \cap \mathcal{J}'$ there exist at least one smooth curve $\gamma_{\mathbf{p} \to \mathbf{y}}(t)$ in $\mathcal{J}'$ connecting $\mathbf{p}$ and $\mathbf{y}$. Applying Lemma 2.1 to this curve, and plugging the result into (2.1), we find

$$(2.2) \quad \kappa(\mathfrak{p}) = \lim_{\epsilon \to 0} \max_{\mathbf{y} \in (B_\epsilon(\mathbf{p}) \cap \mathcal{J}')} \left\| \mathrm{d}_{\mathbf{p}}\Phi_{\mathfrak{p}}^{-1} \frac{\mathrm{d}_0 \gamma_{\mathbf{p} \to \mathbf{y}}}{\|\mathrm{d}_0 \gamma_{\mathbf{p} \to \mathbf{y}}\|} \right\| + \mathcal{O}(\|\mathbf{y} - \mathbf{p}\|).$$

Since the geometric tangent space is precisely

$$\mathrm{T}_{\mathbf{p}}\mathcal{J}' := \left\{ \mathrm{d}_0 \gamma \in \mathbb{R}^N \mid \gamma(t) \subset \mathcal{J}' \text{ is a smooth curve with } \gamma(0) = \mathbf{p} \right\},$$

see, e.g., [34, Proposition 3.23], equation (2.2) is equivalent to

$$\kappa(\mathfrak{p}) = \max_{\mathbf{v} \in \mathbb{S}(\mathrm{T}_{\mathbf{p}}\mathcal{J}')} \|\mathrm{d}_{\mathbf{p}}\Phi_{\mathfrak{p}}^{-1}\mathbf{v}\| = \|\mathrm{d}_{\mathbf{p}}\Phi_{\mathfrak{p}}^{-1}\| = \|(\mathrm{d}_{\mathfrak{p}}\Phi)^{-1}\| = \frac{1}{\varsigma_n(\mathrm{d}_{\mathfrak{p}}\Phi)},$$

where the penultimate step is by invoking the inverse function theorem for manifolds [34, Theorem 4.5]. It remains to prove the case where $\mathrm{d}_{\mathfrak{p}}\Phi$ is not injective. Then, some unit-length vector $\mathbf{v} \in \mathrm{T}_{\mathfrak{p}}\mathcal{P}$ is mapped to 0. Let $\gamma(t) \subset \mathcal{N}$ be a smooth curve with $\gamma(0) = \mathfrak{p}$ and $\mathrm{d}_0\gamma = \mathbf{v}$. From Lemma 2.1 it follows that

$$(2.3) \quad \Phi(\gamma(t)) - \Phi(\gamma(0)) = (\mathrm{d}_{\mathfrak{p}}\Phi \circ P_{\mathrm{T}_{\mathfrak{p}}\mathcal{P}})(\gamma(t) - \gamma(0)) + \mathcal{O}(\|\gamma(t) - \gamma(0)\|^2).$$

Therefore, for sufficiently small t,

$$\kappa(\mathfrak{p}) \geq \lim_{t \to 0} \frac{\|\Phi_{\mathfrak{p}}^{-1}(\Phi(\gamma(t))) - \Phi_{\mathfrak{p}}^{-1}(\Phi(\gamma(0)))\|}{\|\Phi(\gamma(t)) - \Phi(\gamma(0))\|} = \lim_{t \to 0} \left(\frac{\|\Phi(\gamma(t)) - \Phi(\gamma(0))\|}{\|\gamma(t) - \gamma(0)\|} \right)^{-1},$$

where the last step is because $\Phi_{\mathfrak{p}}^{-1}$ is a local inverse of Φ. Plugging (2.3) into the last expression yields $\kappa(\mathfrak{p}) \geq \|(\mathrm{d}_{\mathfrak{p}}\Phi \circ P_{\mathrm{T}_{\mathfrak{p}}\mathcal{P}})\mathbf{v}\|^{-1} = \|\mathrm{d}_{\mathfrak{p}}\Phi\mathbf{v}\|^{-1} = \infty = \frac{1}{\varsigma_n(\mathrm{d}_{\mathfrak{p}}\Phi)}$, where the last equality is precisely because the derivative is not injective. This proves the first equality in (1.5) for all $\mathfrak{p} \in \mathcal{P}$.

Finally, we show that $\varsigma_n(\mathrm{d}_{\mathfrak{p}}\Phi) = \varsigma_n(U)$ for all $\mathfrak{p} \in \mathcal{P}$. The derivative of the addition map $\Phi : \mathcal{P} \to \mathcal{J} \subset \mathbb{R}^N$ at $\mathfrak{p} = (p_1, \ldots, p_r)$ is

$$\mathrm{d}_{\mathfrak{p}}\Phi : \mathrm{T}_{p_1}\mathcal{M}_1 \times \cdots \times \mathrm{T}_{p_r}\mathcal{M}_r \to \mathrm{T}_{\Phi(\mathfrak{p})}\mathbb{R}^N, \quad (\dot{\mathbf{p}}_1, \ldots, \dot{\mathbf{p}}_r) \mapsto \dot{\mathbf{p}}_1 + \cdots + \dot{\mathbf{p}}_r.$$

Letting U_i be as in the theorem, we can uniquely express $\dot{\mathbf{p}}_i = U_i \mathbf{x}_i$. Hence,

$$\mathrm{d}_\mathfrak{p}\Phi(\dot{\mathbf{p}}_1,\ldots,\dot{\mathbf{p}}_r) = U_1\mathbf{x}_1 + \cdots + U_r\mathbf{x}_r = \begin{bmatrix} U_1 & \cdots & U_r \end{bmatrix} [\mathbf{x}_i]_{i=1}^r = U\mathbf{x}.$$

The claim follows from the Courant–Fisher characterization of the least singular value, namely $\varsigma_n(\mathrm{d}_\mathfrak{p}\Phi) = \min_{\mathfrak{r}\in\mathbb{S}(\mathrm{T}_\mathfrak{p}\mathcal{P})}\|(\mathrm{d}_\mathfrak{p}\Phi)(\mathfrak{r})\| = \min_{\mathbf{x}\in\mathbb{S}(\mathbb{R}^n)}\|U\mathbf{x}\| = \varsigma_n(U)$. Note that the second equality is valid only for U_i's with orthonormal columns. $\square$

2.2. Relative condition numbers. *Relative condition numbers* can be obtained as follows. Let $\mathfrak{p} = (\mathbf{p}_1,\ldots,\mathbf{p}_r)$ and $\mathfrak{p}' = (\mathbf{p}_1',\ldots,\mathbf{p}_r')$ with $\mathbf{p}_i, \mathbf{p}_i' \in \mathcal{M}_i$. Then, it follows from (1.4) that

$$\|\mathbf{p}_j - \mathbf{p}_j'\| \le \left\|\begin{bmatrix} \mathbf{p}_1 - \mathbf{p}_1' & \cdots & \mathbf{p}_r - \mathbf{p}_r' \end{bmatrix}\right\|_F \lesssim \kappa(\mathfrak{p}) \cdot \|\Phi(\mathfrak{p}) - \Phi(\mathfrak{p}')\|,$$

provided that $\mathfrak{p}$ and $\mathfrak{p}'$ are close in the product metric on $\mathcal{P}$. Hence, a relative condition number κ_j^{rel} for the jth component of the JDP at $\mathfrak{p}$ could be defined as

$$\kappa_j^{\mathrm{rel}}(\mathfrak{p}) := \kappa(\mathfrak{p})\frac{\|\Phi(\mathfrak{p})\|}{\|\mathbf{p}_j\|}, \quad \text{so that} \quad \frac{\|\mathbf{p}_j - \mathbf{p}_j'\|}{\|\mathbf{p}_j\|} \lesssim \kappa_j^{\mathrm{rel}}(\mathfrak{p})\frac{\|\Phi(\mathfrak{p}) - \Phi(\mathfrak{p}')\|}{\|\Phi(\mathfrak{p})\|}.$$

Remark 2.4. One could complimentarily define a relative condition number for the entire decomposition $\mathfrak{p}$: $\kappa^{\mathrm{rel}}(\mathfrak{p}) := \kappa(\mathfrak{p})\frac{\|\Phi(\mathfrak{p})\|}{\|\mathfrak{p}\|}$. However, we believe that the former definition will usually be the more informative one, especially when the norms of the rank-1 terms appearing in the decomposition $\mathfrak{p}$ are of different orders of magnitude. Consider the following example. Assume that $\mathfrak{p} = (\mathbf{p}_1, \mathbf{p}_2)$ with $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^N$, $\kappa(\mathfrak{p}) = 1$, $\|\mathbf{p}_1\| = 1$ and $\|\mathbf{p}_2\| = 10^{-10}$.[3] Then, $\kappa^{\mathrm{rel}} \approx 1\frac{1}{1} = 1$. One might conclude that the whole decomposition is well-conditioned. If we have an alternative decomposition $\mathfrak{p}'$ with $\|\mathfrak{p} - \mathfrak{p}'\| = 10^{-10}$, then the relative forward error is at most about 10^{-10}. Interpreting this result, it follows that $\mathfrak{p}'$ could well be $(\mathbf{p}_1, 0)$, which has a positive-dimensional family of decompositions, namely $\mathfrak{p}_\alpha = (\alpha\mathbf{p}_1, (1-\alpha)\mathbf{p}_1)$. All of these decompositions have $\kappa(\mathfrak{p}_\alpha) = \infty$. This does not comply well with the intuition that a small relative forward error should imply a numerically stable decomposition. By contrast, the separate relative condition numbers accurately detect that $\mathbf{p}_2$ is not at all stable to small perturbations, as $\kappa_1^{\mathrm{rel}} \approx 1\frac{1}{1} = 1$ and $\kappa_2^{\mathrm{rel}} \approx 1\frac{1}{10^{-10}} = 10^{10}$.

3. Characterization as a distance to ill-posedness. Demmel [21] showed that many condition numbers arising in linear algebra can be interpreted as an inverse distance to a locus of ill-posed inputs to the problem. In this section, we show that the condition number of the JDP is an inverse distance to a locus of ill-posed problems.

Recall that $\mathrm{Gr}(N, m)$ is the Grassmannian manifold of m-dimensional subspaces of $\mathbb{R}^N$. The *projection distance* [24, Section 2.6] between $W, W' \in \mathrm{Gr}(N, m)$ is

$$\mathrm{dist}(W, W') := \|\Pi_W - \Pi_{W'}\|,$$

where Π_W is the orthogonal projection onto W. For brevity, we introduce the next shorthand for a product Grassmannian:

$$\mathrm{Gr}(N, \mathbf{n}) := \mathrm{Gr}(N, n_1) \times \mathrm{Gr}(N, n_2) \times \cdots \times \mathrm{Gr}(N, n_r),$$

where $n_i = \dim \mathcal{M}_i$. The projection distance between $W = (W_1, \ldots, W_r) \in \mathrm{Gr}(N, \mathbf{n})$ and $W' = (W_1', \ldots, W_r') \in \mathrm{Gr}(N, \mathbf{n})$ is then defined as

$$(3.1) \qquad \mathrm{dist}(W, W') := \sqrt{\sum_{i=1}^r \mathrm{dist}(W_i, W_i')^2}.$$

---

[3]Such cases can occur for weak 3-orthogonal tensor rank decompositions; see subsection 5.1.

