Condition numbers for the tensor rank decomposition

Nick Vannieuwenhoven[1]

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

Abstract

The tensor rank decomposition problem consists of recovering the parameters of the model from an identifiable low-rank tensor. These parameters are analyzed and interpreted in many applications. As tensors are often perturbed by measurement errors in practice, one must investigate to what extent the unique parameters change in order to preserve the validity of the analysis. The magnitude of this change can be bounded asymptotically by the product of the condition number and the norm of the perturbation to the tensor. This paper introduces a condition number that admits a closed expression as the inverse of a particular singular value of Terracini's matrix, which represents the tangent space to the set of tensors of fixed rank. A practical algorithm for computing this condition number is presented. The latter's elementary properties such as scaling and orthogonal invariance are established. Rank-1 tensors are always well-conditioned. The class of weak 3-orthogonal tensors, which includes orthogonally decomposable tensors, contains both well-conditioned and ill-conditioned problems. Numerical experiments confirm that the condition number yields a good estimate of the magnitude of the change of the parameters when the tensor is perturbed. They also suggest that the condition number can be inversely related to the distance to ill-posed tensor rank decomposition problems, where the ill-posedness arises either from the nonclosedness of the set of tensors of fixed rank or from the existence of infinitely many decompositions.

Keywords: tensor rank decomposition, CP, condition number, stability analysis, Terracini's matrix
2010 MSC: 65F35, 15A69, 15A72, 65H04, 14Q15, 14N05, 58A05, 58A07

1. Introduction

A *tensor* is an element of $\mathrm{T} = \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d}$, where $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$, $n_k \in \mathbb{N}$, and $\otimes$ denotes the tensor product. A *rank*-1 *tensor* in T is defined as $\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d$ with $\mathbf{a}^k \in \mathbb{F}^{n_k} \setminus \{0\}$. Every tensor is expressible as a linear combination of rank-1 tensors:

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

If the number of terms r is minimal, then this decomposition due to Hitchcock [29] is called a *tensor rank decomposition*. The integer r is then called the *rank* of the tensor.

The *representative* $\mathfrak{p} = (\mathbf{a}^1, \ldots, \mathbf{a}^d)$ of a rank-1 tensor $\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d$ is called *essentially unique* or *identifiable* because the vectors $\mathbf{a}^k$ are unique up to scaling: if $\mathfrak{q} = (\mathbf{b}^1, \ldots, \mathbf{b}^d)$ is such that $\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d = \mathbf{b}^1 \otimes \cdots \otimes \mathbf{b}^d$, then $\mathbf{b}^k = \lambda_k \mathbf{a}^k$ for $k = 1, \ldots, d$, and furthermore $\lambda_1 \cdots \lambda_d = 1$. A rank-$r$ tensor with $d \geq 3$ often admits an *r-identifiable* decomposition [16, 23, 34]: the set of rank-1 tensors appearing in a decomposition is uniquely determined. For instance, if $d = 3$ the well-known criterion due to Kruskal [34] states that (1) is the essentially unique decomposition if $r \leq \frac{1}{2}(k_1 + k_2 + k_3 - 2)$, where k_i is the *Kruskal*

Email address: nick.vannieuwenhoven@cs.kuleuven.be (Nick Vannieuwenhoven)
[1]The author was supported by a Postdoctoral Fellowship of the Research Foundation—Flanders (FWO).

rank of $A_i = \{\mathbf{a}_j^i\}_{j=1}^r$; that is, the integer k_i such that every subset of A_i of cardinality k_i forms a linearly independent set.

Identifiability of a tensor rank decomposition is essential in several applications, such as in the parameter identification problem in latent variable models [3, 4]. Models whose parameters can be inferred with this decomposition and symmetrical variants were recently surveyed in a unified tensor-based framework [4]; they include exchangeable single topic models, hidden Markov models, and Gaussian mixture models. The blind source separation problem, or independent component analysis, in signal processing is another example where the uniqueness of the symmetric tensor rank decomposition of a cumulant tensor is key in identifying the statistically independent signals [18].

In all of the above applications, a mathematical theory guarantees that the tensor $\mathfrak{A}$ has a rank-r decomposition; however, in practice this tensor is often estimated from imperfect measurements. This entails that the true but unknown tensor $\mathfrak{A}$ is approximated by the estimated tensor $\mathfrak{A}''$ whose rank is in general not equal to r. Therefore, it is common practice to approximate $\mathfrak{A}''$ by a nearby rank-r tensor $\mathfrak{A}'$. The latter's rank-r decomposition then serves as a proxy for $\mathfrak{A}$'s decomposition. This is a standard approach in data-analytic applications, yet one primordial question has thus far remained unanswered:

> What is the relationship between the essentially unique decompositions of two rank-r tensors $\mathfrak{A}$ and $\mathfrak{A}'$ when $\|\mathfrak{A} - \mathfrak{A}'\| \leq \epsilon$ for small perturbations $\epsilon > 0$?

The goal of this paper is characterizing that relationship to first order via a study of the *condition number* of the associated computational problem of computing a tensor rank decomposition.

1.1. Condition numbers

The structured absolute condition number of a computational problem g is a standard concept in numerical analysis [13, 28, 48, 60]. It measures the maximum change at the output of a function $g : \mathcal{D} \subset \mathbb{F}^m \to \mathbb{F}^n$ when infinitesimally perturbing the input $\mathbf{x} \in \mathcal{D}$ to another $\mathbf{x} + \Delta\mathbf{x} \in \mathcal{D}$:

$$\kappa_{\alpha,\beta}(\mathbf{x}) := \lim_{\epsilon \to 0} \max_{\substack{\|\Delta\mathbf{x}\| \leq \epsilon, \\ \mathbf{x}+\Delta\mathbf{x} \in \mathcal{D}}} \frac{\|g(\mathbf{x}+\Delta\mathbf{x}) - g(\mathbf{x})\|_\alpha}{\|\Delta\mathbf{x}\|_\beta}, \tag{2}$$

where $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are norms on $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively, and $\mathbf{x} \in \mathcal{D} \subset \mathbb{F}^m$. The condition number yields an asymptotically sharp a posteriori bound on the absolute forward error:

$$\|g(\mathbf{x}+\Delta\mathbf{x}) - g(\mathbf{x})\|_\alpha \lesssim \kappa_{\alpha,\beta}(\mathbf{x}) \cdot \|\Delta\mathbf{x}\|_\beta, \tag{3}$$

where $\mathbf{x}, \mathbf{x}+\Delta\mathbf{x} \in \mathcal{D}$, and where $\|\Delta\mathbf{x}\|_\beta$ should be sufficiently close to zero. That is, *the absolute forward error is asymptotically bounded by the product of the absolute backward error multiplied with the condition number*, which is a well-known rule of thumb in numerical analysis.

Unfortunately, the computational problem of computing a tensor rank decomposition admits no known closed expression in general. Therefore, its condition number will be investigated via the study of its much simpler inverse problem, which asks to compute the tensor given the parameters of its rank-r decomposition (1). This problem can be described explicitly as a "tensor computation" function f. In the literature, the parameters are succinctly represented by the *factor matrices* $A_i = [\mathbf{a}_j^i]_{j=1}^r$; however, dealing with their vectorizations will be more convenient in this paper. Therefore, we define the bijections

$$\begin{aligned} &\operatorname{vec} : \mathbb{F}^{n_1} \times \cdots \times \mathbb{F}^{n_d} \to \mathbb{F}^{\Sigma+d} && \operatorname{vecr} : (\mathbb{F}^{n_1} \times \cdots \times \mathbb{F}^{n_d})^{\times r} \to \mathbb{F}^{r(\Sigma+d)} \\ &\mathfrak{p}_i = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d) \mapsto \begin{bmatrix} \mathbf{a}_i^1 \\ \vdots \\ \mathbf{a}_i^d \end{bmatrix}, \quad \text{and} && (\mathfrak{p}_1, \ldots, \mathfrak{p}_r) \mapsto \begin{bmatrix} \operatorname{vec}(\mathfrak{p}_1) \\ \vdots \\ \operatorname{vec}(\mathfrak{p}_r) \end{bmatrix}, \end{aligned}$$

where $\Sigma = \sum_{k=1}^d (n_k - 1)$. Let $\Pi = \prod_{k=1}^d n_k$ and $M = r \cdot \sum_{k=1}^d n_k$. The tensor computation function is then

$$f : \mathbb{F}^M \to \mathbb{F}^\Pi, \quad \mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_r) \mapsto \sum_{i=1}^r \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d, \tag{4}$$

which takes the *vectorized factor matrices* (VFMs) $\mathbf{p}$ as input and sends them to the tensor they represent.

1.2. The decomposition problem versus the approximation problem

This paper is principally concerned with the condition number of the tensor rank *decomposition* problem:[2]

$$\text{Given a rank-}r\text{ tensor }\mathcal{X}\in\mathbb{F}^{\Pi},\ \text{find }\mathbf{x}\in\mathbb{F}^{M}\text{ so that }\mathcal{X}=f(\mathbf{x}). \tag{TDP}$$

The input to this problem is always a rank-r tensor. Therefore, we study the *structured* condition number, which accounts only for structured perturbations that take one rank-r tensor to another rank-r tensor. Mathematically, we consider the condition number in (2) where $\mathcal{D}$ is the set of tensors of rank equal to r, and $g:\mathcal{D}\to\mathbb{F}^{M}$ maps a rank-r tensor $\mathcal{X}$ to VFMs $\mathbf{x}$ so that $\mathcal{X}=f(\mathbf{x})$.

One could complementarily investigate the condition number of the tensor rank-r *approximation* problem:

$$\text{Given a tensor }\mathcal{X}\in\mathbb{F}^{\Pi},\ \text{find }\mathbf{x}\in\mathbb{F}^{M}\text{ so that }\|\mathcal{X}-f(\mathbf{x})\|^{2}=\min_{\mathbf{y}\in\mathbb{F}^{M}}\|\mathcal{X}-f(\mathbf{y})\|^{2}. \tag{TAP}$$

The input now consists of any element of $\mathbb{F}^{\Pi}$. The corresponding condition number would be as in (2) with $\mathcal{D}=\mathbb{F}^{\Pi}$ and $g:\mathbb{F}^{\Pi}\to\mathbb{F}^{M}$ maps a tensor $\mathcal{X}$ to the parameters of its best rank-r approximation (assuming it exists). Such a condition number measures the sensitivity of the best rank-r approximation of $\mathcal{X}$ to arbitrary perturbations to $\mathcal{X}$.

As this paper focuses on the first condition number, it is important to understand when it is appropriate. Consider the situation where an unknown theoretical tensor $\mathcal{X}$ admits a unique rank-r decomposition, i.e., $\mathcal{X}=f(\mathbf{x})$, but where $\mathcal{X}$ is not known exactly due to the occurrence of measurement, model, representation, or computation errors. An example of this is presented in the next subsection. In this case, only an approximation $\widetilde{\mathcal{X}}\approx\mathcal{X}$ is known. In general, $\widetilde{\mathcal{X}}$ does not admit an exact rank-r decomposition. However, it can be well approximated by one, say $\widehat{\mathcal{X}}=f(\widehat{\mathbf{x}})$. Even though we need to solve the TAP in this step, if we are interested in understanding the true VFMs $\mathbf{x}$ of $\mathcal{X}$, then *the condition number of the TDP at $\widehat{\mathcal{X}}$ is the only quantity of interest* because, by definition, it bounds the error between the computed VFMs $\widehat{\mathbf{x}}$ and the true but unknown VFMs $\mathbf{x}$. The condition number of the TAP with input $\widetilde{\mathcal{X}}$ (or $\widehat{\mathcal{X}}$) offers no information of this kind, as it looks at unstructured perturbations while in this application we know that we seek a nearby tensor with an exact rank-r decomposition. Note that at a rank-r tensor $\mathcal{X}$, the condition number of the TAP at $\mathcal{X}$ is always bounded from below by the condition number of the TDP at $\mathcal{X}$ by definition.

1.3. Applications

The condition number is expected to find application in data analysis for estimating the forward error of the TDP, as is explained next. Recall that decomposition (1) was employed for data analysis in fluorescence spectroscopy by Appellof and Davidson [5]. In a diluted solution, fluorophores will absorb a certain amount of light at the ith wavelength and re-emit a fraction of the absorbed energy as light at the jth wavelength [53, Section 2.3]. Mathematically, $x_{i,j}=\chi\lambda_i\mu_j$, where $x_{i,j}$ is the intensity of the light emitted at the jth wavelength when the fluorophore is excited by light at the ith wavelength, λ_i models the fraction of light absorbed at the ith wavelength, μ_j models the fraction of light emitted at the jth wavelength, and χ is a chemical constant proportional to the concentration of the fluorophore in the solution. If r fluorophores occur jointly in a solution and they are chemically inert, then their signal is additive: $x_{i,j}=\sum_{\ell=1}^{r}\chi_\ell\lambda_{i,\ell}\mu_{j,\ell}$. If several of such mixtures with varying concentrations of fluorophores are simultaneously analyzed, the model for the intensity of the emitted light is a tensor rank decomposition: $x_{i,j,k}=\sum_{\ell=1}^{r}\chi_{k,\ell}\lambda_{i,\ell}\mu_{j,\ell}$. As stated in [53, Section 2.3], "the uniqueness property of [this model] guarantees that, if [it] is adequate, [...] the estimated loading vectors for the ℓth component are readily interpretable as concentration profiles [$\boldsymbol{\chi}_\ell$] and emission and excitation spectra ([$\boldsymbol{\mu}_\ell$] and [$\boldsymbol{\lambda}_\ell$] respectively) of the ℓth fluorophore." This enables an identification of the fluorophores present in a chemical mixture of unknown composition. While the theoretical tensor $\mathcal{X}=f(\mathbf{x})$ admits an exact rank decomposition with rank equal to the number of fluorophores, in practice only an approximation $\widetilde{\mathcal{X}}$ can be obtained from physical measurements. The reason is that such

[2]Ignore for now that this problem is ill-posed due to the non-uniqueness of the representation of rank-1 tensors. In Section 2, this will be addressed in detail.

measurements are always subject to errors because of the finite resolution of the equipment. In addition, the supposed model is only approximately valid in practice, as the assumptions of a sufficiently diluted solution and chemical inertness may not be satisfied exactly. This means that the data tensor $\widetilde{\mathfrak{X}}$ that is available for data analysis must be regarded as a perturbation of the theoretical rank-r tensor $\mathfrak{X}$ with $\|\mathfrak{X} - \widetilde{\mathfrak{X}}\|_F \leq \mu$ and where μ is known from the application domain. The data tensor $\widetilde{\mathfrak{X}}$ is then approximated by a rank-r tensor $\widehat{\mathfrak{X}} = f(\widehat{\mathbf{x}})$. Letting $\nu := \|\widehat{\mathfrak{X}} - \widetilde{\mathfrak{X}}\|_F$, we know that the distance between $\mathfrak{X}$ and the approximation $\widehat{\mathfrak{X}}$ computed from the sampled data $\widetilde{\mathfrak{X}}$ is at most $\mu + \nu$. Armed with the condition number κ of the TDP at $\widehat{\mathfrak{X}} = f(\widehat{\mathbf{x}})$, we can compute the asymptotically sharp bound $\|\mathbf{x} - \widehat{\mathbf{x}}\| \lesssim \kappa \cdot (\nu + \mu)$, which bounds the difference between the computed approximate VFMs $\widehat{\mathbf{x}}$ and the true but unknown VFMs $\mathbf{x}$. If $\kappa \cdot (\nu + \mu)$ is small, then one can safely match the computed rank-1 terms with chemical compounds, as the true rank-1 terms in the decomposition of the true tensor $\mathfrak{X}$ will have been well approximated.

1.4. Contributions

The main contribution of this paper is the introduction of a condition number for the TDP. It is defined for every rank decomposition that satisfies a mild condition called *robust r-identifiability*. This condition is satisfied by generic[3] rank-r decompositions in tensor spaces that are generically r-identifiable; see Section 4.1. The prime result, Theorem 10, can be stated informally as follows.

Theorem 1. *Let $N' = \min\{r(\Sigma + d), \Pi\}$, $N = r(\Sigma + 1) < N' \leq \Pi$, and let $\mathfrak{A} = f(\mathbf{x})$ be a robustly r-identifiable tensor in $\mathbb{F}^{n_1 \times n_2 \times \cdots \times n_d}$. If $J = [\partial f_i / \partial x_j]_{i,j}$ is the $\Pi \times r(\Sigma + d)$ Jacobian matrix of f evaluated at $\mathbf{x}$ and $\varsigma_1 \geq \varsigma_2 \geq \cdots \geq \varsigma_N \geq \varsigma_{N+1} = \cdots = \varsigma_{N'} = 0$ are its singular values, then $\overline{\kappa}(\mathbf{x}) := \varsigma_N^{-1}$ can be interpreted as the absolute condition number of the tensor rank decomposition problem at $\mathbf{x}$.*

This Jacobian matrix is required in every step of gradient-based optimization methods for computing approximate tensor rank decompositions; see, e.g., [2, 27, 39–42, 46, 54]. In these methods, the condition number can thus be obtained with little extra effort by computing the singular values of the Jacobian matrix in the final iteration. Nevertheless, the condition number is a property of a computational problem, so it does not depend on the method for solving it. In particular, the condition number of rank decompositions obtained via ALS-type methods or via direct decomposition algorithms [6, 19, 36] is also governed by aforementioned Jacobian matrix, even though this matrix is nowhere employed in these methods.

1.5. Prior work

The Cramér–Rao bounds that were investigated in [32, 37, 51, 52] measure the stability of a tensor rank decomposition in a statistical framework, wherein additive Gaussian noise is assumed to corrupt the factor matrices. The quantity of interest is the squared angular error between the true parameters and the estimated parameters [52]. The Cramér–Rao lower bound (CRLB) for estimating the parameters $\mathbf{p}$ of the tensor rank decomposition of $\mathfrak{A} = f(\mathbf{p})$ is then defined in [52] as the inverse of the Fisher information matrix $\frac{1}{\sigma^2} J^T J$, where J is as in Theorem 1 and σ^2 is the variance of the Gaussian noise. However, this matrix is not invertible, so [52] suggest to alter the TDP by removing some of the parameters. Hence, the CRLB depends on the particular elimination of the variables that is chosen.