The distance between a point W and a set $\mathcal{W}$ is defined to be the infimum of the distances from W to any of the elements of the set $\mathcal{W}$.

Let the *locus of intersecting subspaces* in $\mathrm{Gr}(N, \mathbf{n})$ be

$$(3.2) \qquad \Sigma_{\mathrm{Gr}} := \{(W_1, \ldots, W_r) \in \mathrm{Gr}(N, \mathbf{n}) \mid \dim(W_1 + \cdots + W_r) < n\}.$$

We can then define the locus of *ill-posed solutions* of the JDP as

$$\begin{aligned}
\Sigma := \{\mathfrak{p} \in \mathcal{P} \mid \kappa(\mathfrak{p}) = \infty\} &= \{\mathfrak{p} \in \mathcal{P} \mid \mathrm{d}_{\mathfrak{p}}\Phi \text{ is not injective}\} \\
&= \{\mathfrak{p} \in \mathcal{P} \mid \mathrm{rk}\,(\mathrm{d}_{\mathfrak{p}}\Phi) < n\} \\
&= \{\mathfrak{p} = (p_1, \ldots, p_r) \in \mathcal{P} \mid (\mathrm{T}_{p_1}\mathcal{M}_1, \ldots, \mathrm{T}_{p_r}\mathcal{M}_r) \in \Sigma_{\mathrm{Gr}}\},
\end{aligned}$$

where the second and third equalities follow from the proof of Theorem 1.1 and the last by Theorem 1.3.

Remark 3.1. If $\Sigma = \mathcal{P}$ with $\dim \mathcal{P} \leq N$, then we call $\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$ *defective* in analogy with defective join and secant varieties of algebraic varieties [26, Example 11.22]. For join sets originating from Segre and Veronese varieties it is known that defective join sets occur only exceptionally; see [1, 3] for details. By definition, the condition number for elements of defective join sets is ∞.

We need two auxiliary results for proving Theorem 1.3. Let $\mathcal{S}^{N \times r} := \mathbb{S}(\mathbb{R}^N)^r \subset \mathbb{R}^{N \times r}$ denote the subset of matrices whose columns are of unit norm, and $\mathcal{S}^{N \times r}_{<r}$ will denote the subset of $\mathcal{S}^{N \times r}$ of matrices whose rank is strictly less than r. The first lemma is an alternative characterization of the smallest singular value of a matrix with unit-norm columns.

LEMMA 3.2. *Let $r \leq N$, and let $Y \in \mathcal{S}^{N \times r}$. Then,*

$$(3.3) \qquad \varsigma_r(Y) = \min_{\substack{X \in \mathbb{R}^{N \times r}, \\ \mathrm{rk}\,(X) < r}} \sqrt{\sum_{i=1}^{r} \|\mathbf{x}_i - \mathbf{y}_i\|^2} = \min_{\substack{X \in \mathbb{R}^{N \times r}, \\ \mathrm{rk}\,(X) < r}} \sqrt{\sum_{i=1}^{r} \bigl(\sin \sphericalangle(\mathbf{x}_i, \mathbf{y}_i)\bigr)^2}.$$

Proof. The first equality is by the Eckart-Young characterization of the smallest singular value. Let X be a minimizer of the middle expression in (3.3). For every $1 \leq i \leq r$, we distinguish between two cases. If $\mathbf{x}_i \neq 0$, let $\mathbf{x}'_i = \alpha \mathbf{x}_i$ denote the orthogonal projection of $\mathbf{y}_i$ onto $\mathbf{x}_i$. By definition, $\|\mathbf{x}_i - \mathbf{y}_i\| \geq \|\mathbf{x}'_i - \mathbf{y}_i\| = \sin \sphericalangle(\mathbf{x}_i, \mathbf{y}_i)$. Otherwise, if $\mathbf{x}_i = 0$, we choose $\mathbf{x}'_i$ as any nonzero vector in $\mathrm{span}\,\{\mathbf{x}_1, \ldots, \mathbf{x}_r\} \cap (\mathbf{y}_i)^\perp$. Then, $\|\mathbf{x}_i - \mathbf{y}_i\| = 1 = \sin \sphericalangle(\mathbf{x}'_i, \mathbf{y}_i)$ by construction. In both cases, the column span of X does not change if we replace the ith column by $\mathbf{x}'_i$. Hence, $X' = [\mathbf{x}'_1 \cdots \mathbf{x}'_r]$ is of rank $< r$. Hence, $\varsigma_r(Y)$ is bounded from below by the right-hand side of (3.3).

Conversely, let $X' = [\mathbf{x}'_1 \cdots \mathbf{x}'_r]$ be a minimizer of the right-hand side of (3.3). Let $\mathbf{x}_i$ be the orthogonal projection of $\mathbf{y}_i$ onto $\mathbf{x}'_i$. Then $X = [\mathbf{x}_1 \cdots \mathbf{x}_r]$ is of rank $< r$ with $\|\mathbf{x}_i - \mathbf{y}_i\| = \sin \sphericalangle(\mathbf{x}'_i, \mathbf{y}_i)$, proving the converse inequality. $\square$

The second auxiliary lemma sufficiently characterizes Σ_{Gr} for our purpose.

LEMMA 3.3. *Let $n = n_1 + \cdots + n_r$. For $X \in \mathcal{S}^{N \times r}$, we define*

$$\mathcal{E}_X := \{(W_1, \ldots, W_r) \in \mathrm{Gr}(N, \mathbf{n}) \mid \forall 1 \leq i \leq r : \mathbf{x}_i \in W_i\}.$$

Then, the following statements hold:
1. Σ_{Gr} *is the union of $\mathcal{E}_X$ over all $X \in \mathcal{S}^{N \times r}_{<r}$ of rank strictly less than r; and*
2. Σ_{Gr} *is an algebraic subvariety of $\mathrm{Gr}(N, \mathfrak{n})$.*

Proof. Part (1) is a direct consequence of the observation

$$W \in \Sigma_{\mathrm{Gr}} \text{ iff } \exists X \in \mathcal{S}^{N \times r}_{<r} : W \in \mathcal{E}_X, \quad \text{and} \quad \forall X \in \mathcal{S}^{N \times r}_{<r} : \mathcal{E}_X \subset \Sigma_{\mathrm{Gr}}.$$

To prove (2) we mildly generalize [26, Example 8.30]: Since $\mathrm{Gr}(N, m) \simeq \mathbb{P}(\wedge^m \mathbb{R}^N)$ via the Plücker embedding [23, Chap. 3.1.C], we see that

$$\Sigma_{\mathrm{Gr}} = \left\{ ([\mathbf{w}_{1,1} \wedge \cdots \wedge \mathbf{w}_{1,n_1}], \ldots, [\mathbf{w}_{r,1} \wedge \cdots \wedge \mathbf{w}_{r,n_r}]) \mid 0 = \bigwedge_{i=1}^{r} \bigwedge_{j=1}^{n_i} \mathbf{w}_{i,j} \right\} \subset \mathrm{Gr}(N, \mathfrak{n})$$

is a projective subvariety cut out set-theoretically by the above equation. $\qquad \square$

An important consequence of Lemma 3.3(2) is that Σ_{Gr} is closed in the Euclidean topology. Hence, for all $W \in \mathrm{Gr}(N, \mathbf{n})$ there exists a closest $W' \in \Sigma_{\mathrm{Gr}}$ with $\mathrm{dist}(W, W') = \mathrm{dist}(W, \Sigma_{\mathrm{Gr}})$; that is, the infimum is attained.

Now we can prove the next result which essentially entails Theorem 1.3.

PROPOSITION 3.4. *Let $W = (W_1, \ldots, W_r) \in \mathrm{Gr}(N, \mathbf{n})$. Let U_i be a matrix whose columns form an orthonormal basis for W_i, and set $U = \begin{bmatrix} U_1 & \cdots & U_r \end{bmatrix} \in \mathbb{R}^{N \times n}$. Then, we have $\mathrm{dist}(W, \Sigma_{\mathrm{Gr}}) = \varsigma_n(U)$.*

Proof. We show first that $\mathrm{dist}(W, \Sigma_{\mathrm{Gr}}) \leq \varsigma_n(U)$. It is easy to verify that

$$(3.4) \qquad \varsigma_n(U) := \min_{\hat{\mathbf{z}} \in \mathbb{S}(\mathbb{R}^n)} \|U\hat{\mathbf{z}}\| = \min_{\mathbf{w}_i \in \mathbb{S}(W_i), \, 1 \leq i \leq r} \min_{\mathbf{z} \in \mathbb{S}(\mathbb{R}^r)} \|z_1 \mathbf{w}_1 + \cdots + z_r \mathbf{w}_r\|$$
$$= \min_{\mathbf{w}_i \in \mathbb{S}(W_i), \, 1 \leq i \leq r} \varsigma_r \left(\begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_r \end{bmatrix} \right).$$

Let $Y^\star = \begin{bmatrix} \mathbf{y}_1^\star & \cdots & \mathbf{y}_r^\star \end{bmatrix}$ with $\mathbf{y}_i^\star \in \mathbb{S}(W_i)$ be a minimizer of the last statement in the above expression, so that $\varsigma_n(U) = \varsigma_r(Y^\star)$. From the Eckart–Young theorem, we find

$$\varsigma_n(U) = \varsigma_r(Y^\star) = \min_{X \in \mathbb{R}^{N \times r}, \, \mathrm{rk}(X) < r} \|X - Y^\star\|;$$

let $X^\star$ be an optimizer of this expression. Then,

$$\varsigma_n(U) = \varsigma_r(Y^\star) \geq \min_{\substack{\mathbf{w}_i \in \mathbb{S}(W_i), \\ 1 \leq i \leq r}} \sqrt{\sum_{i=1}^{r} \|\mathbf{x}_i^\star - \mathbf{w}_i\|^2} \geq \min_{\substack{\mathbf{w}_i \in \mathbb{S}(W_i), \\ 1 \leq i \leq r}} \sqrt{\sum_{i=1}^{r} \|\Pi_{\mathbf{x}_i^\star}(\mathbf{w}_i) - \mathbf{w}_i\|^2}.$$

From [5, Lemma 3.2], we have

$$(3.5) \qquad \mathrm{dist}(W_i, \mathcal{E}_{\mathbf{x}_i^\star}) = \min_{\mathbf{w}_i \in \mathbb{S}(W_i)} \sin \sphericalangle(\mathbf{x}_i^\star, \mathbf{w}_i) = \min_{\mathbf{w}_i \in \mathbb{S}(W_i)} \|\Pi_{\mathbf{x}_i^\star}(\mathbf{w}_i) - \mathbf{w}_i\|,$$

where $\mathcal{E}_{\mathbf{x}_i^\star} := \{ W' \in \mathrm{Gr}(N, n_i) \mid \mathbf{x}_i^\star \in W' \}$. Hence, by the definition of $\mathrm{dist}(W, \mathcal{E}_X)$ and $\mathcal{E}_X$, one finds $\varsigma_r(Y^\star) \geq \mathrm{dist}(W, \mathcal{E}_X)$. Since $\mathcal{E}_X \subset \Sigma_{\mathrm{Gr}}$ by Lemma 3.3(1), we have proven the first bound: $\mathrm{dist}(W, \Sigma_{\mathrm{Gr}}) \leq \mathrm{dist}(W, \mathcal{E}_X) \leq \varsigma_n(U)$.