1.6. Outline

The rest of this paper is structured as follows. In the next section, the classic theory for conditioning of implicit inverse functions is stated. From geometrical insights it is established that this theory does not apply unreservedly. Three strategies are identified for overcoming this problem, two of which are developed in succeeding sections. Strategy I relies on the explicit elimination of variables and is explored in Section 3. The main contribution of this paper consists of developing Strategy II, which applies the classic definition of a condition number in (2) to the TDP, albeit with respect to a suitable premetric; Section 4 presents this

[3] A property that can be admitted by the elements of a set S is called generic if all elements of $S \setminus N$ admit the property, where $N \subset S$ has measure zero in S in some topology.

development which results in a condition number of the TDP at specific VFMs $\mathbf{p}$. Based on these results, the condition number of norm-balanced VFMs is proposed in Section 5 as the condition number of the TDP at a rank-r tensor $\mathfrak{A}$, and its elementary properties are stated. Numerical experiments illustrate the behavior of the proposed condition number in Section 6. The paper is concluded in Section 7 by a discussion of the main conclusions.

1.7. Notation and conventions

Varieties are typeset in a calligraphic upper-case font ($\mathcal{S}, \mathcal{V}$), tensors in an upper-case fraktur font ($\mathfrak{A}$), matrices in upper-case letters (A, U, V), tuples of vectors in a lower-case fraktur font ($\mathfrak{p}$), vectors in boldface lower-case letters ($\mathbf{a}$, $\mathbf{x}$), and scalars and points on varieties in lower-case letters (a, b, λ, p, q). The symbol $\mathbb{F}$ denotes either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. Let $\mathbb{F}_0^m := \mathbb{F}^m \setminus \{0\}$. For a matrix $A \in \mathbb{F}^{m\times n}$, A^T is its transpose, A^H is its conjugate transpose, and $A^\dagger$ is its Moore–Penrose pseudoinverse. The Euclidean norm of $\mathbf{v} \in \mathbb{F}^m$ is $\|\mathbf{v}\| = \sqrt{\mathbf{v}^H\mathbf{v}}$, which is induced from the Hermitian inner product $\langle \mathbf{a}, \mathbf{b}\rangle = \mathbf{b}^H\mathbf{a}$. A block diagonal matrix with diagonal blocks $A_1, \ldots, A_d$ is denoted by $\operatorname{diag}(A_1, \ldots, A_d)$. By $\operatorname{span}(A)$ the column span of the matrix A is meant. The ith largest singular value of a matrix $A \in \mathbb{F}^{m\times n}$ is denoted by $\varsigma_i(A)$.

The prototypical tensor in this paper is the rank-r tensor $\mathfrak{A} \in \mathrm{T} = \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d}$. Hence, the scalar d is used exclusively for the order of $\mathfrak{A}$, r denotes its rank, the scalars n_1, n_2, …, n_d are reserved for $\mathfrak{A}$'s dimensions, and we also define

$$\Sigma = \sum_{k=1}^{d}(n_k - 1), \quad M = r(\Sigma + d), \quad N = r(\Sigma + 1), \quad \text{and} \quad \Pi = \prod_{k=1}^{d} n_k,$$

which are respectively the dimension of the variety of rank-1 tensors in T minus one, the dimension of the domain of f, the expected dimension of the smallest variety containing the tensors of rank r in T, and the dimension of T.

2. The classic approach to conditioning for implicit inverse functions

Given the parameters of a rank-r decomposition, constructing the corresponding tensor amounts to evaluating the smooth "tensor computation" function f in (4). Unfortunately, no explicit description is known of f^{-1} which sends a rank-r tensor to the parameters of its rank-r decomposition. In certain cases, the related function $f^\dagger$ which sends a tensor to the parameters of its best rank-r approximation can be known up to first order by invoking the Inverse Function Theorem [33, 47]. This suffices for characterizing the condition number in the 2-norm, as is shown next.

Let $g : \mathbb{R}^m \to \mathbb{R}^n$ be a smooth function with $m \le n$, and let $J_g(\mathbf{x})$ denote its Jacobian matrix evaluated at $\mathbf{x}$. Then, the local Taylor series expansion of g about $\mathbf{x}$ is $g(\mathbf{x} + \Delta\mathbf{x}) = g(\mathbf{x}) + J_g(\mathbf{x})\Delta\mathbf{x} + \mathcal{O}(\|\Delta\mathbf{x}\|^2)$. From the definition of the structured condition number in (2) it follows immediately that if $\mathcal{D}$ is an open neighborhood of $\mathbf{x}$ and $\alpha = \beta = 2$, then the absolute condition number of g at $\mathbf{x}$ is simply the spectral norm of the Jacobian of g at $\mathbf{x}$, because

$$\kappa(\mathbf{x}) = \lim_{\epsilon\to 0} \max_{\substack{\|\Delta\mathbf{x}\|\le\epsilon \\ \mathbf{x}+\Delta\mathbf{x}\in\mathcal{D}}} \frac{\|g(\mathbf{x}) + J_g(\mathbf{x})\Delta\mathbf{x} + \mathcal{O}(\|\Delta\mathbf{x}\|^2) - g(\mathbf{x})\|}{\|\Delta\mathbf{x}\|} = \|J_g(\mathbf{x})\|.$$

The following variant of the Inverse Function Theorem then results in a characterization of the condition number of the approximation problem associated with g. I was unable to locate this version in the literature, but it is probably known to the experts. For convenience, an extension of Spivak's proof in [47] is included.

Theorem 2. *Let $g : \mathbb{R}^m \to \mathbb{R}^n$ with $m \le n$ be a smooth (C^∞) function. Assume that the Jacobian matrix $J_g(\mathbf{x}_0) \in \mathbb{R}^{n\times m}$ has rank m at $\mathbf{x}_0$. Then, the set-valued map*

$$s : \mathbf{y} \mapsto \arg\min_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{y} - g(\mathbf{x})\|^2 \tag{5}$$

has a unique smooth single-valued localization $g^\dagger_{\mathbf{x}_0} : \mathbb{R}^n \to \mathbb{R}^m$ *in an open neighborhood of* $\mathbf{y}_0 = g(\mathbf{x}_0)$ *for which* $g^\dagger_{\mathbf{x}_0}(g(\mathbf{x}_0)) = \mathbf{x}_0$ *and whose Jacobian is*

$$J_{g^\dagger_{\mathbf{x}_0}}(\mathbf{y}_0) = \big(J_g(\mathbf{x}_0)\big)^\dagger,$$

where $A^\dagger = (A^TA)^{-1}A^T$ *is the Moore-Penrose pseudoinverse of* $A \in \mathbb{R}^{m\times n}$.

Proof. First, we construct an open set where $g^\dagger_{\mathbf{x}_0}$ is defined about $\mathbf{y}_0$. Since g is differentiable in a neighborhood of $\mathbf{x}_0$, we can expand g for sufficiently small Δ as follows:

$$g(\mathbf{x}_0 + \Delta) = g(\mathbf{x}_0) + J\Delta + \mathcal{O}(\|\Delta\|\|r(\Delta)\|),$$

where $J = J_g(\mathbf{x}_0) \in \mathbb{R}^{n\times m}$ is the Jacobian matrix of g at $\mathbf{x}_0$, and $\lim_{\Delta\to 0}\|r(\Delta)\| \to 0$. For every sufficiently small $\tau > 0$, there exists a δ_τ so that $\sup_{\|\Delta\|\le\delta_\tau}\|r(\Delta)\| \le \tau$. Consider then the open ball $U = \mathcal{B}_{\delta_\tau}$ of radius δ_τ centered at $\mathbf{x}_0$. Let $\mathbf{x}_0 + \Delta \in \partial\mathcal{B}_{\delta_\tau}$ be a point on the boundary. Then it follows that

$$\|g(\mathbf{x}_0 + \Delta) - g(\mathbf{x}_0)\| \ge \|J\Delta\| - C\delta_\tau\tau \ge \delta_\tau \cdot (\varsigma_m(J) - C\tau),$$

where C is a constant, and $\varsigma_m(J) > 0$ is the mth singular value of J, which is nonzero by assumption on the rank of J. Since τ was arbitrary, we can assume that $\tau \le \frac{1}{2C}\varsigma_m(J)$. Then,

$$\|g(\mathbf{x}_0 + \Delta) - g(\mathbf{x}_0)\| \ge \frac{1}{2}\delta_\tau\varsigma_m(J) = 2\epsilon > 0.$$

With this definition of ϵ, let $\mathcal{B}_\epsilon$ be the open ball of radius ϵ in $\mathbb{R}^n$ centered at $\mathbf{y}_0$.

Next, we show how $s(\mathbf{y})$ can be characterized locally as a solution of a nonlinear system of equations. Fix a $\mathbf{y} \in \mathcal{B}_\epsilon \subset \mathbb{R}^n$. Then, for every $\mathbf{x} \in \partial\mathcal{B}_\delta \subset \mathbb{R}^m$ we now have the inequality

$$\|\mathbf{y} - g(\mathbf{x})\| \ge \|g(\mathbf{x}_0) - g(\mathbf{x})\| - \|\mathbf{y} - g(\mathbf{x}_0)\| > 2\epsilon - \epsilon > \|\mathbf{y} - g(\mathbf{x}_0)\|,$$

where the first inequality is due to the triangle inequality. This implies that for all $\mathbf{y} \in \mathcal{B}_\epsilon$ the minimizers of $s(\mathbf{y})$ are all attained in the interior of U. This entails that the gradient of the smooth objective function $h_\mathbf{y}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - g(\mathbf{x})\|^2$ vanishes at each minimizer $\mathbf{x}_i \in s(\mathbf{y})$ [38]. The stationary points of $h_\mathbf{y}(\mathbf{x})$ can be obtained as the zero locus of the smooth map

$$\phi : \mathcal{B}_\epsilon \times U \to \mathbb{R}^m,\ (\mathbf{y}, \mathbf{x}) \mapsto (J_g(\mathbf{x}))^T(\mathbf{y} - g(\mathbf{x})),$$

where the Jacobian matrix $J_g(\mathbf{x}) \in \mathbb{R}^{n\times m}$ is of rank m in a neighborhood of $\mathbf{x}_0$. Hence, we can state that

$$s(\mathbf{y}) \subset S(\mathbf{y}) := \{\mathbf{x} \in U \mid \phi(\mathbf{y}, \mathbf{x}) = 0\}$$

for all $\mathbf{y} \in \mathcal{B}_\epsilon \subset \mathbb{R}^n$. By assumption, $\phi(\mathbf{y}_0, \mathbf{x}_0) = 0$. The implicit function theorem [24, Theorem 1B.1] states that if the partial derivative of ϕ with respect to the second argument is of full rank at $(\mathbf{y}_0, \mathbf{x}_0)$, then there exists a smooth function $g^\dagger_{\mathbf{x}_0}$ that corresponds with a single-valued localization around $\mathbf{x}_0$ of the set-valued solution map $S(\mathbf{y})$ in an open neighborhood of $\mathbf{y}_0 \in \mathbb{R}^n$, i.e., $g^\dagger_{\mathbf{x}_0}(\mathbf{y}_0) = \mathbf{x}_0$. In addition, the Jacobian of $g^\dagger_{\mathbf{x}_0}$ is given by

$$J_{g^\dagger_{\mathbf{x}_0}}(\mathbf{y}) = -\big(\nabla_\mathbf{x}\phi(\mathbf{y}, g^\dagger_{\mathbf{x}_0}(\mathbf{y}))\big)^{-1}\nabla_\mathbf{y}\phi(\mathbf{y}, g^\dagger_{\mathbf{x}_0}(\mathbf{y})),$$

where $\nabla_\mathbf{x}\phi$ denotes the partial derivative of ϕ with respect to $\mathbf{x}$ and similarly for $\nabla_\mathbf{y}\phi$. By the chain rule,

$$\nabla_\mathbf{x}\phi(\mathbf{y}_0, \mathbf{x}_0) = \big(\nabla_\mathbf{x} J_g(\mathbf{x})^T\big)(\mathbf{x}_0) \cdot (\mathbf{y}_0 - g(\mathbf{x}_0)) + J_g(\mathbf{x}_0)^T(-J_g(\mathbf{x}_0)) = -(J_g(\mathbf{x}_0))^T J_g(\mathbf{x}_0).$$

As $J_g \in \mathbb{R}^{n\times m}$ is of rank m in a neighborhood of $\mathbf{x}_0$ by semicontinuity of matrix rank, it follows that the partial derivative of ϕ to $\mathbf{x}$ is also of full rank at $(\mathbf{y}_0, \mathbf{x}_0)$ because it is the Gram matrix of J_g. So, the implicit function theorem applies, and we find that there is a local inverse function $g^\dagger_{\mathbf{x}_0}$ defined in a neighborhood of $\mathbf{y}_0$ with $g^\dagger_{\mathbf{x}_0}(\mathbf{y}_0) = \mathbf{x}_0$ and whose Jacobian in $\mathbf{y}_0$ is

$$J_{g^\dagger_{\mathbf{x}_0}}(\mathbf{y}_0) = \big((J_g(\mathbf{x}_0))^T J_g(\mathbf{x}_0)\big)^{-1}(J_g(\mathbf{x}_0))^T = \big(J_g(\mathbf{x}_0)\big)^\dagger,$$

as it is easy to verify that $\nabla_\mathbf{y}\phi(\mathbf{y}_0, \mathbf{x}_0) = J_g(\mathbf{x}_0)^T I_n$. □

The following result is an immediate consequence of the above discussions.

Corollary 3. *If $g : \mathbb{R}^m \to \mathbb{R}^n$ with $m \leq n$ is a smooth function with a full rank Jacobian matrix $J_g(\mathbf{x}) \in \mathbb{R}^{n\times m}$ at $\mathbf{x}$, then the condition number of $g^\dagger_{\mathbf{x}}$ at $g(\mathbf{x})$ is $\kappa = \|(J_g(\mathbf{x}))^\dagger\| = \big(\varsigma_m(J_g(\mathbf{x}))\big)^{-1}$.*

Applying this corollary to the smooth tensor computation function f would thus characterize the condition number of the TAP for real[4] tensors. Unfortunately, it does not apply unreservedly because the Jacobian matrix of f is not of full rank due to the scaling indeterminacies. This is naturally understood from the (semi-)algebraic geometry of the tensor rank decomposition.

2.1. Geometry of the tensor rank decomposition I: Dimension

The relevant basic facts and definitions from (semi-)algebraic geometry are recalled next; the reader is referred to Landsberg [35] for greater detail.

Let $\mathbb{F}^n_0 := \mathbb{F}^n \setminus \{0\}$. The Segre variety over $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$ is the variety of rank-1 tensors. It is denoted by

$$\mathcal{S}_{\mathbb{F}} := \mathrm{Seg}(\mathbb{F}^{n_1}_0 \times \mathbb{F}^{n_2}_0 \times \cdots \times \mathbb{F}^{n_d}_0) \subset \mathbb{F}^{n_1} \otimes \mathbb{F}^{n_2} \otimes \cdots \otimes \mathbb{F}^{n_d} \cong \mathbb{F}^{\Pi},$$

and its elements can be parametrized explicitly via the Segre map:

$$\begin{aligned} \mathrm{Seg} : \mathbb{F}^{n_1} \times \mathbb{F}^{n_2} \times \cdots \times \mathbb{F}^{n_d} &\to \mathbb{F}^{n_1} \otimes \mathbb{F}^{n_2} \otimes \cdots \otimes \mathbb{F}^{n_d} \cong \mathbb{F}^{\Pi} \\ (\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^d) &\mapsto \mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d, \end{aligned}$$

where $\otimes$ is the tensor product and $\Pi = \prod_{k=1}^d n_k$. For notational brevity, I consider the Segre map that embeds into $\mathbb{F}^{\Pi}$ so that *the tensor product can be realized in coordinates as the Kronecker product*. The 2-factor Segre variety embedded in $\mathbb{F}^{n_1} \otimes \mathbb{F}^{n_2} \cong \mathbb{F}^{n_1 \times n_2}$ is the smooth variety of rank-1 matrices. The specific tuple $\mathfrak{p}_i = (\mathbf{a}^1_i, \mathbf{a}^2_i, \ldots, \mathbf{a}^d_i)$ in the representation of a point $p_i = \mathrm{Seg}(\mathfrak{p}_i) \in \mathcal{S}_{\mathbb{F}}$ is called a *representative* of p_i.

Let $r \geq 2$ be strictly *subgeneric*: $r < \frac{\Pi}{\Sigma+1}$. Then, let

$$\sigma^0_r(\mathcal{S}_{\mathbb{F}}) = \{p_1 + p_2 + \cdots + p_r \mid p_i \in \mathcal{S}_{\mathbb{F}}\} = \{\mathfrak{A} \in \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} \mid \mathrm{rank}_{\mathbb{F}}(\mathfrak{A}) \leq r\}$$

denote the set of tensors of $\mathbb{F}$-rank at most r. Its closure in the Euclidean topology is $\sigma_r(\mathcal{S}_{\mathbb{F}}) := \overline{\sigma^0_r(\mathcal{S}_{\mathbb{F}})}$. By definition, $\sigma^0_r(\mathcal{S}_{\mathbb{F}})$ is a dense, constructible subset of $\sigma_r(\mathcal{S}_{\mathbb{F}})$. It is known that $\sigma_r(\mathcal{S}_{\mathbb{F}})$ is an irreducible *algebraic variety*[5] if $\mathbb{F} = \mathbb{C}$ (see, e.g., [35, Corollary 5.1.1.5]) and a *semi-algebraic variety* if $\mathbb{F} = \mathbb{R}$ (see, e.g., [20, Theorem 6.2]). An algebraic variety is the common zero set of a system of polynomial equations over $\mathbb{C}$ (see, e.g., [26]) and a semi-algebraic set is the solution set of a system of polynomial equations and inequalities over $\mathbb{R}$; see [10].

At most points varieties and semi-algebraic sets locally resemble a linear space. One way to describe this local linear approximation to a variety or semi-algebraic set $\mathcal{V} \subset \mathbb{F}^{\Pi}$ at a point $p \in \mathcal{V}$ is the *tangent space*; see respectively [25, pp. 175–176] and [10, Section 3.3]. For a smooth variety $\mathcal{V} \subset \mathbb{F}^{\Pi}$, this vector space can be defined extrinsically as the span of all tangent vectors of analytic curves in $\mathcal{V}$ passing through p:

$$\mathrm{T}_p\mathcal{V} = \mathrm{span}\big(\{\tfrac{d}{dt}p(0) \mid p(t) \subset \mathcal{V},\ t \in (-1,1), p(0) = p\}\big).$$

For general varieties the tangent space is best defined algebraically; see, e.g., [25, 26]. The dimension of a variety $\mathcal{V} \subset \mathbb{F}^{\Pi}$ can be defined geometrically as the number n such that $\dim \mathrm{T}_p\mathcal{V} = n$ for all p in an open neighborhood $\mathcal{N}$ of $p_0 \in \mathcal{V}$. For a semi-algebraic set $\mathcal{V}$ the dimension is defined as the dimension of the Zariski-closure of $\mathcal{V}$ [10, Proposition 2.8.2]. The dimension of the Segre variety $\mathcal{S}_{\mathbb{F}} = \mathrm{Seg}(\mathbb{F}^{n_1}_0 \times \cdots \times \mathbb{F}^{n_d}_0) \subset \mathbb{F}^{\Pi}$ is well-known [26, 35]: $\dim \mathcal{S}_{\mathbb{F}} = \Sigma + 1$. The dimension of $\sigma_r(\mathcal{S}_{\mathbb{F}})$ is known only conjecturally in general. As $\sigma^0_r(\mathcal{S}_{\mathbb{F}})$ is Euclidean-dense in $\sigma_r(\mathcal{S}_{\mathbb{F}})$, an upper bound is obtained by noting that $\sigma_r(\mathcal{S}_{\mathbb{F}})$ is the projection of

$$\overline{\big\{\big((p_1, \ldots, p_r), p\big) \mid p = p_1 + \ldots + p_r,\ p_1, \ldots, p_r \in \mathcal{S}_{\mathbb{F}}\big\}} \subset (\mathcal{S}_{\mathbb{F}} \times \cdots \times \mathcal{S}_{\mathbb{F}}) \times \mathbb{F}^{\Pi}$$

[4] I was unable to generalize Theorem 2 to the complex case.

[5] This variety can be defined in projective space as well and is then called the r-secant variety of the Segre variety $\mathcal{S}_{\mathbb{C}} \subset \mathbb{P}^{\Pi-1}$.

onto the second factor, where the overline indicates the closure in the Euclidean topology. Therefore, $\dim \sigma_r(\mathcal{S}_\mathbb{F}) \leq \min\{\Pi, r(\Sigma+1)\}$, by [10, Proposition 2.8.6] for $\mathbb{F}=\mathbb{R}$ and [26, Theorem 11.12] for $\mathbb{F}=\mathbb{C}$. One expects that equality holds. The problem of determining the dimension of $\sigma_r(\mathcal{S}_\mathbb{F})$ has seen much progress recently, leading up to the conjecturally complete picture in [1].

A point $p \in \mathcal{V}$ on an irreducible complex variety $\mathcal{V}$ where there exists an open neighborhood $\mathcal{N}$ of p such that $\dim \mathrm{T}_q\mathcal{V} = \dim \mathcal{V}$ for all $q \in \mathcal{N}$ is called a *smooth point* of $\mathcal{V}$. The set of smooth points of $\mathcal{V}$ forms an n-dimensional submanifold of $\mathbb{C}^\Pi$; it is Zariski-dense in $\mathcal{V}$. For consistency with the real case, I will call a smooth point of a $\mathbb{C}$-variety a *regular point*. For semi-algebraic sets $\mathcal{V} \subset \mathbb{R}^\Pi$, defining the smooth locus is more subtle, because $\mathcal{V}$ can consist of several semi-algebraically connected components of different dimensions. Every semi-algebraic set $\mathcal{V}$ admits a *Nash stratification*, meaning that it can be decomposed as a finite union of disjoint semi-algebraically connected Nash manifolds [10, Section 9.1]: $\mathcal{V} = \bigcup_i \mathcal{M}_i$, where $\mathcal{M}_i$ is a smooth $\mathbb{R}$-manifold of dimension $d_i = \dim \mathcal{M}_i$ and $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ if $i \neq j$. Let $D = \{i \mid \dim \mathcal{M}_i = \dim \mathcal{V}\}$ be the set of all $\mathcal{M}_i$ of maximal dimension. I will call $\bigcup_{i\in D} \mathcal{M}_i$ the set of *regular points* of the semi-algebraic set $\mathcal{V}$; it forms a (potentially reducible)[6] $\mathbb{R}$-manifold. All regular points are smooth but in general the converse is false. The regular locus is neither Zariski-dense nor Euclidean-dense, in general.[7]

2.2. Why the standard approach fails

We can now understand why the Jacobian of f in (4) is not of full rank. Let $\mathfrak{p}_i = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d)$ be a representative of the rank-1 tensor $p_i = \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathcal{S}_\mathbb{F}$, and consider the decomposition

$$\mathfrak{A} = f(\mathbf{p}) = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d,$$

where $\mathbf{p} = \mathrm{vecr}(\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_r)$. Then, f's Jacobian matrix $T_\mathbf{p} := J_f(\mathbf{p})$ is given explicitly by

$$T_\mathbf{p} = \begin{bmatrix} T_1 & \cdots & T_r \end{bmatrix} \in \mathbb{F}^{\Pi \times M} \text{ with } T_i = \begin{bmatrix} I_{n_1} \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d & \cdots & \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^{d-1} \otimes I_{n_d} \end{bmatrix}, \tag{6}$$

where $M = r(\Sigma + d)$; the tensor products should be interpreted as Kronecker products in this expression. It is well known that the column span of $T_\mathbf{p}$ is contained in (one component of) the tangent space of $\sigma_r(\mathcal{S}_\mathbb{F})$ [1, 35, 58]. Hence, the rank of $T_\mathbf{p}$ is bounded from above by the dimension of $\sigma_r(\mathcal{S}_\mathbb{F})$, which is no more than $r(\Sigma+1)$. Terracini's Lemma [50] provides an even stronger statement, namely that $\mathrm{rank}(T_\mathbf{p}) = \dim \sigma_r(\mathcal{S}_\mathbb{F})$ provided that $\mathbf{p} \in \mathbb{F}^M$ is not contained in some unspecified Zariski-closed subset. For this reason, I refer to $T_\mathbf{p}$ as *Terracini's matrix* associated with the VFMs $\mathbf{p}$.

A *part* of the kernel of Terracini's matrix can be described explicitly. A basis of the kernel of T_i is

$$K_i = \left\{ \begin{bmatrix} \mathbf{a}_i^1 \\ -\mathbf{a}_i^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{a}_i^1 \\ 0 \\ -\mathbf{a}_i^3 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \mathbf{a}_i^1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -\mathbf{a}_i^d \end{bmatrix} \right\} = \{\mathbf{k}_i^2, \mathbf{k}_i^3, \ldots, \mathbf{k}_i^d\}, \tag{7a}$$

which contains exactly $d-1$ linearly independent basis vectors. One then verifies that

$$K = \{\mathbf{k}_i^2 \otimes \mathbf{e}_i, \mathbf{k}_i^3 \otimes \mathbf{e}_i, \ldots, \mathbf{k}_i^d \otimes \mathbf{e}_i\}_{i=1}^r, \tag{7b}$$

where $\mathbf{e}_i$ is the ith standard basis vector of $\mathbb{F}^r$, is a linearly independent set contained in the kernel of $T_\mathbf{p}$.

[6] In this paper, global connectedness of $\sigma_r(\mathcal{S}_\mathbb{F})$ is not of import as the definition of a condition number is inherently local. A neighborhood of a regular point of a variety or semi-algebraic set can always be chosen to be irreducible.

[7] The regular locus suffices for our purpose since the condition number in Theorem 10 will be ∞ for non-regular points.

Remark 4. *A reviewer asked when the rank of Terracini's matrix can be strictly less than $r(\Sigma+1)$. Characterizing such points is an interesting open problem. A partial answer was provided in [17, Lemma 37], which entails that if the VFMs $\mathbf{p}$ representing the tensor $\mathfrak{A} = f(\mathbf{p})$ are contained in a positive-dimensional family of alternative decompositions of $\mathfrak{A}$, then Terracini's matrix $T_{\mathbf{p}}$ is of rank strictly smaller than $r(\Sigma+1)$.*

As Terracini's matrix is of rank at most $N = r(\Sigma+1) < r(\Sigma+d) = M$, the classic theory of conditioning for implicit inverse functions does not apply. This paper nevertheless identifies three candidate strategies for deriving condition numbers in the setting of tensor rank decompositions. The first strategy yields a set of condition numbers for the TAP. The other two strategies result in condition numbers of the TDP.

Strategy I. Terracini's matrix is not of full rank because the domain of f is an over-parameterization for representing tensors of rank r. One could eliminate a subset S of the parameters (with $|S| = r(d-1)$) via an invertible smooth projection function $\pi_S : \mathbb{R}^M \to \mathbb{R}^N$ and consider instead an alternative tensor computation function $f_S = f \circ \pi_S^{-1}$. Then, for real tensors the standard theory, in particular Corollary 3, applies, resulting in a characterization of a condition number of the TAP $f_S^\dagger$. The main disadvantage of this approach is that these representations are usually not *global*, i.e., some rank-r tensors in the image of f are not in the image of f_S. Another disadvantage is that the elimination of parameters can be performed in many, ultimately arbitrary ways, resulting in a multitude of condition numbers all of which are equally valid. This approach is nevertheless explored in Section 3, mainly because its results are required in the proof of Theorem 1.

Strategy II. It is reasonable to conjecture that Corollary 3 should still be valid for functions of *constant rank*, namely the condition number of f^{-1} is expected to be $\kappa = \|(J_f(\mathbf{x}))^\dagger\| = \big(\varsigma_N(J_f(\mathbf{x}))\big)^{-1}$. A complication in generalizing the proof of Theorem 2 consists of ensuring that the boundary $\partial\mathcal{B}_\delta$ is mapped away from $\mathbf{y}_0$; otherwise the condition number is unbounded. It turns out that this can be accomplished by relaxing the requirement that distances between the parameters of two rank-r decompositions are measured by a metric. This approach results in a condition number of the TDP and is developed in detail in Section 4.

Strategy III. We could abandon the classic framework entirely and instead investigate the question of stability in the geometric framework of Blum, Cucker, Shub, and Smale [8], and Bürgisser and Cucker [13] by modeling the TDP as a map between manifolds. This presents its own challenges as the set of tensors of rank at most r is not naturally a manifold, but rather a singular semi-algebraic set if $\mathbb{F} = \mathbb{R}$ or a singular projective algebraic variety if $\mathbb{F} = \mathbb{C}$. It is nevertheless possible to extend this geometric framework to *join sets*, of which the set of tensors of rank at most r is an instance. The geometric condition number of the TDP obtained in this manner was recently introduced by Breiding and the author in [12].

3. Strategy I: Eliminating variables

A common strategy for eliminating the scaling indeterminacies in the representation of a rank-1 tensor consists of explicitly eliminating some variables via a suitable normalization. Because $\mathcal{S}_\mathbb{F}$ is a manifold of dimension $\Sigma + 1$, this idea essentially consists of covering $\mathcal{S}_\mathbb{F}$ by a collection of smooth coordinate charts $\{U_j \subset \mathcal{S}_\mathbb{F}\}_j$ such that $U_j = \phi_j^{-1}(\widehat{U}_j)$ with $\widehat{U}_j \subset \mathbb{F}^{\Sigma+1}$ open. This covering is not unique, and in general a rank-1 tensor $\mathfrak{A} \in \mathcal{S}_\mathbb{F}$ can even be represented on several charts, i.e., $\mathfrak{A} = \phi_i^{-1}(\mathbf{x}_i)$ for multiple i. The gist of this strategy consists of applying the standard theory of conditioning to the function $f \circ \phi_j^{-1}$. By construction, its Jacobian is expected to be of full rank so that Theorem 2 and Corollary 3 apply in the real case, yielding a condition number for the corresponding TAP.

Rank-1 tensors. A covering of $\mathcal{S}_\mathbb{F}$ is easily constructed by noting that a rank-1 tensor can be represented as

$$\mathfrak{p} = (\mathbf{x}^1, \dots, \mathbf{x}^d) \in \mathbb{F}_0^{n_1} \times \cdots \times \mathbb{F}_0^{n_d} \quad \text{with } \|\mathbf{x}^k\|_{\alpha_k} = c_k,\ k = 2, \dots, d.$$

where $\|\cdot\|_{\alpha_k}$ is a norm and $c_k > 0$ are arbitrary constants. The constraints $\|\mathbf{x}^k\|_{\alpha_k} = c_k$ enable the elimination of one variable. For example, an evident choice is taking $\alpha_k = 2$ for all k, so that a dense, open subset $U_{1,\ldots,1}$ of $\mathcal{S}_{\mathbb{F}}$ is the image of the smooth invertible function

$$\begin{aligned}\phi_{1,\ldots,1}^{-1} : \mathbb{F}_0^{n_1} \times \mathcal{B}_{c_2}(0) \times \cdots \times \mathcal{B}_{c_d}(0) &\to \mathcal{S}_{\mathbb{F}} \subset \mathbb{F}^{\Pi} \\ (\mathbf{w}^1, \ldots, \mathbf{w}^d) &\mapsto \mathbf{w}^1 \otimes \begin{bmatrix} \sqrt{c_2^2 - \|\mathbf{w}^2\|^2} \\ \mathbf{w}^2 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \sqrt{c_d^2 - \|\mathbf{w}^d\|^2} \\ \mathbf{w}^d \end{bmatrix},\end{aligned}$$

where $\mathcal{B}_{c_k}(0)$ is the open ball of radius $c_k > 0$ in $\mathbb{F}^{n_k - 1}$ centered at 0. $(U_{1,\ldots,1}, \phi_{1,\ldots,1})$ is a coordinate chart of $\mathcal{S}_{\mathbb{F}}$. Let $\mathbf{x}^k = [x_i^{(k)}]_{i=1}^{n_k}$, and then $\phi_{1,\ldots,1}$ corresponds to the elimination of the variables $x_1^{(k)}$, $k = 2, \ldots, d$. In an analogous manner, one can define the functions $\phi_{i_2,\ldots,i_d}^{-1}$ with $1 \le i_k \le n_k$ for $k = 2, \ldots, d$. Then, it is easy to check that $\{\{\mathbb{F}_0^{n_1} \times \mathcal{B}_{c_2}(0) \times \cdots \times \mathcal{B}_{c_d}(0), \phi_{i_2,\ldots,i_d}\}_{i_2,\ldots,i_d=1,\ldots,1}^{n_2,\ldots,n_d}\}$ is a smooth atlas for $\mathcal{S}_{\mathbb{F}}$. For every $(i_2, \ldots, i_d) \in [1, n_2] \times \cdots \times [1, n_d]$, Corollary 3 then yields the condition number of $(f \circ \phi_{i_2,\ldots,i_d}^{-1})_{\mathbf{y}}^{\dagger}$, i.e., the rank-1 approximation problem $\min_{\mathbf{x} \in \mathbb{R}^{\Sigma+1}} \|\mathfrak{A} - (f \circ \phi_{i_2,\ldots,i_d}^{-1})(\mathbf{x})\|^2$ associated with the function $f \circ \phi_{i_2,\ldots,i_d}^{-1}$ at the real, rank-1 tensor $\mathfrak{A} = f(\phi_{i_2,\ldots,i_d}^{-1}(\mathbf{y}))$.

The general case. The foregoing strategy can be generalized to tensors of rank equal to r if we restrict ourselves to the regular points $\mathcal{N}$ of $\sigma_r^0(\mathcal{S}_{\mathbb{F}})$; this ensures that Terracini's matrix is always of maximal rank $r(\Sigma + 1)$ for all of the VFMs representing a tensor $\mathfrak{A} \in \mathcal{N}$. Then, by definition, $\mathcal{N}$ is an $\mathbb{F}$-manifold so that we can apply the same approach as in the case of rank-1 tensors. Since Strategy I is not the focus of this paper, only the atlas of $\mathcal{N}$ that is required in the proof of the main theorem, namely Theorem 10, is presented here. It is essentially based on normalization in the ∞-norm. The construction proceeds as follows.