For proving the converse inequality, let $W' = (W_1', \ldots, W_r') \in \Sigma_{\mathrm{Gr}}$ be such that $\mathrm{dist}(W, \Sigma_{\mathrm{Gr}}) = \mathrm{dist}(W, W')$. By Lemma 3.3(1) there exists a matrix $X \in \mathcal{S}_{<r}^{N \times r}$, such that $W' \in \mathcal{E}_X$. By definition (3.1) we have

$$\mathrm{dist}(W, \Sigma_{\mathrm{Gr}}) = \mathrm{dist}(W, W') = \left(\sum_{i=1}^{r} \|\Pi_{W_i} - \Pi_{W_i'}\|^2 \right)^{\frac{1}{2}} \geq \left(\sum_{i=1}^{r} \|(\Pi_{W_i} - \Pi_{W_i'})\mathbf{x}_i\|^2 \right)^{\frac{1}{2}}.$$

For $1 \leq i \leq r$, let $\mathbf{w}_i = \Pi_{W_i} \mathbf{x}_i$ and $\mathbf{y}_i := \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \in W_i$ if $\mathbf{w}_i \neq 0$, or, if $\mathbf{w}_i = 0$, let $\mathbf{y}_i$ denote some arbitrary but fixed point in $\mathbb{S}(W_i)$. Then for each i, we have $\|(\Pi_{W_i} - \Pi_{W_i'})\mathbf{x}_i\| = \|\mathbf{w}_i - \mathbf{x}_i\| = \sin \sphericalangle(\mathbf{x}_i, \mathbf{w}_i) = \sin \sphericalangle(\mathbf{x}_i, \mathbf{y}_i)$, and, hence,

$$\mathrm{dist}(W, \Sigma_{\mathrm{Gr}}) \geq \left(\sum_{i=1}^{r} (\sin \sphericalangle(\mathbf{x}_i, \mathbf{y}_i))^2 \right)^{\frac{1}{2}} \geq \varsigma_r \left(\begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_r \end{bmatrix} \right) \geq \varsigma_n(U),$$

where the second step is due to Lemma 3.2 and the last step because of (3.4). $\qquad \square$

We are now ready to wrap up the proof of Theorem 1.3.

Proof of Theorem 1.3. Let $\mathfrak{p} = (p_1, \ldots, p_r) \in \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$. Applying Proposition 3.4 to $W := (\mathrm{T}_{p_1}\mathcal{M}_1, \ldots, \mathrm{T}_{p_r}\mathcal{M}_r) \in \mathrm{Gr}(N, \mathbf{n})$ concludes the proof. □

4. The condition number of open boundary points. As mentioned in the introduction, in applications one seeks to solve the JAP $\min_{\mathfrak{p} \in \mathcal{P}} \frac{1}{2}\|\Phi(\mathfrak{p}) - \widetilde{\mathbf{p}}\|^2$, where $\widetilde{\mathbf{p}} \in \mathbb{R}^N$ and $\mathcal{P} := \mathcal{M}_1 \times \cdots \times \mathcal{M}_r$. Since a closed solution of this optimization problem is unknown, iterative optimization methods are usually employed.

A serious complication occurs in these methods when the JAP has no minimizer. Let $\mathcal{J} \subset \mathbb{R}^N$ be a join set, and let $\overline{\mathcal{J}}$ denote its closure in the Euclidean topology. We call $\mathcal{B} := \overline{\mathcal{J}} \setminus \mathcal{J}$ the set of *open boundary points* of $\mathcal{J}$. Denote by $\Xi \subset \mathbb{R}^N$ the set of all inputs to the JAP such that the infimum cannot be attained; they are ill-posed inputs. It is easy to show that $\mathcal{B} \subset \Xi$, and $\Xi = \emptyset$ if and only if $\mathcal{B} = \emptyset$. If $\mathbf{p} \in \Xi$ is an ill-posed input, then there is no solution to the JAP, but aforementioned optimization methods will nevertheless produce an approximate minimizer $\mathfrak{p}_\star \in \mathcal{P}$. While the backward error $\|\Phi(\mathfrak{p}_\star) - \mathbf{p}\|$ may be small, the components $\mathfrak{p}_\star$ of the computed approximate solution usually admit no interpretation, since $\mathbf{p} \notin \mathcal{J}$ does not satisfy the model!

Remark 4.1. The existence of open boundary points is not an exceptional phenomenon. Consider the case of join sets of projective varieties. Chapter 10 of [33] contains examples of join sets $\mathcal{J} = \mathrm{Join}(\mathcal{M}, \ldots, \mathcal{M})$ where the closure of $\mathcal{J}$ in the Euclidean topology, write $\overline{\mathcal{J}}$, is strictly larger than $\mathcal{J}$; namely, this situation arises for tensor rank decompositions, Waring decompositions, and partially symmetric decompositions. In the context of secant varieties, points in $\overline{\mathcal{J}} \setminus \mathcal{J}$ are said to admit an $\mathcal{M}$-*border rank* [33, Section 5.2] that is strictly less than their rank.

This section is devoted to proving Theorem 1.4 which states that all decompositions $\mathfrak{p}_\star \in \mathcal{P}$ whose corresponding point in the join set $\mathbf{p}_\star = \Phi(\mathfrak{p}_\star)$ is close to the locus of ill-posed inputs Ξ must necessarily have a large condition number. This can then be interpreted in light of (1.4), which would reveal that $\mathfrak{p}_\star$ is not a stable decomposition. The condition number of the computed approximation $\mathfrak{p}_\star$ is thus always a reliable measure of stability, even for ill-posed inputs Ξ of the JAP.

Unfortunately, we could not prove Theorem 1.4 without additional assumptions on the manifolds $\mathcal{M}_1, \ldots, \mathcal{M}_r$ and their join set $\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$. The two extra assumptions that we make are that all of the manifolds $\mathcal{M}_i$ are *cones*, and that

$$\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r) = \mathrm{Join}(\overline{\mathcal{M}}_1, \ldots, \overline{\mathcal{M}}_r),$$

where the overline denotes the Euclidean closure. Recall that we call a set $C \subset \mathbb{R}^N$ a cone if for every $p \in C$ also $\alpha p \in C$ for all $\alpha \in \mathbb{R}_0$.

It is natural to wonder whether aforementioned assumptions are necessary. We give two examples[4] of join sets where one of these assumptions is not satisfied, and show that the conclusion of Theorem 1.4 does not hold. The first example features an open boundary point for which the condition number converges to a finite value.

Example 4.2. Consider the following smooth manifolds embedded in $\mathbb{R}^3$:

$$\mathcal{M}_1 = (\mathbb{R}\setminus\{0\}) \times \mathbb{R} \times \{0\} \quad \text{and} \quad \mathcal{M}_2 = \{0\} \times \{0\} \times \mathbb{R}.$$

These manifolds violate the second assumption because $(0, 1, 1)$ is in $\mathrm{Join}(\overline{\mathcal{M}}_1, \overline{\mathcal{M}}_2)$ but not in $\mathrm{Join}(\mathcal{M}_1, \mathcal{M}_2)$.

Let $p_1(t) = (\sin(t), \cos(t), 0)$ and $p_2(t) = (0, 0, 1)$ for $0 < t < \pi$. Then, we have

$$\lim_{t \to 0}(p_1(t) + p_2(t)) = (0, 1, 1) \notin \mathrm{Join}(\mathcal{M}_1, \mathcal{M}_2).$$

---

[4]We would like to thank an anonymous referee for pointing out the first example to us.

The tangent spaces to $\mathcal{M}_1$ and $\mathcal{M}_2$ are orthogonal to each other. Using this information, we can determine the condition number easily via Theorem 1.1: the matrix U in the statement of the theorem is a matrix with orthonormal columns, so that $\varsigma_2(U) = 1$. Hence, $\kappa((p_1(t), p_2(t))) = 1^{-1}$. Consequently, $\lim_{t \to 0} \kappa(p_1(t), p_2(t)) = 1$.

The second example shows a join set that has an open boundary point where the condition number neither converges nor diverges towards ∞.

Example 4.3. Consider the embedded smooth submanifolds of $\mathbb{R}^3$,

$$\mathcal{M}_1 = \{(-t, 0, 0) \mid 1 \leq t < \infty\} \quad \text{and} \quad \mathcal{M}_2 = \left\{\left(t, \tfrac{\cos(t)}{t}, \tfrac{\sin(t^2)}{t}\right) \mid 1 \leq t < \infty\right\}.$$

It is easy to see that they are not cones.

Let $p_1(t) = (-t, 0, 0)$ and $p_2(t) = \left(t, \tfrac{\cos(t)}{t}, \tfrac{\sin(t^2)}{t}\right)$ for $1 \leq t \leq \infty$. Then,

$$\lim_{t \to \infty} p_1(t) + p_2(t) = (0, 0, 0),$$

which is not a point in $\mathrm{Join}(\mathcal{M}_1, \mathcal{M}_2)$. The tangent spaces to $\mathcal{M}_1$ and $\mathcal{M}_2$ are given respectively by the spans of

$$\mathbf{v}(t) = (1, 0, 0) \text{ and } \mathbf{w}(t) = \left(1, \tfrac{-\sin(t)t - \cos(t)}{t^2}, 2\cos(t^2) - \tfrac{\sin(t^2)}{t^2}\right).$$

By Theorem 1.1, $\kappa((p_1(t), p_2(t)))^{-1} = \varsigma_2\left(\begin{bmatrix} \mathbf{v}(t) & \mathbf{w}(t)/\|\mathbf{w}(t)\| \end{bmatrix}\right) = \varsigma_2(U(t))$. Let $z(t)$ be such that $\|\mathbf{w}(t)\|^2 = 1 + z(t)$. It is straightforward to show that

$$\varsigma_2(U(t)) = \sqrt{1 - \frac{1}{\sqrt{1 + z(t)}}} = \sqrt{2} \sin\left(\tfrac{\theta(t)}{2}\right),$$

where $\theta(t) = \arccos\left((1 + z(t))^{-\frac{1}{2}}\right)$ happens to be the angle between $\mathrm{T}_{p_1(t)}\mathcal{M}_1$ and $\mathrm{T}_{p_2(t)}\mathcal{M}_2$. For large t we have $\theta(t) \sim \arccos\left((1 + (2\cos(t^2))^2)^{-\frac{1}{2}}\right)$, from which we see that $\kappa((p_1(t), p_2(t)))$ does not converge.

Note that in this example there exist both convergent subsequences as well as divergent subsequences. Indeed, letting $t_k = \sqrt{2k\pi}$, one finds

$$\lim_{k \to \infty} \kappa((p_1(t_k), p_2(t_k))) = \frac{1}{\sqrt{2}\sin(\arccos(\tfrac{1}{\sqrt{5}}))} = \sqrt{1 - \sqrt{\tfrac{1}{5}}},$$

whereas letting $t_k = \sqrt{k\pi/2}$ yields

$$\lim_{k \to \infty} \kappa((p_1(t_k), p_2(t_k))) \to \frac{1}{\sqrt{2}\sin(\arccos(1))} = \infty.$$

This situation is troublesome because it is a theoretical possibility that an iterative algorithm for solving the associated JAP with $\mathbf{p} = \mathbf{0} \in \overline{\mathcal{J}} \setminus \mathcal{J}$ as input yields only iterates for which the condition number is bounded by some small constant, so that the condition number cannot detect that the final iterate has no meaningful interpretation.