Choose $C \in (\mathbb{F}_0)^{(d-1)\times r}$, and let $\mathcal{C}$ be a set of $r(d-1)$ integers

$$\mathcal{C} := \{j_{k,i} \mid k = 2, \ldots, d,\ i = 1, \ldots, r\} \text{ with } j_{k,i} \in \{j \mid a_{j,i}^{(k)} \neq 0\} \text{ a fixed choice.} \tag{8}$$

Let $\mathbf{x} = \operatorname{vecr}(\mathfrak{x}_1, \ldots, \mathfrak{x}_r) \in \mathbb{F}^M$ be variables, where $\mathfrak{x}_i = (\mathbf{x}_i^1, \ldots, \mathbf{x}_i^d) \in \mathbb{F}_0^{n_1} \times \cdots \times \mathbb{F}_0^{n_d}$. We will represent tensors $\mathfrak{X} = f(\mathbf{x})$ in an open subset of $\mathcal{N}$ using the expected number of parameters N by choosing VFMs that lie in the subspace given by the $r(d-1)$ equations

$$\widehat{U}_{\mathcal{C},C} := \left\{ x_{j_{k,i},i}^{(k)} = c_{k,i} \ \mid \ j_{k,i} \in \mathcal{C} \right\} \subset \mathbb{F}^M,$$

where $x_{\ell,i}^{(k)}$ denotes the ℓth element of $\mathbf{x}_i^k$. Let $\mathfrak{Z} \in f(\widehat{U}_{\mathcal{C},C}) \cap \mathcal{N}$ be any choice. Then, there exists an open neighborhood $\mathcal{N}_{\mathfrak{Z}} \subset \mathcal{N}$ of $\mathfrak{Z}$ such that $\mathcal{N}_{\mathfrak{Z}}$ is contained in $f(\widehat{U}_{\mathcal{C},C})$. This follows from $\mathfrak{X} = f(\mathbf{x}) = f(\eta_{\mathcal{C},C}(\mathbf{x}))$ with normalization function $\eta_{\mathcal{C},C} : \mathbf{x} \mapsto \dot{\mathbf{x}} = \operatorname{vecr}(\dot{\mathfrak{x}}_1, \ldots, \dot{\mathfrak{x}}_r)$, where

$$\dot{\mathfrak{x}}_i = \left(\mathbf{x}_i^1 \cdot \prod_{k=2}^{d} \left(\frac{c_{k,i}}{x_{j_{k,i},i}^{(k)}} \right)^{-1}, \mathbf{x}_i^2 \cdot \frac{c_{2,i}}{x_{j_{2,i},i}^{(2)}}, \ldots, \mathbf{x}_i^d \cdot \frac{c_{d,i}}{x_{j_{d,i},i}^{(d)}} \right) \in \mathbb{F}_0^{n_1} \times \cdots \times \mathbb{F}_0^{n_d};$$

this normalization is well-defined in an open neighborhood of $\widehat{U}_{\mathcal{C},C}$ in $\mathbb{F}^M$, and the image $\eta_{\mathcal{C},C}$ is contained in $\widehat{U}_{\mathcal{C},C}$. The definition of $\eta_{\mathcal{C},C}$ is such that it simply chooses an alternative representative for each of the rank-1 tensors $\mathbf{x}_i^1 \otimes \cdots \otimes \mathbf{x}_i^d$. Define for $k = 2, 3, \ldots, d$ and $i = 1, 2, \ldots, r$, the functions

$$\begin{aligned}\pi_{j_{k,i}} &: \mathbb{F}^{n_k} \to \mathbb{F}^{n_k - 1},\ \mathbf{z} \mapsto \begin{bmatrix} z_1 & \cdots & z_{j_{k,i}-1} & z_{j_{k,i}+1} & \cdots & z_{n_k} \end{bmatrix}^T, \\ \pi_{j_{k,i},C}^{-1} &: \mathbb{F}^{n_k - 1} \to \mathbb{F}^{n_k},\ \mathbf{z} \mapsto \begin{bmatrix} z_1 & \cdots & z_{j_{k,i}-1} & c_{k,i} & z_{j_{k,i}} & \cdots & z_{n_k - 1} \end{bmatrix}^T,\end{aligned}$$

where $j_{k,i} \in \mathcal{C}$. Define also

$$\pi_{\mathcal{C}} : \mathbb{F}^M \to \mathbb{F}^N \qquad\qquad \pi_{\mathcal{C},C}^{-1} : \mathbb{F}^N \to \mathbb{F}^M$$

$$\begin{bmatrix} \mathbf{z}_i^1 \\ \mathbf{z}_i^2 \\ \vdots \\ \mathbf{z}_i^d \end{bmatrix}_{i=1,\dots,r} \mapsto \begin{bmatrix} \mathbf{z}_i^1 \\ \pi_{j_{2,i}}(\mathbf{z}_i^2) \\ \vdots \\ \pi_{j_{d,i}}(\mathbf{z}_i^d) \end{bmatrix}_{i=1,\dots,r}, \text{ and } \quad \begin{bmatrix} \mathbf{z}_i^1 \\ \mathbf{z}_i^2 \\ \vdots \\ \mathbf{z}_i^d \end{bmatrix}_{i=1,\dots,r} \mapsto \begin{bmatrix} \mathbf{z}_i^1 \\ \pi^{-1}_{j_{2,i},C}(\mathbf{z}_i^2) \\ \vdots \\ \pi^{-1}_{j_{d,i},C}(\mathbf{z}_i^d) \end{bmatrix}_{i=1,\dots,r}.$$

Then, we define the smooth "local tensor computation" function associated with $\mathcal{C}$ and C as

$$f_{\mathcal{C},C} : \mathbb{F}^N \to \mathbb{F}^\Pi,\ \mathbf{z} \mapsto (f \circ \pi^{-1}_{\mathcal{C},C})(\mathbf{z}), \tag{9}$$

where it is recalled that $N = r(\Sigma + 1)$. For example, in the simplest case E is the $(d-1) \times r$ matrix of ones and $\mathcal{C} = \{j_{k,i} = 1 \mid k = 2, \dots, d,\ i = 1, \dots, r\}$, so the corresponding local tensor computation function is

$$\begin{aligned} f_{\mathcal{C},E} : (\mathbb{F}_0^{n_1} \times \mathbb{F}^{n_2-1} \times \cdots \times \mathbb{F}^{n_d-1})^{\times r} &\to \mathbb{F}^\Pi \\ ((\mathbf{w}_i^1, \dots, \mathbf{w}_i^d))_{i=1}^r &\mapsto \sum_{i=1}^r \mathbf{w}_i^1 \otimes \begin{bmatrix} 1 \\ \mathbf{w}_i^2 \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} 1 \\ \mathbf{w}_i^d \end{bmatrix}. \end{aligned}$$

The "local" tensor rank approximation problem with input $\mathfrak{A} \in \mathbb{F}^\Pi$ associated with the foregoing model $f_{\mathcal{C},C}$ is $\min_{\mathbf{x} \in \mathbb{F}^N} \|\mathfrak{A} - f_{\mathcal{C},C}(\mathbf{x})\|^2$. As $f_{\mathcal{C},C}$ is analytic in N variables, it is continuously differentiable. Hence, the Jacobian matrix $\overline{J}_{\overline{\mathbf{p}}} \in \mathbb{F}^{\Pi \times N}$ of $f_{\mathcal{C},C}$ at $\overline{\mathbf{p}} = \pi_{\mathcal{C}}(\mathbf{p})$ with $\mathbf{p} := \eta_{\mathcal{C},C}(\mathbf{p}')$ is defined, and is given by

$$\overline{J}_{\overline{\mathbf{p}}} := J_{f \circ \pi^{-1}_{\mathcal{C},C}}(\overline{\mathbf{p}}) = J_f(\pi^{-1}_{\mathcal{C},C}(\overline{\mathbf{p}})) J_{\pi^{-1}_{\mathcal{C},C}}(\overline{\mathbf{p}}) = J_f(\mathbf{p}) \widehat{I}_{\mathcal{C}} = T_{\mathbf{p}} \widehat{I}_{\mathcal{C}}, \tag{10}$$

where the second equality is by the chain rule and $\widehat{I}_{\mathcal{C}} \in \mathbb{R}^{M \times N}$ is an identity matrix from which certain columns have been removed. Hence, the Jacobian $\overline{J}_{\overline{\mathbf{p}}}$ is formed by a subset of the columns of f's Jacobian, i.e., Terracini's matrix $T_{\mathbf{p}}$. Expressing $\mathbf{p} = \mathrm{vecr}(\mathfrak{p}_1, \dots, \mathfrak{p}_r)$ with $\mathfrak{p}_i = (\mathbf{a}_i^1, \dots, \mathbf{a}_i^d)$, we see that it follows from the definition of $\pi^{-1}_{\mathcal{C},C}$ that $\overline{J}_{\overline{\mathbf{p}}}$ lacks the $r(d-1)$ columns

$$\mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^{k-1} \otimes \mathbf{e}_{j_{k,i}} \otimes \mathbf{a}_i^{k+1} \otimes \cdots \otimes \mathbf{a}_i^d, \quad j_{k,i} \in \mathcal{C},$$

relative to $T_{\mathbf{p}}$, where $\mathbf{e}_i$ is the ith basis vector. One verifies that the columns of $\overline{J}_{\overline{\mathbf{p}}}$ and $T_{\mathbf{p}}$ still span the same space. As the latter has dimension N by the assumption that $\mathcal{N}_{\mathfrak{Z}} \subset \mathcal{N}$, it follows that $\overline{J}_{\overline{\mathbf{p}}} \in \mathbb{F}^{\Pi \times N}$ has linearly independent columns, provided that $\overline{\mathbf{p}} \in \mathbb{F}^N$ is not contained in some unspecified Zariski-closed subset. This leads to the next corollary in the real case.

Corollary 5. *Let $\mathfrak{A} = f_{\mathcal{C},C}(\overline{\mathbf{p}})$ be a real, rank-r tensor. If $\overline{J}_{\overline{\mathbf{p}}} \in \mathbb{R}^{\Pi \times N}$ has full rank, then the condition number of the TAP $(f_{\mathcal{C},C})^\dagger_{\overline{\mathbf{p}}} : \mathfrak{A} \mapsto \arg\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathfrak{A} - f_{\mathcal{C},C}(\mathbf{x})\|^2$ is $\kappa_{\mathcal{C},C}(\mathfrak{A}) = \|\overline{J}^\dagger_{\overline{\mathbf{p}}}\| = \big(\varsigma_N(\overline{J}_{\overline{\mathbf{p}}})\big)^{-1}$.*

The condition number of the TAP associated with the model $f_{\mathcal{C},C}$, obtained via the first strategy of elimination of variables, is *always worse than the corresponding condition number of the TDP in Theorem 1*, which is obtained via the second strategy of relaxing the distance measure. This is proved next.

Proposition 6. *Let $C \in (\mathbb{R}_0)^{(d-1) \times r}$ be arbitrary, and let $\mathcal{C}$ be any set as in (8). Then, the condition number $\kappa_{\mathcal{C},C}(\overline{\mathbf{p}})$ of the TAP $(f_{\mathcal{C},C})^\dagger_{\overline{\mathbf{p}}}$ at $\mathfrak{A} = f_{\mathcal{C},C}(\overline{\mathbf{p}})$ is larger than the condition number $\overline{\kappa}(\mathbf{p})$ with $\mathbf{p} = \pi^{-1}_{\mathcal{C},C}(\overline{\mathbf{p}})$ of the TDP as in Theorem 1:*

$$\kappa_{\mathcal{C},C}(\overline{\mathbf{p}}) := \big(\varsigma_N(\overline{J}_{\overline{\mathbf{p}}})\big)^{-1} \geq \big(\varsigma_N(T_{\mathbf{p}})\big)^{-1} =: \overline{\kappa}(\mathbf{p})$$

Proof. The inequality follows immediately from the min-max characterization of singular values, namely $\varsigma_N(\overline{J}_{\overline{\mathbf{p}}}) = \varsigma_N(\overline{J}^H_{\overline{\mathbf{p}}}) = \varsigma_N(\widehat{I}^H_{\mathcal{C}} T^H_{\mathbf{p}}) \leq \varsigma_1(\widehat{I}^H_{\mathcal{C}}) \varsigma_N(T^H_{\mathbf{p}}) = \varsigma_N(T_{\mathbf{p}})$, where all definitions are as in (10). □

Aside from the very arbitrary nature of the variable elimination, I believe that this proposition provides a good incentive for considering Strategy II, which is developed in the next section.

4. Strategy II: Relaxing the distance measure

The second strategy is derived from the definition of the structured condition number in (2). There is one major complication, namely the TDP is not naturally modeled as a *function*. Indeed, let

$$f^{-1}(\mathfrak{A}) := \{\mathbf{x} \in \mathbb{F}^M \mid f(\mathbf{x}) = \mathfrak{A}\}$$

be the fibre of f at $\mathfrak{A}$. The domain of f^{-1} is $\sigma_r^0(\mathcal{S}_\mathbb{F})$. The fibre $f^{-1}(\mathfrak{A})$ is an affine variety consisting of representatives of the tensor $\mathfrak{A}$; it is well known that $\dim f^{-1}(\mathfrak{A}) \geq r(d-1)$. The next subsection explains how this complication can be overcome via robust r-identifiability and a quotient.

4.1. Geometry of the tensor rank decomposition II: Robust identifiability

The representatives $\mathbf{x} = \operatorname{vecr}(\mathfrak{p}_1, \ldots, \mathfrak{p}_r) \in \mathbb{F}^M$ and $\mathbf{y} = \operatorname{vecr}(\mathfrak{q}_1, \ldots, \mathfrak{q}_r) \in \mathbb{F}^M$ of two rank decompositions are *essentially equal*, denoted $\mathbf{x} \sim \mathbf{y}$, if the multisets of rank-1 tensors $\{\operatorname{Seg}(\mathfrak{p}_1), \ldots, \operatorname{Seg}(\mathfrak{p}_r)\}$ and $\{\operatorname{Seg}(\mathfrak{q}_1), \ldots, \operatorname{Seg}(\mathfrak{q}_r)\}$ are equal including multiplicities.

Proposition 7. *$\sim$ is an equivalence relationship.*

Proof. See Appendix A. □

Let $[\mathbf{p}] := \{\mathbf{q} \in \mathbb{F}^M \mid \mathbf{q} \sim \mathbf{p}\}$ denote the equivalence class of $\mathbf{p} \in \mathbb{F}^M$, and let $q : \mathbb{F}^M \to \mathbb{F}^M/\!\sim$ be the quotient map. By the proposition, the composition $(q \circ f^{-1})(\mathfrak{A}) := \{[\mathbf{p}] \mid f(\mathbf{p}) = \mathfrak{A}\}$ is well defined: it sends a rank-r tensor to its set of equivalence classes of representatives. Let the set of r-identifiable tensors be $\mathcal{I} := \{\mathfrak{A} \mid |(q \circ f^{-1})(\mathfrak{A})| = 1\} \subset \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} \cong \mathbb{F}^\Pi$, where $|S|$ denotes the cardinality of the set S which for infinite sets we define as $|S| = \infty$. It follows by definition that $g = (q \circ f^{-1}|_\mathcal{I}) : \mathcal{I} \to (\mathbb{F}^M/\!\sim)$ is a function. If we additionally obtain some differentiability guarantees for g, then we could hope to apply the standard definition of a condition number in (2). Therefore, the notion of *robust r-identifiability* is introduced next.[8]

Definition 1. *A rank-r tensor $\mathfrak{A} \in \sigma_r^0(\mathcal{S}_\mathbb{F})$ is* robustly r-identifiable *if it is a regular point of $\sigma_r(\mathcal{S}_\mathbb{F})$ and there exists an open neighborhood $\mathcal{N}$ of $\mathfrak{A}$ such that every $\mathfrak{B} \in \mathcal{N} \subset \sigma_r^0(\mathcal{S}_\mathbb{F})$ is r-identifiable.*

A robustly r-identifiable tensor $\mathfrak{A} = f(\mathbf{p})$ thus has an open neighborhood $\mathcal{N} \subset \sigma_r^0(\mathcal{S}_\mathbb{F})$ such that for every $\mathfrak{B} \in \mathcal{N}$ it holds that $\mathfrak{B}$ is r-identifiable and a regular point of $\sigma_r(\mathcal{S}_\mathbb{F})$. Hence, $\mathcal{N}$ is a smooth $\mathbb{F}$-manifold. It follows that the column span of Terracini's matrix $T_\mathbf{p}$ coincides with the tangent space to $\sigma_r(\mathcal{S}_\mathbb{F})$.

Robust r-identifiability is expected to impose only a mild condition on rank-r tensors. A Segre variety $\mathcal{S}_\mathbb{F}$ is called *generically r-identifiable* if there exists a Zariski-closed set $Z \subset \sigma_r(\mathcal{S}_\mathbb{F})$ such that every $\mathfrak{A} \in \sigma_r(\mathcal{S}_\mathbb{F}) \backslash Z$ is r-identifiable. Then, the following basic result applies.

Lemma 8. *Let $\mathcal{S}_\mathbb{F}$ be a generically r-identifiable Segre variety, and let Z be the Zariski-closed set that is the union of the locus of singular points of $\sigma_r(\mathcal{S}_\mathbb{F})$ and the locus where r-identifiability fails. Then, every element of $\sigma_r^0(\mathcal{S}_\mathbb{F}) \setminus Z$ is robustly r-identifiable.*

In the literature, many Segre varieties $\mathcal{S}_\mathbb{F}$ are proven to be generically r-identifiable [9, 15, 23, 49]. The conjecture from [16, 17, 43] paints a nearly complete picture.

Conjecture 1 ([16, 17, 43])**.** *Let $d \geq 3$, let $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$, and let $\mathcal{S}_\mathbb{F} = \operatorname{Seg}(\mathbb{F}_0^{n_1} \times \cdots \times \mathbb{F}_0^{n_d}) \subset \mathbb{F}^\Pi$ be a Segre variety with $n_1 \geq \cdots \geq n_d \geq 2$. Then, $\mathcal{S}_\mathbb{F}$ is generically r-identifiable if $r < \frac{\Pi}{\Sigma+1}$, unless $(n_1, \ldots, n_d)$ is one of the exceptional cases, namely $(4,4,3)$; $(4,4,4)$; $(6,6,3)$; $(2,2,2,2,2)$; $(n,n,2,2)$ with $n \in \mathbb{N}$; or $n_1 > \prod_{k=2}^d n_k - \sum_{k=2}^d (n_k - 1)$. In the exceptional cases there are multiple* complex *rank-r decompositions.*

The computer-assisted proof of [16] combined with the observation of [17, 43] shows that this conjecture is true for all $n_1 \cdots n_d \leq 15000$. Note that if $r > \frac{\Pi}{\Sigma+1}$, then $\mathcal{S}_\mathbb{F}$ is never generically r-identifiable.

[8] A similar notion that bears the same name but imposes no smoothness condition was recently examined in [7].

4.2. Relaxing the metric

We can now apply the standard definition of the condition number in (2) to the function

$$f_{\mathcal{R}}^{-1} := (q \circ f^{-1}|_{\mathcal{R}}) : \mathcal{R} \to (\mathbb{F}^M/\sim),\ f(\mathbf{x}) \mapsto [\mathbf{x}]$$

where $\mathcal{R}$ is the subset of robustly r-identifiable tensors in $\mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} \cong \mathbb{F}^{\Pi}$ and q the quotient map from the previous subsection. This function takes a robustly r-identifiable tensor $\mathfrak{A} = f(\mathbf{x})$ to the affine variety of essentially equal representatives $[\mathbf{x}] = \{\mathbf{y} \mid \mathbf{x} \sim \mathbf{y}\} \subset \mathbb{F}^M$ where $\mathbf{x}$ is any vector satisfying $f(\mathbf{x}) = \mathfrak{A}$.

We still need to choose how to measure errors between equivalence classes of representatives of rank-r decompositions in $\mathbb{F}^M/\sim$ as well as between tensors in $\mathbb{F}^{\Pi}$. The natural choice for the latter is the standard 2-norm on $\mathbb{F}^{\Pi}$. An obvious choice for measuring the error between two equivalence classes of representatives $[\mathbf{x}], [\mathbf{y}] \in \mathbb{F}^M/\sim$ would be the minimum distance between these varieties (considered as subvarieties of $\mathbb{F}^M$): $\mu([\mathbf{x}], [\mathbf{y}]) := \inf_{\mathbf{x}' \in [\mathbf{x}], \mathbf{y}' \in [\mathbf{y}]} \|\mathbf{x}' - \mathbf{y}'\|$. A condition number of $f_{\mathcal{R}}^{-1}$ with respect to μ at the tensor $\mathfrak{A}$ would measure the worst possible fraction

$$\frac{\mu\big(f_{\mathcal{R}}^{-1}(\mathfrak{A}), f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A})\big)}{\|\Delta\mathfrak{A}\|}$$

for infinitesimal $\Delta\mathfrak{A}$ with $\mathfrak{A} + \Delta\mathfrak{A} \in \mathcal{R}$. That is, it looks at the distance between the best representative $\mathbf{p} \in [\mathbf{x}]$ of $\mathfrak{A} = f(\mathbf{x})$ and $\mathbf{q} \in [\mathbf{y}]$ of $\mathfrak{A}+\Delta\mathfrak{A} = f(\mathbf{y})$ relative to $\|\Delta\mathfrak{A}\|$. Unfortunately, μ *does not separate equivalence classes* as $\mu([\mathbf{x}], [\mathbf{y}])$ can be 0 even if $[\mathbf{x}] \neq [\mathbf{y}]$.[9] The basic example is $\mathbf{x} = \mathrm{vec}(\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^d)$ and $\mathbf{y} = \mathrm{vec}(\mathbf{b}^1, \mathbf{b}^2, \mathbf{a}^3, \ldots, \mathbf{a}^d)$. Let $\mathbf{x}_\epsilon = \mathrm{vec}(\epsilon\mathbf{a}^1, \epsilon\mathbf{a}^2, \epsilon^{-2}\mathbf{a}^3, \ldots, \mathbf{a}^d) \in [\mathbf{x}]$ and $\mathbf{y}_\epsilon = \mathrm{vec}(\epsilon\mathbf{b}^1, \epsilon\mathbf{b}^2, \epsilon^{-2}\mathbf{a}^3, \ldots, \mathbf{a}^d) \in [\mathbf{y}]$, and then $0 \leq \mu([\mathbf{x}], [\mathbf{y}]) \leq \|\mathbf{x}_\epsilon - \mathbf{y}_\epsilon\| = \mathcal{O}(\epsilon)$ for all $\epsilon > 0$, hence $\mu([\mathbf{x}], [\mathbf{y}]) = 0$. So if $\mathbf{a}^1 \neq \mathbf{b}^1$ and $\mathbf{a}^2 \neq \mathbf{b}^2$ then $[\mathbf{x}] \neq [\mathbf{y}]$ even though their distance is zero.

The above problem motivates the following alternative way to measure errors between representatives of rank-r tensors. Let $\mathfrak{A} = f(\mathbf{p})$ and $\mathfrak{B} = f(\mathbf{q})$ be two robustly r-identifiable tensors. We can measure the distance between a given representative $\mathbf{p} \in \mathbb{F}^M$ of $\mathfrak{A}$ and the equivalence class of representatives of the rank-r decomposition of $\mathfrak{B}$, i.e., $[\mathbf{q}] = f_{\mathcal{R}}^{-1}(\mathfrak{B})$, as the least distance between $\mathbf{p}$ and the variety $[\mathbf{q}]$:

$$\mathrm{dist}(\mathbf{p}, [\mathbf{q}]) = \inf_{\mathbf{q}' \in [\mathbf{q}]} \|\mathbf{p} - \mathbf{q}'\|. \tag{11}$$

The condition number of $f_{\mathcal{R}}^{-1}$ with respect to dist *at the representative* $\mathbf{p}$ *of* $\mathfrak{A} = f(\mathbf{p})$ measures the worst possible fraction

$$\frac{\mathrm{dist}\big(\mathbf{p}, f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A})\big)}{\|\Delta\mathfrak{A}\|}$$

for infinitesimal $\Delta\mathfrak{A}$ with $\mathfrak{A} + \Delta\mathfrak{A} \in \mathcal{R}$. That is, it looks at the distance between *the given representative* $\mathbf{p} \in \mathbb{F}^M$ of $\mathfrak{A} = f(\mathbf{p})$ and the best representative $\mathbf{q} \in [\mathbf{y}]$ of $\mathfrak{A} + \Delta\mathfrak{A} = f(\mathbf{y})$. In the case where we are agnostic about the perturbed decomposition $f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A})$, there is no reason to prefer some VFMs over others. In this case it is reasonable to measure the error with respect to the best possible representative $\mathbf{q} \in f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A})$ of the perturbed tensor, which is exactly what dist accomplishes.

The function dist does correctly separate equivalence classes of representatives, as is shown next.

Proposition 9. *Let* $\mathbf{x} \in \mathbb{F}^M$ *and* $[\mathbf{y}] \in \mathbb{F}^M/\sim$. *Then, the infimum is attained:*

$$\mathrm{dist}(\mathbf{x}, [\mathbf{y}]) = \min_{\mathbf{y}' \in [\mathbf{y}]} \|\mathbf{x} - \mathbf{y}'\|.$$

Moreover, the distance $\mathrm{dist}(\mathbf{x}, [\mathbf{y}]) = \mathrm{dist}(\mathbf{y}, [\mathbf{x}]) = 0$ *if and only if* $\mathbf{x} \sim \mathbf{y}$.

Proof. See Appendix A. □

In conclusion, dist : $\mathbb{F}^M \times (\mathbb{F}^M/\sim)$ satisfies the properties $\mathrm{dist}(\mathbf{x}, [\mathbf{y}]) \geq 0$, $\mathrm{dist}(\mathbf{x}, [\mathbf{x}]) = 0$, and $\mathrm{dist}(\mathbf{x}, [\mathbf{y}]) = 0$ iff $\mathbf{x} \sim \mathbf{y}$. These conditions suffice for proving the main theorem of the next subsection.

[9] μ is a symmetric premetric satisfying $\mu([\mathbf{x}], [\mathbf{y}]) \geq 0$, $\mu([\mathbf{x}], [\mathbf{x}]) = 0$, and $\mu([\mathbf{x}], [\mathbf{y}]) = \mu([\mathbf{y}], [\mathbf{x}])$.

4.3. A condition number

The main result of this paper states that the inverse of the Nth singular value of Terracini's matrix is the absolute condition number of the TDP $f_{\mathcal{R}}^{-1}$ with respect to the distance measure dist in (11).

Theorem 10. *Let $\mathfrak{p}_i = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d)$ be a representative of $p_i = \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathcal{S}_{\mathbb{F}}$, $i = 1, \ldots, r$, and let $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_r)$ be the given VFMs. Assume that*

$$\mathfrak{A} = f(\mathbf{p}) = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \quad \in \sigma_r^0(\mathcal{S}_{\mathbb{F}}) \subset \mathbb{F}^{\Pi}$$

is robustly r-identifiable, and let ς_N with $N = r(\Sigma + 1)$ denote the Nth largest singular value of Terracini's matrix $T_{\mathbf{p}}$. Then, the absolute condition number *of the tensor rank decomposition problem $f_{\mathcal{R}}^{-1}$ at $\mathbf{p}$ satisfies*

$$\overline{\kappa}_A(\mathbf{p}) := \lim_{\epsilon \to 0} \max_{\substack{\|\Delta\mathfrak{A}\| \leq \epsilon, \\ \mathfrak{A}+\Delta\mathfrak{A} \in \mathcal{R}}} \frac{\operatorname{dist}(\mathbf{p}, f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A}))}{\|\Delta\mathfrak{A}\|} = \varsigma_N^{-1},$$

where $\mathcal{R} \subset \sigma_r(\mathcal{S}_{\mathbb{F}})$ is the locus of robustly r-identifiable tensors. If $\mathfrak{A} \notin \mathcal{R}$, then $\overline{\kappa}_A$ is undefined.

If $\overline{\kappa}_A$ is finite, then the relative condition number *at $\mathbf{p}$ is $\overline{\kappa}(\mathbf{p}) := \varsigma_N^{-1} \|\mathfrak{A}\| \|\mathbf{p}\|^{-1}$.*

Proof. The proof is presented in Appendix C. □

Let $\mathfrak{A} = f(\mathbf{p})$ and $\widehat{\mathfrak{A}} = f(\widehat{\mathbf{p}})$ be robustly r-identifiable tensors. As usual, we have the following asymptotically sharp inequalities for estimating the absolute, respectively relative, forward error between VFMs:

$$\operatorname{dist}(\widehat{\mathbf{p}}, f_{\mathcal{R}}^{-1}(\mathfrak{A})) \lesssim \overline{\kappa}_A(\widehat{\mathbf{p}}) \cdot \|\mathfrak{A} - \widehat{\mathfrak{A}}\| \quad \text{and} \quad \frac{\operatorname{dist}(\widehat{\mathbf{p}}, f_{\mathcal{R}}^{-1}(\mathfrak{A}))}{\|\widehat{\mathbf{p}}\|} \lesssim \overline{\kappa}(\widehat{\mathbf{p}}) \frac{\|\mathfrak{A} - \widehat{\mathfrak{A}}\|}{\|\widehat{\mathfrak{A}}\|},$$

provided that $\|\mathfrak{A} - \widehat{\mathfrak{A}}\|$ is sufficiently small.

5. The condition number of the TDP

The absolute and relative condition numbers considered in the previous subsection are defined at the representative $\mathbf{p} \in \mathbb{F}^M$ rather than at the tensor $\mathfrak{A} \in \mathbb{F}^{\Pi}$. These condition numbers depend on the particular $\mathbf{p} \in f_{\mathcal{R}}^{-1}(\mathfrak{A})$ that was chosen. If there is no choice of VFMs that is suggested by the application at hand, then where should the condition number from Theorem 10 be evaluated?

Consider the rank-1 tensor $p_i = \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d \in \mathcal{S}_{\mathbb{F}}$. All of the representatives of p_i can be parameterized as $\mathbf{p}_{\boldsymbol{\alpha}}^i := \operatorname{vec}(\alpha_{i,1}\mathbf{u}_i^1, \ldots, \alpha_{i,d}\mathbf{u}_i^d)$ where $\|\mathbf{u}_i^k\| = 1$, $\alpha_{i,1} \cdots \alpha_{i,d} = 1$, and $\alpha_{i,k} > 0$. Terracini's matrix is

$$\begin{aligned} T_{\mathbf{p}_{\boldsymbol{\alpha}}^i} &= \begin{bmatrix} (\alpha_{i,2} \cdots \alpha_{i,d}) I_{n_1} \otimes \mathbf{u}_i^2 \otimes \cdots \otimes \mathbf{u}_i^d & \cdots & (\alpha_{i,1} \cdots \alpha_{i,d-1}) \mathbf{u}_i^1 \otimes \cdots \otimes \mathbf{u}_i^{d-1} \otimes I_{n_d} \end{bmatrix} \\ &= \begin{bmatrix} I_{n_1} \otimes \mathbf{u}_i^2 \otimes \cdots \otimes \mathbf{u}_i^d & \cdots & \mathbf{u}_i^1 \otimes \cdots \otimes \mathbf{u}_i^{d-1} \otimes I_{n_d} \end{bmatrix} \cdot \operatorname{diag}((\alpha_{i,2} \cdots \alpha_{i,d}) I_{n_1}, \ldots, (\alpha_{i,1} \cdots \alpha_{i,d-1}) I_{n_d}) \\ &=: \widehat{T}_{\mathbf{p}_{\boldsymbol{\alpha}}^i} \cdot D_{\boldsymbol{\alpha}}^i. \end{aligned}$$

From the min-max characterization of singular values it follows that $\varsigma_{\Sigma+1}(T_{\mathbf{p}_{\boldsymbol{\alpha}}^i}) \leq \varsigma_1(\widehat{T}_{\mathbf{p}_{\boldsymbol{\alpha}}^i}) \cdot \varsigma_{\Sigma+1}(D_{\boldsymbol{\alpha}}^i)$; this is essentially [30, Ex. 18 in Section 7.3]. From Corollary 18 below, we have $\varsigma_1(\widehat{T}_{\mathbf{p}_{\boldsymbol{\alpha}}^i}) = \sqrt{d}$. If we assume that all $n_k \geq d$, then it follows from the structure of $D_{\boldsymbol{\alpha}}^i$ that $\varsigma_{\Sigma+1}(D_{\boldsymbol{\alpha}}^i) = \min_{1 \leq k \leq d} \prod_{1 \leq j \neq k \leq d} \alpha_{i,j}$. Hence,

$$\overline{\kappa}_A(\mathbf{p}_{\boldsymbol{\alpha}}^i) \geq \frac{1}{\sqrt{d}} \max_{1 \leq k \leq d} \prod_{1 \leq j \neq k \leq d} \alpha_{i,j}^{-1}.$$

One verifies that this lower bound is minimized for $\alpha_{i,1} = \cdots = \alpha_{i,d} = \|p_i\|^{1/d}$. On the other hand, it is easy to choose representatives $\mathbf{p}_{\boldsymbol{\alpha}}^i$ of p_i that are ill-conditioned, for example by choosing $w \gg 1$ and setting $\alpha_{i,1} = 2^w$, $\alpha_{i,2} = \cdots = \alpha_{i,d} = 2^{-w/(d-1)}$, the condition number becomes $\overline{\kappa}_A(\mathbf{p}_{\boldsymbol{\alpha}}^i) \geq d^{-1/2} 2^w$.

Continuing from the above observations, we can also bound the condition number of $f_{\mathcal{R}}^{-1}$ at the VFMs $\mathbf{p}_{\boldsymbol{\alpha}}^T = \begin{bmatrix}(\mathbf{p}_{\boldsymbol{\alpha}}^1)^T & \cdots & (\mathbf{p}_{\boldsymbol{\alpha}}^r)^T\end{bmatrix}$, which represent the tensor $p = p_1 + \cdots + p_r$. Terracini's matrix is

$$T_{\mathbf{p}_{\boldsymbol{\alpha}}} = \begin{bmatrix}T_{\mathbf{p}_{\boldsymbol{\alpha}}^1} & \cdots & T_{\mathbf{p}_{\boldsymbol{\alpha}}^r}\end{bmatrix} = \begin{bmatrix}\widehat{T}_{\mathbf{p}_{\boldsymbol{\alpha}}^1} & \cdots & \widehat{T}_{\mathbf{p}_{\boldsymbol{\alpha}}^r}\end{bmatrix} \cdot \operatorname{diag}(D_{\boldsymbol{\alpha}}^1, \ldots, D_{\boldsymbol{\alpha}}^r).$$

Assume that $T_{\mathbf{p}_{\boldsymbol{\alpha}}}$ is of the maximal rank N. By the min-max characterization of singular values

$$\varsigma_N(T_{\mathbf{p}_{\boldsymbol{\alpha}}}) = \min_{\substack{\mathbf{v}\perp\ker(T_{\mathbf{p}_{\boldsymbol{\alpha}}}),\\ \|\mathbf{v}\|=1}} \|T_{\mathbf{p}_{\boldsymbol{\alpha}}}\mathbf{v}\| \leq \min_{\substack{\mathbf{v}\perp\ker(T_{\mathbf{p}_{\boldsymbol{\alpha}}}),\\ \|\mathbf{v}\|=1,\\ \mathbf{v}^T=\begin{bmatrix}\mathbf{v}_1^T & \cdots & \mathbf{v}_r^T\end{bmatrix}}} \Big(\|T_{\mathbf{p}_{\boldsymbol{\alpha}}^1}\mathbf{v}_1\| + \cdots + \|T_{\mathbf{p}_{\boldsymbol{\alpha}}^r}\mathbf{v}_r\|\Big),$$

where the inequality is due to the triangle inequality, and $\ker(A)$ denotes the kernel of $A \in \mathbb{F}^{m\times n}$. The assumption on the rank of $T_{\mathbf{p}_{\boldsymbol{\alpha}}}$ causes $\mathbf{v} \perp \ker(T_{\mathbf{p}_{\boldsymbol{\alpha}}})$ to entail $\mathbf{v}_i \perp \ker(T_{\mathbf{p}_{\boldsymbol{\alpha}}^i})$ for all $i = 1, \ldots, r$. Hence, choosing $\mathbf{v}_1 = \cdots = \mathbf{v}_{i-1} = \mathbf{v}_{i+1} = \cdots = \mathbf{v}_d = 0$, $\|\mathbf{v}_i\| = 1$, and $\mathbf{v}_i \perp \ker(T_{\mathbf{p}_{\boldsymbol{\alpha}}^i})$, the right hand side of the above inequality is bounded by $\|T_{\mathbf{p}_{\boldsymbol{\alpha}}^i}\mathbf{v}_i\|$. Since this bound is valid for every $i = 1, \ldots, d$, we find

$$\varsigma_N(T_{\mathbf{p}_{\boldsymbol{\alpha}}}) \leq \min_{1\leq i\leq r} \min_{\substack{\mathbf{v}_i\perp\ker(T_{\mathbf{p}_{\boldsymbol{\alpha}}^i}),\\ \|\mathbf{v}_i\|=1}} \|T_{\mathbf{p}_{\boldsymbol{\alpha}}^i}\mathbf{v}_j\| = \min_{1\leq i\leq r} \varsigma_{\Sigma+1}(T_{\mathbf{p}_{\boldsymbol{\alpha}}^i}).$$

The condition number of $f_{\mathcal{R}}^{-1}$ at the VFMs $\mathbf{p}_{\boldsymbol{\alpha}}$ thus satisfies

$$\overline{\kappa}_A(\mathbf{p}_{\boldsymbol{\alpha}}) \geq \frac{1}{\sqrt{d}} \cdot \max_{1\leq i\leq r} \max_{1\leq k\leq d} \prod_{1\leq j\neq k\leq d} \alpha_{i,j}^{-1}.$$

As before, taking $\alpha_{i,1} = \cdots = \alpha_{i,d} = \|p_i\|^{1/d}$ is a minimizer of the expression on the right hand side.