The examples illustrate that additional conditions on the manifolds $\mathcal{M}_i$ need to be imposed to continuously extend the condition number to open boundary points. However, we do not know which are the minimal restrictions for reaching this goal.

We proceed with the proof of Theorem 1.4. The key property that we exploit is the fact that the condition number of the JDP at $(p_1, \ldots, p_r)$ is invariant under scaling of the individual factors if the $\mathcal{M}_i$'s are cones. The reason for this is that the tangent space to a point of a cone is invariant under scaling of that point.

11

PROPOSITION 4.4. *Let $\mathcal{M}_i$ be an n_i-dimensional smooth submanifold embedded in $\mathbb{R}^N$. Assume that $\mathcal{M}_i$ is a cone. Let $p_i \in \mathcal{M}_i$, then*

$$\kappa((p_1, p_2, \ldots, p_r)) = \kappa((\beta_1 p_1, \beta_2 p_2, \ldots, \beta_r p_r)) \text{ for all } \beta_i \in \mathbb{R}_0, i = 1, 2, \ldots, r.$$

We can now present the proof of Theorem 1.4.

Proof of Theorem 1.4. For brevity, let $\mathfrak{p}(t) := (\mathbf{p}_1(t), \ldots, \mathbf{p}_r(t))$. Assume that the curve $\mathfrak{p}(t)$ admits a bounded subsequence, then this subsequence has a limit $\mathfrak{p}_\star \in \overline{\mathcal{M}_1} \times \ldots \times \overline{\mathcal{M}_r}$. As Φ is a continuous map and $\mathcal{J} = \mathrm{Join}(\overline{\mathcal{M}_1}, \ldots, \overline{\mathcal{M}_r})$, it follows that $\mathbf{p}_\star = \Phi(\mathfrak{p}_\star) \in \mathcal{J}$, which contradicts the assumption in the formulation of the theorem.

The foregoing shows that $\mathfrak{p}(t)$ becomes unbounded, i.e., $\lim_{t\to 0} \|\mathfrak{p}(t)\| \to \infty$, in the product metric on $\mathcal{M}_1 \times \ldots \times \mathcal{M}_r$, i.e., $\|\mathfrak{p}(t)\| = \sqrt{\sum_{i=1}^r \|\mathbf{p}_i(t)\|^2}$. Hence, there is at least one $i_\star \in [1, d]$ such that $\lim_{t\to 0} \|\mathbf{p}_{i_\star}(t)\| \to \infty$. The sequence $\mathbf{p}(t) := \Phi(\mathfrak{p}(t))$ is assumed to converge in $\mathbb{R}^N$, so it is bounded for small t. Since $\mathbf{p}(t) = \mathbf{p}_1(t) + \cdots + \mathbf{p}_r(t)$, there exists another $j_\star \in [1, d]$ with $j_\star \neq i_\star$ so that $\lim_{t\to 0} \|\mathbf{p}_{j_\star}(t)\| \to \infty$.

Let $U(t) = [U_1(t), \ldots, U_r(t)]$, where $U_i(t)$ is a matrix with orthonormal columns that form a basis for $\mathrm{T}_{\mathbf{p}_i(t)}\mathcal{M}_i$. Then Theorem 1.1 states that $\kappa(\mathfrak{p}(t)) = \frac{1}{\varsigma_n(U(t))}$. For proving that $\lim_{t\to 0} \kappa(\mathfrak{p}(t)) \to \infty$ we will construct an explicit $\mathbf{x}(t) \in \mathrm{T}_{\mathbf{p}_1(t)}\mathcal{M}_1 \times \cdots \times \mathrm{T}_{\mathbf{p}_r(t)}\mathcal{M}_r$ and show $\|U(t)\mathbf{x}(t)\| \to 0$, while $\|\mathbf{x}(t)\| \geq 1$, implying $\varsigma_n(U(t)) \to 0$.

Let $\mathbf{q}(t), \widetilde{\mathbf{q}}(t) \subset \mathbb{S}(\mathbb{R}^N)$ be the two curves with

$$(4.1) \qquad \alpha(t)\mathbf{q}(t) = \mathbf{p}_{i_\star}(t), \ \alpha(t) > 0, \quad \text{and} \quad \beta(t)\widetilde{\mathbf{q}}(t) = \sum_{j \neq i_\star} \mathbf{p}_j(t), \ \beta(t) > 0.$$

Since $\alpha(t)\mathbf{q}(t) + \beta(t)\widetilde{\mathbf{q}}(t)$ converges for $t \to 0$, there is a constant $c > 0$ with

$$c \geq \|\alpha(t)\mathbf{q}(t) + \beta(t)\widetilde{\mathbf{q}}(t)\|^2 = \alpha(t)^2 + 2\alpha(t)\beta(t)\langle \mathbf{q}(t), \widetilde{\mathbf{q}}(t)\rangle + \beta(t)^2.$$

Since $\alpha(t)$ is unbounded and bounded away from 0 for $t \approx 0$, $\beta(t)$ is unbounded as well. Write $\beta(t) = \gamma(t)\alpha(t)$ with $\gamma(t) > 0$. Then, we have

$$(4.2) \qquad c\alpha(t)^{-2} \geq 1 + 2\gamma(t)\langle \mathbf{q}(t), \widetilde{\mathbf{q}}(t)\rangle + \gamma(t)^2.$$

Real solutions $\gamma(t)$ exist only if $\langle \mathbf{q}(t), \widetilde{\mathbf{q}}(t)\rangle^2 \geq 1 - c\alpha(t)^{-2}$. In fact, by the Cauchy-Schwartz inequality $\langle \mathbf{q}(t), \widetilde{\mathbf{q}}(t)\rangle^2 \leq 1$, the only solution consistent with (4.2) satisfies

$$\lim_{t\to 0} \langle \mathbf{q}(t), \widetilde{\mathbf{q}}(t)\rangle = -1.$$

Similar to (4.1), for all $j \neq i_\star$, write $\mathbf{p}_j(t) = \nu_j(t)\mathbf{q}_j(t)$ so that $\mathbf{q}_j(t) \in \mathbb{S}(\mathbb{R}^N)$ and $\nu_j(t) \geq 0$. Since the $\mathcal{M}_i$'s are all cones we have $\mathbf{q}_i(t) \in \mathcal{M}_i$ and, because of homogeneity, $\mathrm{T}_{\mathbf{p}_i(t)}\mathcal{M}_i = \mathrm{T}_{\mathbf{q}_i(t)}\mathcal{M}_i$ and $\mathbf{q}_i(t) \in \mathrm{T}_{\mathbf{q}_i(t)}\mathcal{M}_i$. For all $1 \leq i \leq r$ we write $\mathbf{q}_i(t) = U_i(t)\mathbf{x}_i(t)$. Consequently, $\|\mathbf{x}_i(t)\| = 1$. Then

$$\widetilde{\mathbf{q}}(t) = \beta(t)^{-1} \sum_{j \neq i_\star} \nu_j(t)\mathbf{q}_j(t) = \beta(t)^{-1} \sum_{j \neq i_\star} \nu_j(t)U_j(t)\mathbf{x}_j(t).$$

Let $(\mathbf{z}_1, \ldots, \mathbf{z}_k) := [\mathbf{z}_1^T \ \ldots \ \mathbf{z}_k^T]^T$ denote the vertical stacking of the vectors. Then,

$$\beta(t)\mathbf{x}(t) =$$
$$\left(\nu_1(t)\mathbf{x}_1(t), \ldots, \nu_{i_\star-1}(t)\mathbf{x}_{i_\star-1}(t), \beta(t)\mathbf{x}_{i_\star}(t), \nu_{i_\star+1}(t)\mathbf{x}_{i_\star+1}(t), \ldots, \nu_r(t)\mathbf{x}_r(t)\right).$$

For all $t \approx 0$ we have $\|\mathbf{x}(t)\| \geq \|\mathbf{x}_{i_\star}(t)\| = 1$. Since $\langle \mathbf{q}(t), \widetilde{\mathbf{q}}(t)\rangle$ tends to -1 as $t \to 0$, it follows by construction that $\lim_{t\to 0} U(t)\mathbf{x}(t) = 0$. $\qed$

12

Let $\mathcal{X}$ be a smooth projective variety. Then, combining the foregoing theorem with [37, Theorem 2.1] shows that all points on the algebraic boundary of the real rank-2 boundary $\rho(\mathcal{X})$ inside of $\mathrm{Join}(\mathcal{X}, \mathcal{X})$ are always ill-posed for the JDP.

We conclude this section by showing that the several join sets arising in tensor decomposition problems satisfy the assumptions of Theorem 1.4.

PROPOSITION 4.5. *Let $\mathcal{X} \subset \mathbb{PR}^N$ be a smooth projective variety, and let $\widehat{\mathcal{X}} \subset \mathbb{R}^N$ be the affine cone over $\mathcal{X}$, excluding zero. Then, $\sigma_r(\widehat{\mathcal{X}}) = \mathrm{Join}(\widehat{\mathcal{X}}, \ldots, \widehat{\mathcal{X}})$ satisfies the assumptions in Theorem 1.4.*

Proof. The fact that $\widehat{\mathcal{X}}$ is a smooth, analytic manifold is by definition, because the vertex at zero, which could be singular, is removed from our definition of a cone. The closure of $\widehat{\mathcal{X}}$ in the Euclidean topology is $\overline{\mathcal{X}} := \widehat{\mathcal{X}} \cup \{0\}$. To show that $\sigma_r(\overline{\mathcal{X}}) \subset \sigma_r(\widehat{\mathcal{X}})$ (the converse inclusion is trivial), let $\mathbf{p} = \mathbf{p}_1 + \cdots + \mathbf{p}_r$ with $\mathbf{p}_i \in \overline{\mathcal{X}}$. Assume without loss of generality that the first k, $1 \le k \le r$, $\mathbf{p}_i$'s are zero. If $k = 1$, take $\mathbf{q} = \mathbf{p}_2$ and then $\mathbf{p} = 2\mathbf{p}_2 - \mathbf{p}_2 + \mathbf{p}_3 + \cdots + \mathbf{p}_r \in \sigma_r(\widehat{\mathcal{X}})$ as well. If $k > 1$, take $\mathbf{q} \in \widehat{\mathcal{X}}$ arbitrary and note that $\sum_{i=1}^{k} \mathbf{p}_i = 0 = \sum_{i=1}^{k} \alpha_i \mathbf{q}$ with $\sum_{i=1}^{k} \alpha_i = 0$ and all $\alpha_i \ne 0$. Hence, $\mathbf{p} = \mathbf{p}_1 + \cdots + \mathbf{p}_r = \alpha_1 \mathbf{q} + \cdots + \alpha_k \mathbf{q} + \mathbf{p}_{k+1} + \cdots + \mathbf{p}_r \in \sigma_r(\widehat{\mathcal{X}})$ as well. $\square$

In particular, this result covers the next r-secant sets: tensors of canonical rank bounded by r; symmetric tensors of Waring rank bounded by r; and partially symmetric tensors of partially symmetric rank bounded by r. However, the proposition does not apply to block term decompositions, because the set of tensors of multilinear rank at most $(r_1, \ldots, r_d)$ is not a smooth projective variety [32, Section 3]. We do not know if a result analogous to Theorem 1.4 holds for block term decompositions.