The above evidence strongly suggests to evaluate the condition number of $f_{\mathcal{R}}^{-1}$ at *norm-balanced* VFMs.

Definition 2. *The VFMs* $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \ldots, \mathfrak{p}_r)$ *with* $\mathfrak{p}_i = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d)$ *are called* norm-balanced *if for every* $i = 1, \ldots, r$ *we have* $\|\mathbf{a}_i^1\| = \cdots = \|\mathbf{a}_i^d\|$.

We can then define the condition number of the TDP $f_{\mathcal{R}}^{-1}$ at *the tensor* $\mathfrak{A}$ as follows.

Definition 3 (Condition number of the TDP)**.** *Let* $\mathfrak{A} \in \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} \cong \mathbb{F}^{\Pi}$ *be a robustly* r*-identifiable rank decomposition. Let* $\mathbf{p} \in f_{\mathcal{R}}^{-1}(\mathfrak{A})$ *be norm-balanced. Then, the condition number of the TDP* $f_{\mathcal{R}}^{-1}$ *at* $\mathfrak{A}$ *is defined to be the condition number of* $f_{\mathcal{R}}^{-1}$ *at a norm-balanced representative of* $\mathfrak{A}$*:*

$$\kappa_A(\mathfrak{A}) := \overline{\kappa}_A(\mathbf{p}) \quad \textit{and} \quad \kappa(\mathfrak{A}) := \overline{\kappa}(\mathbf{p}).$$

Remark 11. *The fibre* $f_{\mathcal{R}}^{-1}$ *contains several norm-balanced VFMs. Indeed, if* $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \ldots, \mathfrak{p}_r) \in f_{\mathcal{R}}^{-1}(\mathfrak{A})$*, then for every permutation* π *of* $\{1, \ldots, r\}$ *we have* $\mathbf{p}_\pi := \operatorname{vecr}(\mathfrak{p}_{\pi_1}, \ldots, \mathfrak{p}_{\pi_r}) \in f_{\mathcal{R}}^{-1}(\mathfrak{A})$ *as well. It is nevertheless easy to verify that the condition numbers of norm-balanced VFMs of* $\mathfrak{A}$ *are all equal.*

The condition number $\kappa(\mathfrak{A})$ is computed straightforwardly if the VFMs $\mathbf{p} \in \mathbb{F}^M$ representing the rank-r decomposition of $\mathfrak{A}$ are given, requiring only the Nth largest singular value of Terracini's matrix corresponding to the norm-balanced VFMs $\mathbf{p}' \in [\mathbf{p}]$.

Remark 12. *A reviewer observed that choosing norm-balanced VFMs in Strategy II, as in Definition 3, can also be interpreted as a specific approach for variable elimination that could be employed in Strategy I. The reviewer asked if there is any obvious relationship between these two approaches. Note that in this approach one would still have to choose which of the variables are eliminated, resulting in a set of condition numbers as before. It seems unlikely that all of these condition numbers would coincide, as this is also not the case for the approach based on normalization in the* ∞*-norm that was described in Section 3. The precise relationship is left as an open problem.*

5.1. Elementary properties

We start by proving three desirable properties of the relative condition number κ that are also exhibited by the condition number of a matrix, namely its continuity (under mild conditions), and its invariance under scaling and orthogonal change of bases.

Proposition 13. *The absolute and relative condition numbers $\kappa_A(\mathfrak{A})$ and $\kappa(\mathfrak{A})$ are continuous in a neighborhood of a robustly r-identifiable tensor $\mathfrak{A}$.*

Proof. This follows immediately from the continuity of Terracini's matrix $T_{\mathbf{p}}$ in the parameters $\mathbf{p}$, which is a consequence from the assumption of regularity in the definition of robust r-identifiability. □

Proposition 14. *The relative condition number κ is scale-invariant: $\kappa(\mathfrak{A}) = \kappa(\beta\mathfrak{A})$ for all $\beta \in \mathbb{F}_0$.*

Proof. Let $\mathbf{p} \in f_{\mathcal{R}}^{-1}(\mathfrak{A})$ be norm-balanced. For $\mathbb{F} = \mathbb{C}$, we can write $\beta\mathfrak{A} = \alpha^d\mathfrak{A} = f(\alpha\mathbf{p})$. Terracini's matrix corresponding to the norm-balanced VFMs $\mathbf{q} = \alpha\mathbf{p}$ is $T_{\mathbf{q}} = \alpha^{d-1}T_{\mathbf{p}}$. Hence, $(\varsigma_N(T_{\mathbf{q}}))^{-1} = (|\alpha|^{d-1}\varsigma_N(T_{\mathbf{p}}))^{-1}$. Since $\|\mathfrak{B}\|/\|\mathbf{q}\| = \|\alpha^d\mathfrak{A}\|/\|\alpha\mathbf{p}\| = |\alpha|^{d-1}\|\mathfrak{A}\|/\|\mathbf{p}\|$, we get

$$\kappa(\mathfrak{B}) = \overline{\kappa}_A(\mathfrak{B})\frac{\|\mathfrak{B}\|}{\|\mathbf{q}\|} = (\varsigma_N(T_{\mathbf{q}}))^{-1}\frac{\|\mathfrak{B}\|}{\|\mathbf{q}\|} = (|\alpha|^{d-1}\varsigma_N(T_{\mathbf{p}}))^{-1}|\alpha|^{d-1}\frac{\|\mathfrak{A}\|}{\|\mathbf{p}\|} = \kappa(\mathfrak{A}).$$

In the real case, whenever d is even and $\beta < 0$, one should exploit $\beta\mathfrak{A} = -\alpha^d\mathfrak{A} = f(\alpha\mathbf{p}')$, where $\mathbf{p}' = \operatorname{vecr}(\mathfrak{p}'_1, \ldots, \mathfrak{p}'_r)$ with $\mathfrak{p}'_i = (-\mathbf{a}_i^1, \mathbf{a}_i^2, \ldots, \mathbf{a}_i^d)$. Then, Terracini's matrix corresponding to $\mathbf{q} = \alpha\mathbf{p}'$ is $T_{\mathbf{q}} = \alpha^{d-1}T_{\mathbf{p}}S$, where S is a diagonal matrix whose diagonal entries are ± 1. Since S is an orthogonal matrix, $\varsigma_N(T_{\mathbf{p}}) = \varsigma_N(T_{\mathbf{p}}S)$. □

Proposition 15. *The absolute and relative condition numbers κ_A and κ are orthogonally invariant.*

Proof. Considering $\mathfrak{A} \in \mathbb{F}^{\Pi}$, the claim is that $\kappa(\mathfrak{A}) = \kappa((Q_1 \otimes \cdots \otimes Q_d)\mathfrak{A})$, where $Q_i \in O(n_i)$ are orthogonal matrices in $\mathbb{F}^{n_i \times n_i}$ with respect to the Hermitian inner product. As $Q = Q_1 \otimes \cdots \otimes Q_d$ is orthogonal, $\|\mathfrak{A}\| = \|Q\mathfrak{A}\|$. Since

$$Q\mathfrak{A} = (Q_1 \otimes \cdots \otimes Q_d)\sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d = \sum_{i=1}^{r} Q_1\mathbf{a}_i^1 \otimes \cdots \otimes Q_d\mathbf{a}_i^d,$$

it follows that $Q\mathfrak{A} = f(\operatorname{diag}(Q_1, \ldots, Q_d, \ldots, Q_1, \ldots, Q_d)\mathbf{p}) = f(U\mathbf{p})$. As U is orthogonal, $\|\mathbf{p}\| = \|U\mathbf{p}\|$. Terracini's matrix corresponding to the norm-balanced VFMs $\mathbf{p}' = U\mathbf{p}$ is $T_{\mathbf{p}'} = \begin{bmatrix} T'_1 & \cdots & T'_r \end{bmatrix}$, where

$$T'_i = \begin{bmatrix} I \otimes Q_2\mathbf{a}_i^2 \otimes \cdots \otimes Q_d\mathbf{a}_i^d & \cdots & Q_1\mathbf{a}_i^1 \otimes \cdots \otimes Q_{d-1}\mathbf{a}_i^{d-1} \otimes I \end{bmatrix}.$$

Multiplying T'_i on the right by $D = \operatorname{diag}(Q_1, Q_2, \ldots, Q_d)$ results in $T'_iD = QT_i$, and hence $QT_{\mathbf{p}} = T_{\mathbf{p}'}\operatorname{diag}(D, \ldots, D)$. Since Q and $\operatorname{diag}(D, \ldots, D)$ are orthogonal, one finds

$$\varsigma_N(T_{\mathbf{p}}) = \varsigma_N(QT_{\mathbf{p}}) = \varsigma_N(T_{\mathbf{p}'}\operatorname{diag}(D, \ldots, D)) = \varsigma_N(T_{\mathbf{p}'}),$$

concluding the proof. □

The relative condition number of a robustly r-identifiable tensor can be bounded from below.

Proposition 16. *The relative condition number $\kappa(\mathfrak{A})$ of a robustly r-identifiable tensor $\mathfrak{A} \in \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d}$ is bounded from below by d^{-1}.*

Proof. Let $\mathbf{p} \in f_{\mathcal{R}}^{-1}(\mathfrak{A})$ be norm-balanced. Then,

$$\kappa(\mathfrak{A}) = \varsigma_N(T_{\mathbf{p}})^{-1}\frac{\|\mathfrak{A}\|}{\|\mathbf{p}\|} = \varsigma_N(T_{\mathbf{p}})^{-1}\frac{\|d^{-1}T_{\mathbf{p}}\mathbf{p}\|}{\|\mathbf{p}\|} = d^{-1}\varsigma_N(T_{\mathbf{p}})^{-1}\frac{\|T_{\mathbf{p}}\mathbf{p}\|}{\|\mathbf{p}\|}.$$

Let K be as in (7). Then by assumption on the rank of $T_{\mathbf{p}}$, $KK^{\dagger} = K(K^HK)^{-1}K^H$ is a projector onto the kernel of $T_{\mathbf{p}}$. Let $\mathbf{p}_i^T = \begin{bmatrix} (\mathbf{a}_i^1)^T & \cdots & (\mathbf{a}_i^d)^T \end{bmatrix}$, and then it follows that $K_i^H\mathbf{p}_i = 0$. Hence, $K^H\mathbf{p} = 0 = KK^{\dagger}\mathbf{p}$. So $\mathbf{p}$ is contained in $\operatorname{span}(K)^{\perp}$. Therefore, $\|T_{\mathbf{p}}\mathbf{p}\|/\|\mathbf{p}\| \geq \varsigma_N(T_{\mathbf{p}})$ and the result follows. □

It is unknown if this lower bound is sharp. In the next subsection, it is shown that this bound is not sharp for rank-1 tensors and certain weak 3-orthogonal tensors, both of which have condition number $\sqrt{d^{-1}}$.

5.2. Weak 3-orthogonal rank decompositions

The relative condition number of a rank-1 tensor depends only on the order d of the tensor and *decreases* as d increases.

Proposition 17. *A rank-1 tensor $\mathfrak{A} = \alpha \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d \in \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d}$ with $\|\mathbf{a}^k\| = 1$ and $\alpha \in \mathbb{R}_+$ strictly positive has absolute condition number $\kappa_A(\mathfrak{A}) = \alpha^{1/d-1}$ and relative condition number $\kappa(\mathfrak{A}) = \sqrt{d^{-1}}$.*

Proof. Let $\mathbf{p}' = \operatorname{vec}(\alpha^{1/d}\mathbf{a}^1, \alpha^{1/d}\mathbf{a}^2, \ldots, \alpha^{1/d}\mathbf{a}^d)$ and $\mathbf{p} = \operatorname{vec}(\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^d)$. Note that they are both norm-balanced. One verifies that $T_{\mathbf{p}'} = \alpha^{1-1/d} T_{\mathbf{p}}$. Hence, finding the singular values of $T_{\mathbf{p}}$ suffices. The largest singular value of $T_{\mathbf{p}}$ is

$$\begin{aligned}
\varsigma_1(T_{\mathbf{p}}) = \max_{\|\mathbf{v}\|=1} \|T_{\mathbf{p}}\mathbf{v}\| &= \max_{\|\mathbf{v}\|=1} \|\mathbf{v}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d + \cdots + \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^{d-1} \otimes \mathbf{v}^d\| \\
&\leq \max_{\|\mathbf{v}\|=1} \big(\|\mathbf{v}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d\| + \cdots + \|\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^{d-1} \otimes \mathbf{v}^d\|\big) \\
&= \max_{\|\mathbf{v}\|=1} \big(\|\mathbf{v}^1\| + \cdots + \|\mathbf{v}^d\|\big) = \max_{\|\mathbf{v}^1\|^2 + \cdots + \|\mathbf{v}^d\|^2 = 1} \big(\|\mathbf{v}^1\| + \cdots + \|\mathbf{v}^d\|\big) \\
&= \max_{\nu_1^2 + \cdots + \nu_d^2 = 1} (|\nu_1| + \cdots + |\nu_d|) = \max_{\|\boldsymbol{\nu}\|=1} \|\boldsymbol{\nu}\|_1 \leq \sqrt{d}\|\boldsymbol{\nu}\|_2 = \sqrt{d},
\end{aligned}$$

where $\mathbf{v}^T = \begin{bmatrix} (\mathbf{v}^1)^T \cdots (\mathbf{v}^d)^T \end{bmatrix}$ and $\boldsymbol{\nu} = \begin{bmatrix} \nu_1 & \cdots & \nu_d \end{bmatrix}$. If we take $\mathbf{v}_1 = \sqrt{d^{-1}} \operatorname{vec}(\mathbf{a}^1, \cdots, \mathbf{a}^d)$, then it follows that $\|T_{\mathbf{p}}\mathbf{v}\| = \|d\sqrt{d^{-1}}\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d\| = \sqrt{d}$. That is, $\mathbf{v}_1$ is a right singular vector corresponding to the largest singular value $\varsigma_1(T_{\mathbf{p}}) = \sqrt{d}$. Since $T_{\mathbf{p}}\mathbf{v}_1 = \sqrt{d}\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d = \sqrt{d}\mathbf{u}_1$, it follows that $\mathbf{u}_1$ is a left singular vector corresponding to $\varsigma_1(T_{\mathbf{p}})$. Deflating the largest singular tuple from $T_{\mathbf{p}}$, we find

$$\widehat{T}_{\mathbf{p}} = T_{\mathbf{p}} - \sqrt{d}\mathbf{u}_1\mathbf{v}_1^H = \begin{bmatrix} (I - \mathbf{a}^1(\mathbf{a}^1)^H) \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d & \cdots & \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^{d-1} \otimes (I - \mathbf{a}^d(\mathbf{a}^d)^H) \end{bmatrix}.$$

The singular values of the above matrix are the square roots of the eigenvalues of its Gram matrix, which has a particularly pleasing structure as it is block diagonal,

$$\widehat{T}_{\mathbf{p}}^H \widehat{T}_{\mathbf{p}} = \operatorname{diag}\Big((I - \mathbf{a}^1(\mathbf{a}^1)^H)^2, (I - \mathbf{a}^2(\mathbf{a}^2)^H)^2, \ldots (I - \mathbf{a}^d(\mathbf{a}^d)^H)^2\Big),$$

which is a consequence of the fact that $I - \mathbf{a}^k(\mathbf{a}^k)^H$ is a projector onto the orthogonal complement of $\mathbf{a}^k$, so that applying it to the latter yields 0. Since $I - \mathbf{a}^k(\mathbf{a}^k)^H$ is a projector, it is idempotent, and so the eigenvalues of $\widehat{T}_{\mathbf{p}}^H \widehat{T}_{\mathbf{p}}$ are the union of the eigenvalues of the $I - \mathbf{a}^k(\mathbf{a}^k)^H$ for $k = 1, \ldots, d$. The only possible eigenvalues of a projector are 1 and 0; so eigenvalue 1 has multiplicity equal to the rank of the projector, which is readily verified to be $n_k - 1$. Hence, the singular values of $\widehat{T}_{\mathbf{p}}$ are 1 with multiplicity Σ and 0 with multiplicity d. The singular values of $T_{\mathbf{p}}$ are $\sqrt{d}$, 1 with multiplicity Σ, and 0 with multiplicity $d - 1$. We can conclude that the $N = (\Sigma + 1)$th singular value of $T_{\mathbf{p}'}$ is $\alpha^{1-1/d}$, so that the absolute condition number is $\kappa_A = \alpha^{1/d-1}$. Finally,

$$\kappa = \kappa_A \frac{\|\alpha\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d\|}{\|\mathbf{p}'\|} = \alpha^{1/d-1} \cdot \alpha \cdot (\sqrt{d}\alpha^{1/d})^{-1} = \sqrt{d^{-1}},$$

concluding the proof. □

The following result characterizes the singular values of Terracini's matrix corresponding to a norm-balanced rank-1 tensor; it is a consequence of the foregoing proof.