5. Examples. This section presents explicit expressions for the matrices U_i in Theorem 1.4 for two well-studied tensor-related JDPs, so that the condition number can be computed efficiently by computing the least singular value of $[\, U_1 \ U_2 \ \cdots \ U_r \,]$. The first example is the *tensor rank decomposition problem* that consists of computing the rank-1 tensors comprising the CP decomposition[5] (CPD) of an $m_1 \times \cdots \times m_d$ tensor. The second example is the problem of computing the symmetric rank-1 tensors appearing in a Waring decomposition of a symmetric $m \times \cdots \times m$ tensor.

5.1. The CP decomposition. Recall that the Segre manifold $\mathcal{S} \subset \mathbb{R}^N$, where $N = m_1 \cdots m_d$, is the analytic manifold of rank-1 tensors.[6] Let $\otimes$ denote the tensor product. Then, the r-secant set of $\mathcal{S}$ is

$$\sigma_r(\mathcal{S}) = \mathrm{Join}(\mathcal{S}, \ldots, \mathcal{S}) = \Big\{ \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \mid \mathbf{a}_i^k \in \mathbb{R}_0^{m_k} \Big\},$$

which is also a semi-algebraic set [20]. This class appears in a wide variety of applications, such as psychometrics [30], chemical sciences [39], theoretical computer science [13], signal processing [38], and machine learning [38], among others. In several of these applications, the uniquely determined rank-1 terms appearing in this CPD admit an interpretation in the application domain; see [38, Section IX]. Hence, the condition number is of natural interest [40].

In the remainder of this section, let $p_i = \mu_i \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathcal{S}$ with $\mathbf{a}_i^k \in \mathbb{S}(\mathbb{R}^{m_k})$ and $\mu_i > 0$ be rank-1 tensors. Then, an orthogonal basis for $\mathrm{T}_{p_i}\mathcal{S} \subset \mathbb{R}^N$ is obtained by considering the derivative of the surjective map

$$\Psi : \mathbb{R} \times \mathbb{S}(\mathbb{R}^{m_1}) \times \cdots \times \mathbb{S}(\mathbb{R}^{m_d}) \to \mathcal{S}, \quad (\mu_i, \mathbf{a}_i^1, \ldots, \mathbf{a}_i^d) \mapsto \mu_i \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d,$$

---

[5]This decomposition is also called CANDECOMP/PARAFAC, canonical polyadic decomposition, or tensor rank decomposition in the literature.

[6]For consistency with the rest of the article, we exploit the natural isomorphism between $\mathbb{R}^{m_1 \times \cdots \times m_d}$ and $\mathbb{R}^N$, given by *vectorizing* the tensor.

where $\otimes$ should be interpreted as the Kronecker product. Hence, the basis vectors are given by the columns of

$$(5.1) \qquad U_i := \begin{bmatrix} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d & Q_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d & \cdots & \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^{d-1} \otimes Q_i^d \end{bmatrix},$$

where $Q_i^k \in \mathbb{R}^{m_k \times (m_k-1)}$ is a matrix whose columns form an orthonormal basis of $\mathrm{T}_{\mathbf{a}_i^k} \mathbb{S}(\mathbb{R}^{m_k})$, i.e., the orthogonal complement of $\mathbf{a}_i^k$. From Theorem 1.1 it follows that the condition number of the CPD problem at $\mathfrak{p} = (p_1, \ldots, p_r)$ is $\kappa(\mathfrak{p}) = (\varsigma_n([U_1 \cdots U_r]))^{-1}$, where $n = \dim \mathcal{S}^{\times r} = r \cdot \dim \mathcal{S} = r(1 - d + \sum_{k=1}^d m_k)$.

Using substantially the same arguments as in [40, Section 5], the following key properties of the condition number are straightforward to prove.

PROPOSITION 5.1. *The condition number is scale and orthogonally invariant: for every $\mathfrak{A} = \Phi(\mathfrak{p}) \in \sigma_r(\mathcal{S})$, all $\alpha \in \mathbb{R}_0$, and all $Q = Q_1 \otimes \cdots \otimes Q_d$ with Q_k an $m_k \times m_k$ orthogonal matrix, it holds that $\kappa((p_1, \ldots, p_r)) = \kappa((\alpha Q p_1, \ldots, \alpha Q p_r))$.*

Proof. Invariance with respect to α was shown in Proposition 4.4. Orthogonal invariance follows by noting that $Q_k Q_i^k$ is an orthogonal basis of the orthogonal complement of $Q_k \mathbf{a}_i^k$ so that an orthogonal basis of $\mathrm{T}_{Q p_i} \mathcal{S}$ is $Q U_i$ with U_i as in (5.1). The result follows by the orthogonal invariance of singular values. $\square$

Recall from [40, Section 5.2] that a CPD $\mathfrak{A} = \sum_{i=1}^r \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$ is called weak 3-orthogonal if for every $1 \le i < j \le d$ there exist $1 \le k_1 < k_2 < k_3 \le d$, depending on i and j, such that there is orthogonality in the corresponding factors: $\mathbf{a}_i^{k_1} \perp \mathbf{a}_j^{k_1}$, $\mathbf{a}_i^{k_2} \perp \mathbf{a}_j^{k_2}$, and $\mathbf{a}_i^{k_3} \perp \mathbf{a}_j^{k_3}$. This class includes orthogonally decomposable tensors [29, 44]. It turns out that weak 3-orthogonal tensors are very well conditioned.

PROPOSITION 5.2. *The condition number of a weak 3-orthogonal CPD is 1.*

Proof. One verifies that $[U_1 \cdots U_r]$ is a matrix with orthonormal columns due to the weak 3-orthogonality. The result follows from the fact that a matrix with orthonormal columns has all singular values equal to 1. $\square$

COROLLARY 5.3. *The condition number of a rank-1 tensor is 1.*

5.1.1. Relation to the norm-balanced condition number. One can ask how the proposed condition number relates to the one that was recently proposed in [40]. The *norm-balanced* condition number of a decomposition $\mathfrak{p} = (p_1, \ldots, p_r)$, where the p_i's are as above, is defined as $\widetilde{\kappa}(\mathfrak{p}) := (\varsigma_n([\widetilde{U}_1 \widetilde{U}_2 \cdots \widetilde{U}_r]))^{-1}$, where

$$\widetilde{U}_i := \mu_i^{1-1/d} \begin{bmatrix} I_{m_1} \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d & \cdots & \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^{d-1} \otimes I_{m_d} \end{bmatrix}.$$

The condition number $\widetilde{\kappa}$ measures the sensitivity of $\{\mu_i^{1/d} \mathbf{a}_i^k \in \mathbb{R}^{m_k}\}_{i,k}$ with respect to perturbations of the tensor $\mathfrak{A} = \Phi(\mathfrak{p}) \in \sigma_r(\mathcal{S})$, whereas the proposed condition number κ measures the sensitivity of the rank-1 tensors $(p_1, \ldots, p_r)$ in $\mathcal{S}^{\times r}$. The appropriate choice of condition number depends on the specific application.

The two condition numbers κ and $\widetilde{\kappa}$ are related as follows. Let

$$f : (\mathbb{R}_0^{m_1} \times \cdots \times \mathbb{R}_0^{m_d})^{\times r} \to \mathbb{R}^N, \quad \mathfrak{a} = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d)_{i=1,\ldots,r} \mapsto \sum_{i=1}^r \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d.$$

From [40, Theorem 1] we know that for a *norm-balanced* $\mathfrak{a}$, i.e., $\|\mathbf{a}_i^1\| = \cdots = \|\mathbf{a}_i^d\|$, the condition number is $\widetilde{\kappa}(\mathfrak{p}) = (\varsigma_n(\mathrm{d}_\mathfrak{a} f))^{-1}$, where $\mathfrak{p} = (p_1, \ldots, p_r)$ with $p_i = \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$, and $n = r \cdot \dim \mathcal{S}$ as before. We have the following diagram:

$$(5.2)$$

$$\begin{array}{ccc} (\mathbb{R}_0^{m_1} \times \cdots \times \mathbb{R}_0^{m_d})^{\times r} & & \\ \Big\downarrow{\scriptstyle \sigma^{\times r}} & \searrow^{f} & \\ (\mathbb{R}^{m_1} \otimes \cdots \otimes \mathbb{R}^{m_d})^{\times r} \supset \mathcal{S}^{\times r} & \xrightarrow[\Phi]{} & \sigma_r(\mathcal{S}) \end{array}$$

Here $\sigma^{\times r}$ denotes the r-fold product of the Segre map $\sigma : (\mathbf{a}^1, \dots, \mathbf{a}^d) \mapsto \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d$. Since $f = \Phi \circ \sigma^{\times r}$, we have $\mathrm{d}_{\mathfrak{a}} f = \mathrm{d}_{\mathfrak{p}} \Phi \circ \mathrm{d}_{\mathfrak{a}} \sigma^{\times r}$. Note that $\mathrm{d}_{\mathfrak{a}} \sigma^{\times r} \in \mathbb{R}^{n \times r(m_1 + \cdots + m_d)}$ is of full row rank because $\mathcal{S}$ is a smooth manifold. If $\mathrm{d}_{\mathfrak{p}} \Phi \in \mathbb{R}^{N \times n}$ is of maximal column rank then $(\mathrm{d}_{\mathfrak{a}} f)^\dagger = (\mathrm{d}_{\mathfrak{a}} \sigma^{\times r})^\dagger (\mathrm{d}_{\mathfrak{p}} \Phi)^\dagger$, by [25, Section 5.7, Fact 16]. Hence,

$$(5.3) \qquad \widetilde{\kappa}(\mathfrak{p}) := \|(\mathrm{d}_{\mathfrak{a}} f)^\dagger\|_2 \le \|(\mathrm{d}_{\mathfrak{a}} \sigma^{\times r})^\dagger\|_2 \|(\mathrm{d}_{\mathfrak{p}} \Phi)^\dagger\|_2 = \left(\varsigma_n(\mathrm{d}_{\mathfrak{a}} \sigma^{\times r}) \right)^{-1} \cdot \kappa(\mathfrak{p}).$$

Observe that if $\mathrm{d}_{\mathfrak{p}} \Phi$ is not injective then $\widetilde{\kappa}(\mathfrak{p}) = \kappa(\mathfrak{p}) = \infty$.

PROPOSITION 5.4. *Let* $\mathfrak{a} = (\mathbf{a}_i^1, \dots, \mathbf{a}_i^d)_{i=1}^r$ *be norm-balanced. Then, inequality (5.3) is sharp for weak* 3*-orthogonal CPDs* $\mathfrak{p} = \sigma^{\times r}(\mathfrak{a})$.