Corollary 18. *Let $\mathfrak{p} = (\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^d)$ with $\|\mathbf{a}^1\| = \cdots = \|\mathbf{a}^d\| = \alpha^{1/d} > 0$, and let $\mathbf{p} = \operatorname{vec}(\mathfrak{p})$. Then, the singular values of Terracini's matrix are*

$$\varsigma(T_{\mathbf{p}}) = \{\sqrt{d}\alpha^{1-1/d}, \underbrace{\alpha^{1-1/d}, \ldots, \alpha^{1-1/d}}_{\Sigma}, \underbrace{0, \ldots, 0}_{d-1}\}.$$

Based on the foregoing characterization of the singular values of Terracini's matrix associated with one rank-1 tensor, we can determine explicitly the condition number of the class of *weak k-orthogonal tensors* with $k \geq 3$. Both "strong" and "weak" 2-orthogonality were considered in [56]; hence the additional qualifier. A tensor rank decomposition is said to be weak k-orthogonal if for every pair of rank-1 tensors, there is orthogonality in k factors, but the factors in which orthogonality occurs need not be the same for different pairs. The next result shows that tensors with a weak 3-orthogonal decomposition admit a closed expression for their condition number.

Proposition 19. *Let $\alpha_i \in \mathbb{R}_+$ be sorted as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$, and let*

$$\mathfrak{A} = \sum_{i=1}^{r} \alpha_i \mathbf{v}_i^1 \otimes \cdots \otimes \mathbf{v}_i^d \quad \textit{with } \|\mathbf{v}_i^k\| = 1$$

be a robustly r-identifiable weak 3-orthogonal tensor:

$$\forall i < j : \exists 1 \leq k_1 < k_2 < k_3 \leq d : \langle \mathbf{v}_i^{k_1}, \mathbf{v}_j^{k_1} \rangle = \langle \mathbf{v}_i^{k_2}, \mathbf{v}_j^{k_2} \rangle = \langle \mathbf{v}_i^{k_3}, \mathbf{v}_j^{k_3} \rangle = 0,$$

where $\langle \cdot, \cdot \rangle$ is the Hermitian inner product. Then,

$$\kappa_A(\mathfrak{A}) = \alpha_r^{-1+1/d} \quad \textit{and} \quad \kappa(\mathfrak{A}) = \alpha_r^{-1+1/d} \sqrt{\sum_{i=1}^{r} \alpha_i^2} \Big/ \sqrt{\sum_{i=1}^{r} d\alpha_i^{2/d}}.$$

Proof. Let $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \ldots, \mathfrak{p}_r)$ with $\mathfrak{p}_i = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d)$ be norm-balanced. By weak 3-orthogonality, we have for every $1 \leq i < j \leq r$ and every $1 \leq k_1, k_2 \leq d$ that

$$(\mathbf{v}_i^1 \otimes \cdots \otimes \mathbf{v}_i^{k_1 - 1} \otimes I \otimes \mathbf{v}_i^{k_1 + 1} \otimes \cdots \otimes \mathbf{v}_i^d)^H (\mathbf{v}_j^1 \otimes \cdots \otimes \mathbf{v}_j^{k_2 - 1} \otimes I \otimes \mathbf{v}_j^{k_2 + 1} \otimes \cdots \otimes \mathbf{v}_j^d) = 0.$$

Hence, $T_i^H T_j = 0$ for all $i \neq j$, so that $T_{\mathbf{p}}^H T_{\mathbf{p}} = \operatorname{diag}(T_1^H T_1, T_2^H T_2, \ldots, T_r^H T_r)$, where T_i is as in (6). The singular values of $T_{\mathbf{p}}$ thus coincide with the union of the set of singular values of each of the T_i's. From Corollary 18 it follows that

$$\varsigma(T_{\mathbf{p}}) = \bigcup_{i=1}^{r} \{\sqrt{d}\alpha_i^{1-1/d}, \alpha_i^{1-1/d}, \ldots, \alpha_i^{1-1/d}, 0, \ldots, 0\};$$

the Nth singular value is thus $\varsigma_N(T_{\mathbf{p}}) = \alpha_r^{1-1/d}$ by our assumption on the order of the α_i. The squared norm $\|\mathfrak{A}\|^2 = \sum_{i=1}^{r} \alpha_i^2$ because of the orthogonality of the $\mathbf{v}_i^1 \otimes \cdots \otimes \mathbf{v}_i^d$'s. The norm of $\mathbf{p}$ is clear. □

Corollary 20. *If all coefficients α_i in Proposition 19 are equal, then the relative condition number is $\sqrt{d^{-1}}$.*

Remark 21. *The class of weak* 3*-orthogonal tensors includes rank-*1 *tensors and* orthogonally decomposable tensors *(odeco) of order $d \geq 3$, which have been studied extensively in [11, 14, 31, 44, 56, 61].*

6. Numerical experiments

In this section, the condition number is computed numerically for some tensors. The implementation of this algorithm in Matlab/Octave that was used, is provided at `https://arxiv.org/abs/1604.00052`. All of the experiments were performed using Matlab R2015a on a computer system consisting of an Intel Core i7-5600U CPU, clocked at 2.6GHz, and 8GB of main memory. Tensorlab v3 [59] was employed for computing approximate tensor rank decompositions. Additional experiments can be found in [55].

6.1. The main theorem

For illustrating Theorem 10, the quantity on the left hand side of

$$\frac{\|\mathbf{p}-\widehat{\mathbf{p}}\|/\|\mathbf{p}\|}{\|f(\mathbf{p})-f(\widehat{\mathbf{p}})\|/\|f(\mathbf{p})\|} \geq \frac{\operatorname{dist}(\mathbf{p},[\widehat{\mathbf{p}}])/\|\mathbf{p}\|}{\|f(\mathbf{p})-f(\widehat{\mathbf{p}})\|/\|f(\mathbf{p})\|} \tag{12}$$

will be investigated as a proxy for the right hand side, where $\mathbf{p}$ and $\widehat{\mathbf{p}}$ are both norm-balanced VFMs. We can estimate the condition number κ by generating a large number of vectors $\widehat{\mathbf{p}}$ such that $\|f(\mathbf{p})-f(\widehat{\mathbf{p}})\| \leq \varepsilon\|f(\mathbf{p})\|$, for some small value of ε, then computing the left-hand side of (12), and finally taking the maximum value over all samples. For example, a (positive) rank-2 decomposition in $\mathbb{R}^{3\times3\times2}$ was randomly chosen with norm-balanced representatives

$$\mathfrak{p}_1 = 10^{-1}\left(\begin{bmatrix}5.1518\\4.9802\\5.0806\end{bmatrix}, \begin{bmatrix}1.9032\\5.4218\\6.6436\end{bmatrix}, \begin{bmatrix}7.2302\\4.9879\end{bmatrix}\right) \text{ and } \mathfrak{p}_1 = 10^{-1}\left(\begin{bmatrix}8.8821\\3.6941\\1.1117\end{bmatrix}, \begin{bmatrix}7.5082\\1.6653\\5.8845\end{bmatrix}, \begin{bmatrix}6.9447\\6.7487\end{bmatrix}\right).$$

Let $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \mathfrak{p}_2)$ be the VFMs, and let $\mathfrak{A} = f(\mathbf{p})$ be the corresponding tensor. The relative condition number is $\kappa(\mathfrak{A}) \approx 18.410787$. Let $\mathfrak{p}_i = (\mathbf{a}_i^1, \mathbf{a}_i^2, \mathbf{a}_i^3)$. Then, the factor matrices $F_k = [\,\mathbf{a}_1^k\ \mathbf{a}_2^k\,]$ corresponding to $\mathbf{p}$ were perturbed randomly as follows: $\widehat{F}_k = F_k + 10^{-4}\cdot\|F_k\|_F \cdot X_k$ where X_k has elements sampled from the standard uniform distribution $N(0,1)$. Starting from these factor matrices, the `cpd_nls` function in Tensorlab computed an approximate tensor rank decomposition of $\mathfrak{A}$. Let $\widehat{\mathbf{p}}$ denote the norm-balanced VFMs obtained by the `cpd_nls` function. This computation is repeated 1 million times; for every instance where $\|\mathfrak{A}-f(\widehat{\mathbf{p}})\| \leq 10^{-14}\cdot\|\mathfrak{A}\|$, the value on the left-hand side of (12) is computed. This resulted in $278,527$ valid samples. Taking their maximum value yielded approximately 10.7102 as estimate for the condition number. This is actually a poor approximation to κ. In fact, the mean estimate of the condition number over all valid trials was only 1.63618, suggesting that the average small perturbation to $\mathfrak{A}$ only causes the parameters $\mathbf{p}$ to change only about twice as much as the tensor.

Section Appendix C.3 proves that the right singular vector $\mathbf{v}_{12}$ corresponding to the $N = 12$th singular value of $T_{\mathbf{p}}$ is the worst direction of perturbation. If we consider a small perturbation in this direction, e.g., $\widehat{\mathbf{p}} = \mathbf{p} + 10^{-8}\mathbf{v}_{12}$, then we get a substantially more accurate estimate, namely 18.410764, which is a relative difference of about $1\cdot10^{-6}$ with respect to the true condition number.

6.2. An ill-conditioned example

Consider the following sequence of tensors

$$\mathfrak{A}_s = f(\mathbf{p}_s) = \mathbf{a}_1^1\otimes\mathbf{a}_1^2\otimes(\mathbf{x}+2^{-s}\mathbf{a}_1^3) + \mathbf{a}_2^1\otimes\mathbf{a}_2^2\otimes(\mathbf{x}+2^{-s}\mathbf{a}_2^3), \tag{13}$$

where $\mathbf{a}_i^k \in \mathbb{R}^{n_k}$ are random vectors sampled from a normal distribution, $\mathbf{x} \in \mathbb{R}^{n_3}$ is any non-zero vector, and $\mathbf{p}_s$ are norm-balanced VFMs. Every tensor $\mathfrak{A}_s$ on this sequence is robustly 2-identifiable by Kruskal's theorem with probability 1. However, $\lim_{s\to\infty}\mathfrak{A}_s = (\mathbf{a}_1^1\otimes\mathbf{a}_1^2 + \mathbf{a}_2^1\otimes\mathbf{a}_2^2)\otimes\mathbf{x}$, which is not 2-identifiable. In fact, $\mathfrak{A}_\infty$ has infinitely many decompositions of length 2. Recall from Lemma 37 of [17] that Terracini's matrix $T_{\mathbf{p}}$ is of rank strictly less than N, where $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1, \mathfrak{p}_2)$ with $\mathfrak{p}_i = (\mathbf{a}_i^1, \mathbf{a}_i^2, \mathbf{x})$. By continuity in a neighborhood of $\mathfrak{A}_\infty$ this entails that $\kappa(\mathfrak{A}_s) \to \infty$ as $s \to \infty$, which we now confirm numerically.

The decomposition of $\mathfrak{A}_s$ can be computed by a direct decomposition algorithm [19, 22, 36]. Let $\widehat{\mathbf{p}}_s$ be the norm-balanced VFMs that were computed by such a numerical algorithm when $\mathfrak{A}_s$ is provided as input. Because of roundoff errors $\widehat{\mathbf{p}}_s \neq \mathbf{p}_s$ in general, so that the algorithm actually computed the decomposition of a nearby tensor $\widehat{\mathfrak{A}}_s = f(\widehat{\mathbf{p}}_s)$. The relative backward error is $\|\mathfrak{A}_s - \widehat{\mathfrak{A}}_s\|/\|\mathfrak{A}_s\|$, which measures how close the rank decomposition problem that was solved by the numerical algorithm was to the true rank decomposition problem. The condition number allows us to asymptotically bound the relative forward error between the computed (norm-balanced) and the true VFMs:

$$\frac{\operatorname{dist}(\widehat{\mathbf{p}}_s, f_{\mathcal{R}}^{-1}(\mathfrak{A}_s))}{\|\widehat{\mathbf{p}}_s\|} \lesssim \kappa(\widehat{\mathfrak{A}}_s)\cdot\frac{\|\mathfrak{A}_s-\widehat{\mathfrak{A}}_s\|}{\|\widehat{\mathfrak{A}}_s\|}.$$

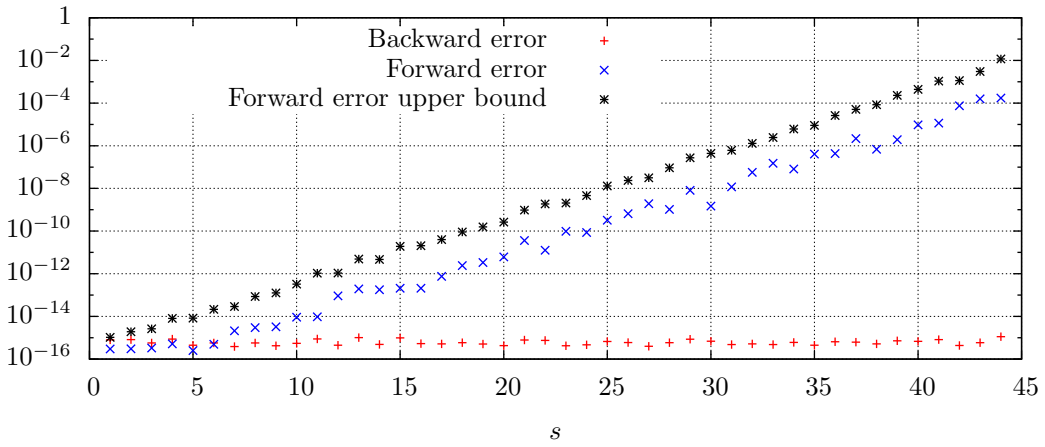

Figure 1: The relative backward error, relative forward error, and relative condition number multiplied with the relative backward error of a particular instance of the sequence in (13).

We will investigate four quantities, namely the relative forward error, the condition number, the relative backward error, and the asymptotic upper bound on the right hand side in the last inequality.

As a concrete problem instance, consider rank-2 tensors in $\mathbb{R}^{13\times 11\times 7}$ whose factor matrices `F(s)` were generated as follows:

```
A = randn(13,2); B = randn(11,2); C = randn(7,2); x = randn(7,1);
F = @(s) {A, B, x*[1 1] + 2^(-s) * C};
```

The `cpd_gevd` function in Tensorlab was used for computing a direct decomposition of $\mathfrak{A}_s$ for the values $s = 1, 2, \ldots, 45$. For simplicity, $\|\widehat{\mathbf{p}}_s - \mathbf{p}_s\|$ is used as a proxy for the forward error $\operatorname{dist}(\widehat{\mathbf{p}}_s, [\mathbf{p}_s])$. The backward error, the proxy of the forward error, and estimated forward error obtained by multiplying the condition number with the backward error are all plotted in function of s in Fig. 1. The figure shows that the forward error grows as s increases, which is driven by the increase of the condition number from roughly $\mathcal{O}(1)$ to $\mathcal{O}(10^{13})$, confirming the implication of [17, Lemma 37] for condition numbers.

Fig. 1 illustrates the perils of investigating only the relative backward error. One sees that the backward error of `cpd_gevd` is of the order of the machine precision. The forward error, however, is several orders of magnitude larger. In this example, there are no telltale signs that the stability of the computed solution as measured by the forward error is doubtful for $s = 1, \ldots, 44$. The `cpd_gevd` algorithm behaves exactly the same for $s = 1$ and $s = 44$; there is no increase of the computation time or backward error, and the computed decomposition is in most regards normal except for the near rank deficiency in the third factor matrix. Nevertheless, the condition number correctly reveals the instability of the solution. It is only for $s \geq 45$ that Tensorlab's implementation of the ST-HOSVD algorithm [57] detects that the multilinear rank of $\mathfrak{A}_s$ is very close to $(2, 2, 1)$, at which point the software prudently refuses to employ the generalized eigenvalue decomposition algorithm, hinting at an ill-conditioned TDP.

6.3. An ill-posed example

Recall that for both $\mathbb{F} = \mathbb{C}$ and $\mathbb{R}$ the set of tensors of $\mathbb{F}$-rank bounded by r is in general not closed [20]. This entails that there exist tensors of rank strictly greater than r that can nevertheless be approximated arbitrarily well by tensors of rank r. Such tensors cause problems when trying to approximate them by a rank-r tensor [20]. According to Demmel [21], a problem that is close to an ill-posed problem is often ill-conditioned. We investigate numerically whether this property is admitted by the proposed condition number by considering the example of [20]:

$$f(\mathbf{p}_s) = 2^{s/5}(\mathbf{a}^1 + 2^{-s/5}\mathbf{b}^1) \otimes (\mathbf{a}^2 + 2^{-s/5}\mathbf{b}^2) \otimes (\mathbf{a}^3 + 2^{-s/5}\mathbf{b}^3) - 2^{s/5}\mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \mathbf{a}^3, \tag{14}$$

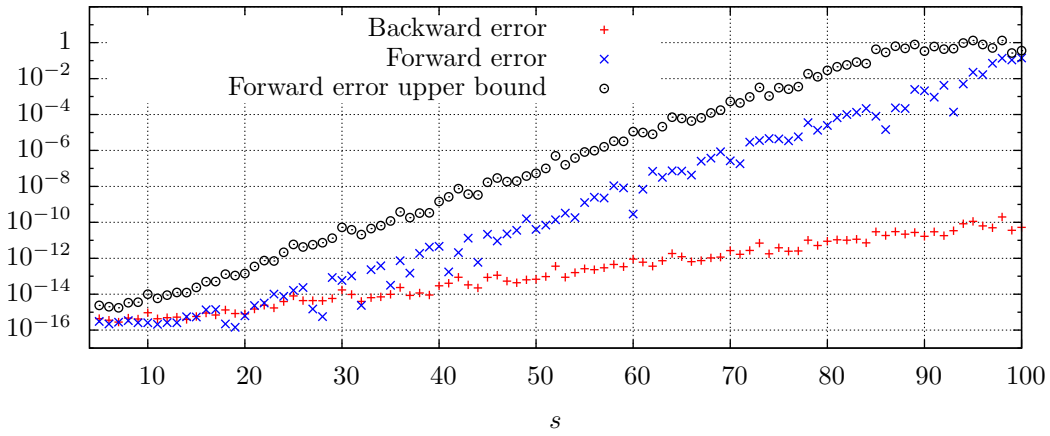

Figure 2: The relative backward error, relative forward error, and relative condition number multiplied with the relative backward error of a particular instance of the sequence in (14).

where $\mathbf{a}^k, \mathbf{b}^k \in \mathbb{F}^{n_k}$ are linearly independent vectors and $\mathbf{p}_s$ are norm-balanced VFMs. Every tensor on this sequence is robustly 2-identifiable by Kruskal's criterion, and the limit of this sequence of rank-2 tensors is the rank-3 tensor $\lim_{s\to\infty} \mathfrak{A}_s = \mathbf{b}^1 \otimes \mathbf{a}^2 \otimes \mathbf{a}^3 + \mathbf{a}^1 \otimes \mathbf{b}^2 \otimes \mathbf{a}^3 + \mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \mathbf{b}^3$. As in the previous example, the tensor rank decomposition of $\mathfrak{A}_s = f(\mathbf{p}_s)$ can be computed by a direct algorithm.