Proof. By Proposition 5.2, $\kappa(\mathfrak{p}) = 1$, whereas [40, Proposition 19] states that $\widetilde{\kappa}(\mathfrak{p})$ is $(\mu_{\min})^{1/d-1}$, where $\mu_{\min}$ is the smallest value in $\{\mu_i := \|p_i\|\}_{i=1}^r$. One verifies that the matrix of $\mathrm{d}_{\mathfrak{a}} \sigma^{\times r}$ with respect to the correct bases is $\mathrm{diag}(\widetilde{U}_1, \dots, \widetilde{U}_r)$. Consequently, $\varsigma_n(\mathrm{d}_{\mathbf{a}} \sigma^r) = (\mu_{\min})^{1-1/d}$. $\square$

There is an interesting difference between the two condition numbers when measuring distances to $\sigma_{r-1}(\mathcal{S})$. Assume that the the norms of the rank-1 tensors p_i are sorted: $\mu_1 \ge \cdots \ge \mu_r > 0$. If μ_r is small then a small perturbation applied to $\mu_r p_r$ sends it to 0, hereby sending the perturbed tensor to σ_{r-1}. The condition number at the perturbed point is $\widetilde{\kappa} = \varsigma_n([\tilde{U}_1 \cdots \tilde{U}_{r-1} \; 0])^{-1} = \infty$. It follows from the continuity of singular values that as $\mu_r \to 0$, the condition number $\widetilde{\kappa} \to \infty$ so that for small μ_r the absolute norm-balanced condition number of [40] is automatically large. By contrast, κ is insensitive to the norm of the points p_i, as was shown in Proposition 4.4, hence $\mu_r \approx 0$ does not automatically imply a large condition number. It also entails that if a perturbation of magnitude less than μ_r is applied to $p = p_1 + \cdots + p_r$ in the direction of p_r, then the condition number remains constant. This type of perturbation does not decrease the projection distance from $(\mathrm{T}_{p_1} \mathcal{S}, \dots, \mathrm{T}_{p_r} \mathcal{S})$ to Σ_{Gr}, which is exactly the inverse of $\kappa(\mathfrak{p})$ according to Theorem 1.4. The only way to perturb p towards $\sigma_{r-1}(\mathcal{S})$ so that the condition number increases is to change the angles between the p_i's, which causes the angles between $\mathrm{T}_{p_i} \mathcal{S}$ to shift, hereby changing the projection distance to Σ_{Gr}.

5.2. The Waring decomposition. A symmetric tensor $\mathfrak{A} \in \mathbb{R}^{m \times \cdots \times m}$ is a tensor whose entries are invariant under a permutation of indices: $a_{i_1, \dots, i_d} = a_{\pi_1, \dots, \pi_d}$ where π is any permutation of $\{i_1, \dots, i_d\}$. Such tensors can be decomposed as

$$\mathfrak{A} = \sum_{i=1}^r \mathbf{a}_i^{\otimes d} = \sum_{i=1}^r \mathbf{a}_i \otimes \cdots \otimes \mathbf{a}_i \quad \text{with } \mathbf{a}_i \in \mathbb{R}^m;$$

if r is minimal, this expression is called a Waring decomposition. The summands in generic Waring decompositions are uniquely determined [18]. This type of decompositions occurs often in latent variable models [6], where the individual rank-1 symmetric tensors $\mathbf{a}_i$ correspond to the quantities of interest.

The set of rank-1 symmetric tensors is the affine cone over a smooth projective variety called the Veronese variety [33]; hence, it is a manifold. It can be constructed as the image of the surjective map

$$\Psi : \mathbb{R}_0 \times \mathbb{S}(\mathbb{R}^m) \to \mathcal{V}, \quad (\mu_i, \mathbf{a}_i) \mapsto \mu \mathbf{a}_i^{\otimes d} = \mu_i \mathbf{a}_i \otimes \cdots \otimes \mathbf{a}_i.$$

The join set $\mathcal{J} = \mathrm{Join}(\mathcal{V}, \dots, \mathcal{V})$ is then the set of all symmetric tensors that have a Waring decomposition of rank bounded by r.

The condition number of the Waring decomposition problem (WDP) is determined analogously to the previous example. The derivative of Ψ is

$$\mathrm{d}_{(\mu_i, \mathbf{a}_i)} \Psi : \mathrm{T}_{\mu_i} \mathbb{R}_0 \times \mathrm{T}_{\mathbf{a}_i} \mathbb{S}(\mathbb{R}^m) \to \mathrm{T}_{\Psi(\mu_i, \mathbf{a}_i)} \mathcal{V}$$

$$(\dot{\mu}_i, \dot{\mathbf{a}}_i) \mapsto \dot{\mu}_i \mathbf{a}_i^{\otimes d} + \mu_i \dot{\mathbf{a}}_i \otimes \mathbf{a}_i^{\otimes d-1} + \cdots + \mathbf{a}_i^{\otimes d-1} \otimes \dot{\mathbf{a}}_i.$$

Let Q_i be an $m \times (m-1)$ matrix whose columns form an orthonormal basis of $\mathrm{T}_{\mathbf{a}_i} \mathbb{S}(\mathbb{R}^m)$. Let $p_i = \mu_i \mathbf{a}_i^{\otimes d}$. Then, we see that an orthonormal basis of $\mathrm{T}_{p_i} \mathcal{V}$ is given by the columns of the $m^d \times m$ matrix[7]

$$V_i := \begin{bmatrix} \mathbf{a}_i^{\otimes d} & \frac{1}{d}(Q_i \otimes \mathbf{a}_i^{\otimes d-1} + \cdots + \mathbf{a}_i^{\otimes d-1} \otimes Q_i) \end{bmatrix};$$

the fact that $V_i^T V_i = I_m$ can be verified readily. Then it follows from Theorem 1.1 that the condition number of the WDP at $\mathfrak{p} = (p_1, \ldots, p_r)$ is $\kappa(\mathfrak{p}) = (\varsigma_n([V_1 \cdots V_r]))^{-1}$, where $n = \dim \mathcal{V}^{\times r} = rm$.

The following results are obtained by arguments very similar to those employed in the corresponding results in subsection 5.1.

PROPOSITION 5.5. *The condition number is scale and orthogonally invariant: for every $\mathfrak{A} = \Phi(\mathfrak{p}) \in \sigma_r(\mathcal{V})$, all $\alpha \in \mathbb{R}_0$, and all $Q = Q' \otimes \cdots \otimes Q'$ with Q' an $m \times m$ orthogonal matrix, it holds that $\kappa((p_1, \ldots, p_r)) = \kappa((\alpha Q p_1, \ldots, \alpha Q p_r))$.*

Recall that a symmetric *odeco* tensor $\mathfrak{A}$ can be expressed as $\sum_{i=1}^r \mathbf{a}_i^{\otimes d}$ with $\mathbf{a}_i \perp \mathbf{a}_j$ for all $i \neq j$ and $r \leq n$. Since Kruskal's lemma [28] applies, the set of symmetric rank-1 summands in the decomposition $\mathfrak{p} \in \mathcal{V} \times \cdots \times \mathcal{V}$ is uniquely determined. We can thus call $\kappa(\mathfrak{p})$ *the* condition number of a symmetric odeco tensor $\Phi(\mathfrak{p})$.

PROPOSITION 5.6. *The condition number of a symmetric odeco tensor is 1.*

COROLLARY 5.7. *The condition number of a symmetric rank-1 tensor is 1.*

6. Numerical experiments. We illustrate the theoretical results in the setting of CPDs. We performed several experiments in Matlab R2016b, employing some features of Tensorlab v3.0 [41]. The code implementing the computation of the condition number presented in subsection 5.1 is included in the ancillary files of the arXiv version of this paper.

In this section, low-rank tensors are specified by *factor matrices* $A_k = [\mathbf{a}_i^k]_{i=1}^r$ with $\mathbf{a}_i^k \in \mathbb{R}^{m_k}$; the associated tensor is $\mathfrak{A} := [\![A_1, \ldots, A_d]\!] := \sum_{i=1}^r \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$. By $X \sim N(0,1)$ we mean that all elements of $X \in \mathbb{R}^{m \times n}$ are standard normally distributed with zero mean and unit standard deviation.

6.1. Estimating typical forward errors. While the bound in (1.4), namely $\|\mathfrak{p} - \Phi_{\mathfrak{p}}^{-1}(\mathbf{w})\| \lesssim \kappa(\mathfrak{p}) \cdot \|\Phi(\mathfrak{p}) - \mathbf{w}\|$, is asymptotically sharp, it may be questioned insofar this is a *typical* result. Perhaps the estimate obtained from the condition number is overly pessimistic compared to the errors that one encounters in practice?

For investigating the above question, we conducted the following experiment. A model for generating potentially ill-conditioned CPD problems is the tensor in $\mathbb{R}^{m_1 \times \cdots \times m_d}$ whose kth factor matrix is

(6.1) $\qquad A_k(s) = C_k(2^{-as} \cdot I_r + X_k Y_k^T) \quad \text{with } C_k, X_k, Y_k \sim N(0,1),$

and where $a > 0$, $C_k \in \mathbb{R}^{m_k \times r}$ and $X_k, Y_k \in \mathbb{R}^{r \times r_k}$ with $r_k < m_k$. These tensors are of rank r and maximal multilinear rank with probability 1, and as $s \to \infty$ the factor matrix $A_k(s)$ is tending to the rank-r_k matrix $C_k X_k Y_k^T$. The tensor $\mathfrak{A}(s) = [\![A_1(s), \ldots, A_d(s)]\!]$ will tend to a tensor whose multilinear rank is at most $(r_1, \ldots, r_d)$ as $s \to \infty$. Recall that a tensor of multilinear rank $(r_1, \ldots, r_d)$ in $\mathbb{R}^{m_1 \times \cdots \times m_d}$ is r-identifiable if and only if the $r_1 \times \cdots \times r_d$ core array in its higher-order singular value decomposition (HOSVD) [19] is r-identifiable [16, Theorem 4.1]. In particular if r is strictly larger than the generic complex rank of $\mathbb{R}^{r_1 \times \cdots \times r_d}$ [8], then $\mathfrak{A}(s)$ tends to a

---

[7]In fact, $\mathcal{V} \subset S^d \mathbb{R}^m \simeq \mathbb{R}^{\binom{d+m-1}{d}}$, so it is possible to reduce the size of this matrix to $\binom{d+m-1}{d} \times m$. For simplicity we ignore this optimization.

FIGURE 6.1. *Distribution of the scaling factor* $\frac{\|B_1(s)\odot\cdots\odot B_d(s)-\widetilde{B}_1(s)\odot\cdots\odot\widetilde{B}_d(s)\|_F}{\kappa(s)\cdot\|\mathfrak{A}(s)-\widetilde{\mathfrak{A}}(s)\|_F} \lesssim 1$ *for 250 random samples of model* (6.1).
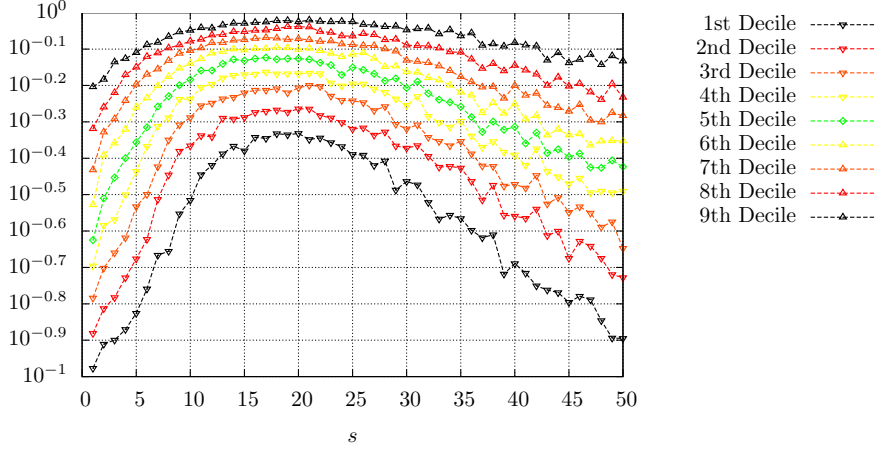

TABLE 6.1
The quartiles of the observed condition numbers in the experiment described in subsection 6.1.

s	1	10	20	30	40	50
1st quartile	$1.27\cdot10^1$	$5.40\cdot10^1$	$2.60\cdot10^2$	$1.22\cdot10^3$	$5.03\cdot10^3$	$2.00\cdot10^4$
median	$2.04\cdot10^1$	$1.02\cdot10^2$	$4.88\cdot10^2$	$2.09\cdot10^3$	$8.55\cdot10^3$	$3.44\cdot10^4$
3rd quartile	$3.27\cdot10^1$	$1.77\cdot10^2$	$8.68\cdot10^2$	$3.65\cdot10^3$	$1.50\cdot10^4$	$6.03\cdot10^4$

tensor with infinitely many decompositions with probability 1. We expect that the condition number of $\mathfrak{A}(s)$ increases with s because of Corollary 1.2.

We take $(m_1, m_2, m_3, m_4) = (6, 5, 4, 4)$, $(r_1, r_2, r_3, r_4) = (1, 2, 3, 4)$, $r = 6$, $a = \frac{1}{5}$, and $s = 1, 2, \ldots, 50$ as an instance of the above model. Then, the tensor

$$\mathfrak{A}(s) = [\![\|A_1(s)\|_F^{-1}A_1(s), \ldots, \|A_d(s)\|_F^{-1}A_d(s)]\!] = [\![B_1(s), \ldots, B_d(s)]\!]$$

was computed. For every s, we tried to recover the CPD of $\mathfrak{A}(s)$ using the `cpd_nls` optimization method from Tensorlab. It was halted either if the value of the objective function $\frac{1}{2}\|\Phi(\mathfrak{p}(s)) - \mathfrak{A}(s)\|^2 \leq 10^{-14}$ or after 1000 iterations. As initial factor matrices from whence `cpd_nls` started we chose $\{B_1(s) + \tau R_1, \ldots, B_d(s) + \tau R_d\}$, where $R_k \sim N(0, 1)$ and $\tau = 5\cdot10^{-4}$. Let $\widetilde{\mathfrak{A}}(s) = [\![\widetilde{B}_1(s), \ldots, \widetilde{B}_d(s)]\!]$ denote the CPD computed by `cpd_nls`. Then, the following data was recorded: the backward error $\|\mathfrak{A}(s) - \widetilde{\mathfrak{A}}(s)\|_F$, the forward error $\|B_1(s) \odot \cdots \odot B_d(s) - \widetilde{B}_1(s) \odot \cdots \odot \widetilde{B}_d(s)\|_F$, and the condition number $\kappa(s)$ of the problem at the computed CPD $\widetilde{B}(s)$. For every $s = 1, 2, \ldots, 50$, we took 250 independent samples from model (6.1) with the aforementioned choice of parameters. Out of the 12500 experiments, 16 were discarded because the `cpd_nls` method did not converge within 1000 iterations.

The obtained results are summarized in Figure 6.1 and Table 6.1. The quartiles of the observed condition numbers are shown in Table 6.1; note that they indeed increase with s. Figure 6.1 plots the deciles of the forward error divided by the asymptotic upper bound on the forward error, i.e., the backward error multiplied with the condition number. The figure thus shows for every s that in at least 90% of the cases the asymptotic upper bound on the forward error was at most 10 times

FIGURE 6.2. *The condition number for the sequences (6.2) and (6.3).*
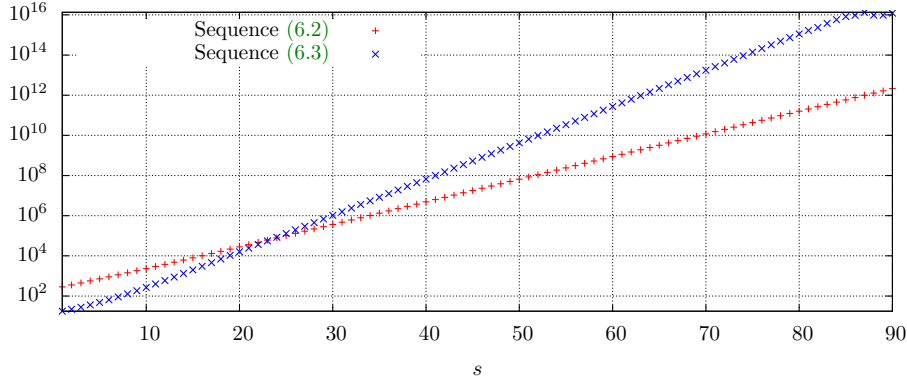
larger than the actual forward error. The curve of the 9th decile even shows that in 10% of the cases for every s the observed forward error was at least $10^{-0.2} \approx 0.6$ times the asymptotic upper bound on the forward error. This provides solid evidence that the condition number multiplied with the backward error *is not an overly pessimistic estimate of the forward error.* In fact, in this particular model *the forward error of the computed CPD is only fractionally less than the asymptotic upper bound.* It follows from Table 6.1 that the backward error, which was approximately $\sqrt{2} \cdot 10^{-7}$ in every experiment, multiplied with the condition number more accurately estimates the forward error than the backward error—the latter is still the dominant criterion for evaluating the quality of CPDs in absence of the forward error.

6.2. Open boundary points. In the next experiments the condition number is computed for a sequence of rank-r tensors that converges to an open boundary point, i.e., a tensor of rank at least $r + 1$. According to Theorem 1.4, the condition number should increase without bound along this sequence.

For brevity we consider the examples of Paatero [35, Section 7], and de Silva and Lim [20, Theorem 1.1]. The first sequence consists of the rank-3 tensors

$$\mathfrak{A}(s) = -\, 2^{\frac{3s}{16}} (\mathbf{a}_1^1 + \mathbf{a}_2^1) \otimes \mathbf{a}_1^2 \otimes \mathbf{a}_1^3 + 2^{\frac{3s}{16}} \mathbf{a}_2^1 \otimes (\mathbf{a}_1^2 + 2^{-\frac{3s}{16}-1} \mathbf{a}_3^2) \otimes (\mathbf{a}_1^3 + 2^{-\frac{3s}{16}-1} \mathbf{a}_3^3)$$

$$(6.2) \qquad + 2^{\frac{3s}{16}} (\mathbf{a}_1^1 + 2^{-\frac{3s}{16}-1} \mathbf{a}_3^1) \otimes (\mathbf{a}_1^2 + 2^{-\frac{3s}{16}-1} \mathbf{a}_2^2) \otimes (\mathbf{a}_1^3 + 2^{-\frac{3s}{16}-1} \mathbf{a}_2^3),$$

where $A_k = [\mathbf{a}_i^k]_{i=1}^3 \in \mathbb{R}^{m_k \times 3}$ and $A_k \sim N(0, 1)$. We took $(m_1, m_2, m_3) = (5, 4, 3)$ so that $\mathfrak{A}(s)$ has a unique rank-3 decomposition for all s with probability 1. As $s \to \infty$, $\mathfrak{A}(s)$ tends to a rank-5 tensor, which even has a positive-dimensional family of rank-5 CPDs [35, Section 7]. The sequence of rank-2 tensors from [20] is

$$(6.3) \quad \mathfrak{B}(s) = 2^{\frac{s}{5}} (\mathbf{b}_1^1 + 2^{-\frac{s}{5}} \mathbf{b}_2^1) \otimes (\mathbf{b}_1^2 + 2^{-\frac{s}{5}} \mathbf{b}_2^2) \otimes (\mathbf{b}_1^3 + 2^{-\frac{s}{5}} \mathbf{b}_2^3) - 2^{\frac{s}{5}} \mathbf{b}_1^1 \otimes \mathbf{b}_1^2 \otimes \mathbf{b}_1^3,$$

where $B_k = [\mathbf{b}_i^k]_{i=1}^2 \in \mathbb{R}^{m_k \times 2}$ and $B_k \sim N(0, 1)$. We took $(m_1, m_2, m_3) = (5, 3, 2)$ so that $\mathfrak{B}(s)$ has a unique rank-2 decomposition for all s with probability 1. The limit for $s \to \infty$ of this sequence is a rank-3 tensor. In both cases, we randomly sampled the factor matrices A_k and B_k and then varied s. For every $s = 1, 2, \ldots, 90$, the condition number was computed.

The results are summarized in Figure 6.2. Both graphs show greatly increasing condition numbers as s increases, consistent with Theorem 1.4. Comparing Figure 6.2 with Figure 2 from [40] reveals qualitatively similar behavior of the condition numbers in [40, Theorem 1] and Theorem 1.1.

FIGURE 6.3. *The base-10 logarithm $\log_{10}$ of the condition number for Paatero's model* (6.4). *The grid consists of 100×100 points equally spaced in $[-0.4, 0.4] \times [-0.4, 0.4]$.*
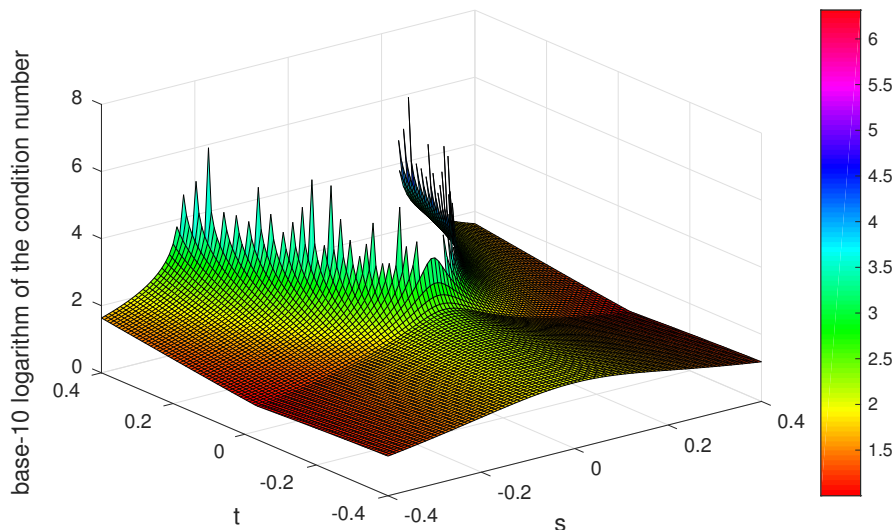
From plotting the condition numbers of the sequence (6.2) and (6.3) we are convinced that for increasing s the respective decompositions become ill-posed. But *how* they become ill-posed is not so evident. To get a clearer picture we consider another experiment. In [35, Equation (21)] Paatero gives the following parametrized family of $2 \times 2 \times 2$ tensors

$$(6.4) \quad \mathfrak{C}(s,t) = (\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1 + s\,\mathbf{e}_2 \otimes \mathbf{e}_2) \otimes \mathbf{e}_1 + (30\,\mathbf{e}_1 \otimes \mathbf{e}_1 - t\,\mathbf{e}_2 \otimes \mathbf{e}_2) \otimes \mathbf{e}_2,$$

where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the standard basis vectors in $\mathbb{R}^2$. Paatero shows that for $\frac{15}{2}s^2 < t$ the tensor $\mathfrak{C}(s,t)$ has rank 3, whereas for $\frac{15}{2}s^2 > t$ the tensor $\mathfrak{C}(s,t)$ has rank 2. Note that $\dim \mathbb{R}^{2\times2\times2} = 8$ and $8 = 2\dim\mathcal{S}$. The generic fibers of $\Phi : \mathcal{S}^{\times 3} \to \mathbb{R}^{2\times2\times2}$ are positive-dimensional and, consequently, the condition number of a decomposition of a generic rank-3 tensor is ill-posed. The parabola $\frac{15}{2}s^2 = t$ consists of open boundary points. Hence, the condition number must increase when heading towards this parabola from the side $\frac{15}{2}s^2 > t$.

The experiment we performed is as follows. On a 100×100 uniformly spaced grid in $[-0.4, 0.4] \times [-0.4, 0.4]$, we computed CPDs for $\mathfrak{C}(s,t)$ using the direct decomposition algorithm `cpd_gevd` from Tensorlab v3.0 [41]. Then we computed the condition number of the decompositions. The outcome of this experiment can be seen in Figure 6.3. Note in particular spikes close to the parabola $\frac{15}{2}s^2 = t$.

7. Conclusions. A local condition number for the join decomposition problem on $\mathcal{J} = \mathrm{Join}(\mathcal{M}_1, \ldots, \mathcal{M}_r)$ was presented in this paper. We gave both an easily computable spectral characterization as well as a characterization as an inverse projection distance to a locus of ill-posed problems in an auxiliary product Grassmannian. We believe that one main application of the condition number lies in the infamous rule of thumb of numerical analysis in (1.4).

An immediate consequence of Theorem 1.1 is that a join decomposition $\mathfrak{p}$ has $\kappa(\mathfrak{p}) = \infty$ when $\Phi(\mathfrak{p})$ locally has a smooth positive-dimensional family of alternative decompositions $\mathcal{E}$ so that $\Phi(\mathfrak{p}) = \Phi(\mathcal{E})$; this was stated as Corollary 1.2.

The JDP is ill-posed at open boundary points. Provided that the join set satisfies some technical conditions, Theorem 1.4 proved that open boundary points are completely surrounded by ill-conditioned JDPs: on a sequence of join decompositions tending to an open boundary point the condition number tends to infinity.

19

Two examples of join decompositions in the context of tensors were investigated more closely: the CPD and the Waring decomposition. For these examples, a closed expression of U in Theorem 1.1 was given, which is easily amenable to a computer implementation. We additionally presented closed expressions of the condition number for certain subclasses of tensors.

We hope that the proposed condition number will find application in the analysis of join decompositions and their JDPs. In fact, we have recently shown in [11] that the condition number naturally appears in the convergence analysis of certain Riemannian optimization methods for solving (JAP). Moreover, in [12] we demonstrated that great computational savings can be realized for solving the canonical tensor rank approximation problem compared to state-of-the-art algorithms in Tensorlab v3.0 by exploiting information about the condition number of the CPD in Riemannian optimization algorithms.

REFERENCES

[1] H. ABO, G. OTTAVIANI, AND C. PETERSON, *Induction for secant varieties of Segre varieties*, Trans. Amer. Math. Soc., 361 (2009), pp. 767–792, https://doi.org/10.1090/S0002-9947-08-04725-9.

[2] P.-A. ABSIL AND J. MALICK, *Projection-like retractions on matrix manifolds*, SIAM J. Optim., 22 (2012), pp. 135–158, https://doi.org/10.1137/100802529.

[3] J. ALEXANDER AND A. HIRSCHOWITZ, *Polynomial interpolation in several variables*, J. Algebraic Geom., 4 (1995), pp. 201–222.

[4] E. S. ALLMAN, C. MATIAS, AND J. A. RHODES, *Identifiability of parameters in latent structure models with many observed variables*, Ann. Statist., 37 (2009), pp. 3099–3132, https://doi.org/10.1214/09-AOS689.

[5] D. AMELUNXEN AND P. BÜRGISSER, *A coordinate-free condition number for convex programming*, SIAM J. Optim., 22 (2012), pp. 1029–1041, https://doi.org/10.1137/110835177.

[6] 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), pp. 2773–2832.

[7] V. BENIC AND S. GORJANC, *$(1, n)$ Congruences*, KoG: Scientific and Professional Journal of the Croatian Society for Geometry and Graphics, 10 (2007), pp. 5–12.

[8] G. BLEKHERMAN AND Z. TEITLER, *On maximum, typical, and generic ranks*, Math. Ann., 362 (2015), pp. 1021–1031, https://doi.org/10.1007/s00208-014-1150-3.

[9] L. BLUM, F. CUCKER, M. SHUB, AND S. SMALE, *Complexity and real computation*, Springer-Verlag, New York, 1998.

[10] J. BOCHNAK, M. COSTE, AND M. ROY, *Real Algebraic Geometry*, Springer–Verlag, 1998.

[11] P. BREIDING AND N. VANNIEUWENHOVEN, *Convergence analysis of Riemannian Gauss–Newton methods and its connection with the geometric condition number*, Appl. Math. Letters, (2017).

[12] P. BREIDING AND N. VANNIEUWENHOVEN, *A Riemannian trust region method for the canonical tensor rank approximation problem*, arXiv:1709.00033, (2017).

[13] P. BÜRGISSER, M. CLAUSEN, AND M. A. SHOKROLLAHI, *Algebraic Complexity Theory*, vol. 315 of Grundlehren der mathematischen Wissenshaften, Springer, Berlin, Germany, 1997.

[14] P. BÜRGISSER AND F. CUCKER, *Condition: The Geometry of Numerical Algorithms*, vol. 349 of Grundlehren der mathematischen Wissenschaften, Springer, Heidelberg, 2013.

[15] L. CHIANTINI AND G. OTTAVIANI, *On generic identifiability of 3-tensors of small rank*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1018–1037, https://doi.org/10.1137/110829180.

[16] 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), pp. 1265–1287, https://doi.org/10.1137/140961389.

[17] L. CHIANTINI, G. OTTAVIANI, AND N. VANNIEUWENHOVEN, *Effective criteria for specific identifiability of tensors and forms*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 656–681, https://doi.org/10.1137/16M1090132.

[18] L. CHIANTINI, G. OTTAVIANI, AND N. VANNIEUWENHOVEN, *On generic identifiability of symmetric tensors of subgeneric rank*, Trans. Amer. Math. Soc., 369 (2017), pp. 4021–4042, https://doi.org/10.1090/tran/6762.

[19] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1033–1066, https://doi.org/10.1137/070690729.

[20] V. de Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127, https://doi.org/10.1137/06066518X.

[21] J. W. Demmel, *The geometry of ill-conditioning*, J. Complexity, 3 (1987), pp. 201–229, https://doi.org/10.1016/0885-064X(87)90027-6.

[22] I. Domanov and L. De Lathauwer, *On the uniqueness of the canonical polyadic decomposition of third-order tensors—part II: Uniqueness of the overall decomposition*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 876–903, https://doi.org/10.1137/120877258.

[23] I. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, Resultants and Multidimensional Determinants*, Modern Birkhäuser Classics, Birkhäuser, 1994.

[24] G. H. Golub and C. van Loan, *Matrix Computations*, The John Hopkins University Press, 1996.

[25] L. Han and M. Neumann, *Handbook of Linear Algebra*, Discrete Mathematics and its Applications, Chapman & Hall/CRC, Boca Raton, FL, 2007, ch. 5: Inner Product Spaces, Orthogonal Projection, Least Squares, and Singular Value Decomposition, pp. 5–1–5–16.

[26] J. Harris, *Algebraic Geometry, A First Course*, vol. 133 of Graduate Text in Mathematics, Springer-Verlag, 1992.

[27] N. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 1996.

[28] T. Jiang and N. D. Sidiropoulos, *Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints*, IEEE Trans. Signal Process., 52 (2004), pp. 2625–2636, https://doi.org/10.1109/TSP.2004.832022.

[29] T. G. Kolda, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255, https://doi.org/10.1137/S0895479800368354.

[30] P. M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley series in probability and statistics, John Wiley & Sons, Hoboken, New Jersey, 2008.

[31] J. B. Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138, https://doi.org/10.1016/0024-3795(77)90069-6.

[32] J. Landsberg and J. Weyman, *On the ideals and singularities of secant varieties of segre varieties*, Bull. London Math. Soc., 39 (2007), pp. 685–697, https://doi.org/10.1112/blms/bdm049.

[33] J. M. Landsberg, *Tensors: Geometry and Applications*, vol. 128 of Graduate Studies in Mathematics, AMS, Providence, Rhode Island, 2012.

[34] J. M. Lee, *Introduction to Smooth Manifolds*, vol. 218 of Graduate Texts in Mathematics, Springer, New York, USA, 2 ed., 2013.

[35] P. Paatero, *Construction and analysis of degenerate PARAFAC models*, J. Chemometr., 14 (2000), pp. 285–299, https://doi.org/10.1002/1099-128X(200005/06)14:3⟨285::AID-CEM584⟩3.0.CO;2-1.

[36] J. Ponce, B. Sturmfels, and M. Trager, *Congruences and concurrent lines in multi-view geometry*, Avd. Appl. Math., 88 (2017), pp. 62–91, https://doi.org/10.1016/j.aam.2017.01.001.

[37] A. Seigal and B. Sturmfels, *Real rank two geometry*, Journal of Algebra, 484 (2017), pp. 310–333, https://doi.org/10.1016/j.jalgebra.2017.04.014.

[38] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Trans. Signal Proc., 65 (2017), pp. 3551–3582, https://doi.org/10.1109/TSP.2017.2690524.

[39] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis: Applications in the Chemical Sciences*, John Wiley & Sons, Hoboken, New Jersey, 2004.

[40] N. Vannieuwenhoven, *A condition number for the tensor rank decomposition*, Linear Algebra Appl., 353 (2017), pp. 35–86, https://doi.org/10.1016/j.laa.2017.08.014.

[41] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab v3.0*, March 2016, http://www.tensorlab.net.

[42] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

[43] M. Yang, *On partial and generic uniqueness of block term tensor decompositions*, Ann. Univ. Ferrara Sez. VII (N.S.), (2013), pp. 1–29, https://doi.org/10.1007/s11565-013-0190-z.

[44] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550, https://doi.org/10.1137/S0895479899352045.