As a particular instance, let $\mathbf{a}^1, \mathbf{b}^1 \in \mathbb{R}^5$, $\mathbf{a}^2, \mathbf{b}^2 \in \mathbb{R}^4$ and $\mathbf{a}^3, \mathbf{b}^3 \in \mathbb{R}^3$ be vectors whose entries were sampled from a standard normal distribution. For all $s = 5, 6, \ldots, 100$, the rank-2 decomposition was computed with `cpd_gevd`. Let $\widehat{\mathbf{p}}_s$ denote the norm-balanced VFMs obtained from applying this numerical algorithm to $\mathfrak{A}_s$, and let $\widehat{\mathfrak{A}}_s = f(\widehat{\mathbf{p}}_s)$. The relative backward error, the proxy of the relative forward error, and the relative condition number at $\widehat{\mathfrak{A}}_s$ multiplied with the relative backward error were recorded. These quantities are plotted in Fig. 2.

From the figure it can be deduced that the condition number increases from $\mathcal{O}(1)$ to $\mathcal{O}(10^{10})$ as s increases from 5 to 90, suggesting that the condition number indeed deteriorates as one moves closer to the ill-posed tensor rank decomposition problem. Notice that the upper bound on the forward error seems to stagnate around $s = 85$. *This is a numerical issue in the computation of the condition number.* It can be verified that the condition number keeps increasing as s increases when employing variable precision arithmetic in Matlab. Fortunately, the occurrence of these numerical difficulties in computing the condition number can be detected by verifying that the Nth singular value of Terracini's matrix $T_{\mathbf{p}_s}$ is larger than a small constant multiple of the largest singular value $\varsigma_1(T_{\mathbf{p}_s})$ multiplied with the machine precision ($\epsilon_{\text{mach}} \approx 2.2 \cdot 10^{-16}$ in standard double precision in Matlab).

The behavior of the condition number is similar in other known examples of sequences tending to an ill-posed tensor rank decomposition problem; see Section 9.5 of [55].

7. Conclusions

Two strategies for defining condition numbers of respectively the TAP and TDP based on the classic framework of conditioning were investigated. A third strategy for the TDP based on the geometric framework of [8, 13] is investigated in a follow-up article by Breiding and the author [12].

Strategy I is based on the modification of the model by explicit elimination of variables, whereas Strategy II relaxes the metric for measuring distances between representatives so that the classic framework applies to the usual over-parameterized model. Terracini's matrix lies at the heart of the conditioning of robustly r-identifiable tensors for both strategies I and II. It was shown that the condition number of the TDP derived via Strategy II is always less, i.e., better, than the condition number of the TAP derived via Strategy I. In addition, the elimination of variables in Strategy I is an arbitrary choice. For these reasons, only

Strategy II was developed in substantial detail. The corresponding condition number equals the inverse of the Nth largest singular value of Terracini's matrix. Provided that the input tensor is robustly r-identifiable, the (relative) condition number multiplied with the (relative) backward error yields an asymptotically sharp upper bound on the (relative) forward error. The analysis of Terracini's matrix shows that rank-1 tensors are always well-conditioned; their condition number admits a closed expression. The condition number of weak 3-orthogonal tensors was completely described; they are well-conditioned only if the norms of the individual rank-1 terms are approximately of the same order of magnitude. Finally, the numerical experiments provide some preliminary evidence suggesting that tensors in a sequence of robustly r-identifiable tensors whose limit is a tensor of rank strictly larger (see Section 6.3) are ill-conditioned near the limit.

The study of conditioning for the tensor rank decompositions is far from complete. A natural question concerns the condition number of more structured decompositions such as nonnegative and (partially) symmetric decompositions. The condition number derived in Theorem 10 provides a natural upper bound for the condition number of such decompositions, but it is unclear to what extent imposing constraints can improve conditioning. The last part of the proof of the main theorem shows that the worst perturbation to the VFMs $\mathbf{p}$ is given by the right singular vector corresponding to the Nth largest singular value of Terracini's matrix $T_{\mathbf{p}}$; this essentially enables an identification of the components of the VFMs that are most sensitive to perturbations of the tensor. I believe that this is a useful property in data analysis applications.

Acknowledgements

This paper benefited immensely from the discussions I had with B. Jeuris, K. Meerbergen, G. Ottaviani, and G. Tomasi. I am particularly indebted to J. Nicaise for inquiring about an intrinsic definition of the condition number. I thank P. Breiding and L.-H. Lim for detailed comments they provided on an earlier version of this manuscript. An anonymous reviewer is heartily thanked for his or her detailed comments and questions that greatly improved this lengthy paper.

Appendix A. Proof of the identifiability propositions

Proof of Proposition 7. Let

$$\mathcal{D} = \big\{\operatorname{diag}\big((\alpha_2 \cdots \alpha_d)^{-1} I_{n_1}, \alpha_2 I_{n_2}, \ldots, \alpha_d I_{n_d}\big) \;\big|\; \alpha_2, \ldots, \alpha_d \in \mathbb{F}_0\big\}, \tag{A.1a}$$

$$\mathcal{B} = \big\{\operatorname{diag}(D_1, D_2, \ldots, D_r) \mid D_1, D_2, \ldots, D_r \in \mathcal{D}\big\}, \tag{A.1b}$$

$$\mathcal{P} = \big\{P \otimes I_{\Sigma+d} \mid P \text{ is an } r \times r \text{ permutation matrix}\big\}, \text{ and} \tag{A.1c}$$

$$\mathcal{T} = \big\{PB \mid P \in \mathcal{P} \text{ and } B \in \mathcal{B}\big\}. \tag{A.1d}$$

It is straightforward to verify that both $\mathcal{B}$ and $\mathcal{P}$ are multiplicative groups. For every $P \otimes I_{\Sigma+d} \in \mathcal{P}$ and diagonal matrices $D_i \in \mathbb{F}^{\Sigma+d \times \Sigma+d}$ we have

$$(P \otimes I_{\Sigma+d}) \cdot \operatorname{diag}(D_1, D_2, \ldots, D_r) \cdot (P \otimes I_{\Sigma+d})^T = \operatorname{diag}(D_{\pi_1}, D_{\pi_2}, \ldots, D_{\pi_r}), \tag{A.2}$$

where π is the permutation represented by P. Hence, (A.2) defines a map $\psi : \mathcal{P} \to \operatorname{Aut}(\mathcal{B})$ that takes $x \in \mathcal{P}$ to $\psi_x \in \operatorname{Aut}(\mathcal{B})$ defined by $\psi_x : y \mapsto xyx^{-1}$. It follows (see, e.g., [45, Chapter 7]) that $\mathcal{T}$ is the semidirect product $\mathcal{T} = \mathcal{P} \rtimes \mathcal{B}$, which is thus a multiplicative group. Writing $\mathbf{p} \sim \mathbf{q}$ if and only if $\mathbf{p} = T\mathbf{q}$ for some $T \in \mathcal{T}$, it follows that $\sim$ is an equivalence relationship and that f is a morphism for $\sim$. □

Proof of Proposition 9. Write $\mathbf{x} = \operatorname{vecr}(\mathfrak{p}_1, \ldots, \mathfrak{p}_r)$ with $\mathfrak{p}_i = (\mathbf{a}_i^1, \ldots, \mathbf{a}_i^d)$, and $\mathbf{y} = \operatorname{vecr}(\mathfrak{q}_1, \ldots, \mathfrak{q}_r)$ with $\mathfrak{q}_i = (\mathbf{b}_i^1, \ldots, \mathbf{b}_i^d)$. Let $\mathcal{T}$ be as in the proof of Proposition 7. Note that $h(T) = \|\mathbf{x} - T\mathbf{y}\|$ is a coercive function in the parameters defining $T \in \mathcal{T}$, so that for a fixed input the minimizer of the optimization problem is always attained. Indeed, since $I \in \mathcal{T}$, it follows that the optimal $(P \otimes I_{\Sigma+d})B \in \mathcal{T}$ satisfies

$$\|\mathbf{x} - (P \otimes I_{\Sigma+d})B\mathbf{y}\|^2 = \sum_{i=1}^{r} \|\operatorname{vec}(\mathfrak{p}_i) - D_{\pi_i} \operatorname{vec}(\mathfrak{q}_{\pi_i})\|^2$$

$$= \sum_{i=1}^{r}\sum_{k=2}^{d} \|\mathbf{a}_i^k - \theta_{k,\pi_i}\mathbf{b}_{\pi_i}^k\|^2 + \sum_{i=1}^{r} \|\mathbf{a}_i^1 - (\theta_{2,\pi_i}\cdots\theta_{d,\pi_i})^{-1}\mathbf{b}_{\pi_i}^1\|^2 \leq \|\mathbf{x}-\mathbf{y}\|^2 < \infty$$

where $B = \operatorname{diag}(D_1, \ldots, D_r)$, $D_i = \operatorname{diag}(\theta_{2,i}^{-1}\cdots\theta_{d,i}^{-1} I_{n_1}, \theta_{2,i} I_{n_2}, \ldots, \theta_{d,i} I_{n_d})$ and π is the permutation represented by the permutation matrix $P \in \mathbb{F}^{r\times r}$. Since we have a sum of positive reals, it follows that all $|\theta_{k,i}|$ are uniformly bounded from above by some positive constant C, and, hence there is also a uniform lower bound $0 < c$. So we can optimize $\theta_{k,i}$ over the bounded intervals $[-C,-c]\cup[c,C]$ leading to the conclusion that the infimum is attained.

Now the last claim about $\operatorname{dist}(\mathbf{x},[\mathbf{y}]) = 0$ is trivial. □

Appendix B. The Iterated Scaling Lemma

For proving the main result, we need the following technical lemma. Let $\mathcal{D}$ be as in (A.1). I will abuse notation, writing $\mathfrak{q} \in \mathcal{D}\mathfrak{p}$ when $\operatorname{vec}(\mathfrak{q}) \in \mathcal{D}\operatorname{vec}(\mathfrak{p})$ is meant, where $\mathfrak{q},\mathfrak{p} \in \mathbb{F}^{n_1}\times\cdots\times\mathbb{F}^{n_d}$.

Lemma 22 (Iterated Scaling)**.** *Let $\mathcal{S}_{\mathbb{F}}$ be a Segre variety. For $i = 1,2,\ldots,r$, let $\mathfrak{p}_i = (\mathbf{a}_i^1,\ldots,\mathbf{a}_i^d)$, $\nabla_i = (\mathbf{n}_i^1,\ldots,\mathbf{n}_i^d)$, and $\mathfrak{q}_i = \mathfrak{p}_i + \nabla_i = (\mathbf{a}_i^1+\mathbf{n}_i^1,\ldots,\mathbf{a}_i^d+\mathbf{n}_i^d)$, where $\operatorname{Seg}(\mathfrak{p}_i), \operatorname{Seg}(\mathfrak{q}_i) \in \mathcal{S}_{\mathbb{F}}$. Assume that the perturbation $\boldsymbol{\nabla} = \operatorname{vecr}(\nabla_1,\nabla_2,\ldots,\nabla_r)$ is of sufficiently small norm:*

$$\|\boldsymbol{\nabla}\| \leq \frac{1}{2}\lambda^{-1} = \frac{1}{2^{d+4}(d-1)^{3/2}} \cdot \min_{1\leq i\leq r}\min_{1\leq k\leq d}\big(\|\mathbf{a}_i^1\|^{-1}\|\mathbf{a}_i^k\|^2\big). \tag{B.1}$$

If Terracini's matrix $T_{\mathbf{p}}$ associated with the VFMs $\mathbf{p} = \operatorname{vecr}(\mathfrak{p}_1,\ldots,\mathfrak{p}_r)$ has rank $N = r(\Sigma+1)$ and if $\boldsymbol{\nabla}$ is contained in its kernel, i.e., $T_{\mathbf{p}}\boldsymbol{\nabla} = 0$, then there is a representative $\dot{\mathfrak{p}}_i \in \mathcal{D}\mathfrak{p}_i$ such that $\mathfrak{p}_i + \nabla_i = \mathfrak{q}_i = \dot{\mathfrak{p}}_i + \Delta_i$ with $\boldsymbol{\Delta}_i = \operatorname{vec}(\Delta_i) \in \operatorname{span}(K_i)^{\perp}$ and $\|\boldsymbol{\Delta}_i\| \leq 2\lambda\|\boldsymbol{\nabla}_i\|^2 \leq \|\boldsymbol{\nabla}_i\|$; herein, $\boldsymbol{\nabla}_i = \operatorname{vec}(\nabla_i)$ and K_i as in (7). In other words, letting K be as in (7), there exists a factorization

$$\mathbf{p} + \boldsymbol{\nabla} = \dot{\mathbf{p}} + \boldsymbol{\Delta} \quad \text{with } \dot{\mathbf{p}} \sim \mathbf{p},\ \boldsymbol{\Delta} \in \operatorname{span}(K)^{\perp},\ \text{and } \|\boldsymbol{\Delta}\| \leq 2\lambda\|\boldsymbol{\nabla}\|^2 \leq \|\boldsymbol{\nabla}\|.$$

Furthermore, letting $T_{\dot{\mathbf{p}}}$ denote Terracini's matrix in $\dot{\mathbf{p}} = \operatorname{vecr}(\dot{\mathfrak{p}}_1,\ldots,\dot{\mathfrak{p}}_r)$, then

$$T_{\dot{\mathbf{p}}} = T_{\mathbf{p}}\dot{D} = T_{\mathbf{p}}(I + \dot{E}),$$

where the diagonal matrix $\dot{D}$ tends to the identity as $\|\boldsymbol{\nabla}\|$ tends to zero; specifically, there is a constant $C > 0$ such that $\|\dot{E}\|_2 = \|\dot{D} - I\|_2 \leq C\|\boldsymbol{\nabla}\|$.

The rest of this appendix is devoted to the proof. We show the existence of a linearly convergent sequence of representatives $\mathfrak{p}_i^{(k)} \in \mathcal{D}\mathfrak{p}_i$, $k = 1,2,\ldots$, with the following properties

$$\mathfrak{q}_i - \mathfrak{p}_i^{(k)} = \Delta_i^{(k)} + \nabla_i^{(k)} \text{ with } \boldsymbol{\Delta}_i^{(k)} = \operatorname{vec}(\Delta_i^{(k)}) \in \operatorname{span}(K_i)^{\perp} \text{ and } \boldsymbol{\nabla}_i^{(k)} = \operatorname{vec}(\nabla_i^{(k)}) \in \operatorname{span}(K_i). \tag{H0a}$$

The sequence will be constructed in such a way that

$$\lim_{k\to\infty} \|\boldsymbol{\nabla}_i^{(k)}\| \to 0. \tag{H0b}$$

Under the assumptions of the lemma both $\mathbf{p}_i^{(k)} = \operatorname{vec}(\mathfrak{p}_i^{(k)})$ and $\boldsymbol{\Delta}_i^{(k)}$ will be of uniformly bounded norm, so that a convergent subsequence exists for which both $\lim_{k\to\infty}\boldsymbol{\Delta}_i^{(k)} \to \boldsymbol{\Delta}_i = \operatorname{vec}(\Delta_i)$ and $\lim_{k\to\infty}\mathbf{p}_i^{(k)} \to \dot{\mathbf{p}}_i = \operatorname{vec}(\dot{\mathfrak{p}}_i)$ are well defined. It remains to show that a sequence satisfying (H0a) and (H0b) exists.

Appendix B.1. A recurrence relation

Let $\mathfrak{p}_i^{(1)} = \mathfrak{p}_i$ and $\Delta_i^{(1)} = 0$, and define $\nabla_i^{(1)} = \nabla_i$. Since $\boldsymbol{\nabla}^{(1)} = \mathrm{vecr}(\nabla_1^{(1)}, \ldots, \nabla_r^{(1)}) = \boldsymbol{\nabla}$ is contained in the kernel, we have $T_{\mathbf{p}}\boldsymbol{\nabla}^{(1)} = \sum_{i=1}^r T_i \boldsymbol{\nabla}_i^{(1)} = 0$, where $T_{\mathbf{p}}$ and T_i are as in (6). Suppose that $T_i\boldsymbol{\nabla}_i^{(1)} \neq 0$, so that $\boldsymbol{\nabla}_i^{(1)}$ would not be contained in the span of K_i. Then $T_{\mathbf{p}}$'s rank would be strictly less than the expected value $r(\Sigma+1)$, which is a contradiction. Hence, the base case $k=1$ of (H0a) is true. Assume now that the statement is true for all $l = 1, 2, \ldots, k$, and then we show that it holds for $k+1$ as well. Since K_i is a basis of the kernel of T_i, we can express

$$\boldsymbol{\nabla}_i^{(k)} = \begin{bmatrix}\mathbf{k}_i^2 & \cdots & \mathbf{k}_i^d\end{bmatrix} \mathbf{v}_i^{(k)} = K_i \mathbf{v}_i^{(k)} \tag{B.2}$$

for some $\mathbf{v}_i^{(k)} \in \mathbb{F}^{d-1}$. By (H0a), $\mathfrak{p}_i^{(k)} \in \mathcal{D}\mathfrak{p}_i$, so that we can write it explicitly as

$$\mathfrak{p}_i^{(k)} = (\gamma_{1,i}^{(k)}\mathbf{a}_i^1, \gamma_{2,i}^{(k)}\mathbf{a}_i^2, \ldots, \gamma_{d,i}^{(k)}\mathbf{a}_i^d); \tag{B.3}$$

this is true with $\gamma_{j,i}^{(1)} = 1$ for the base case, and will follow shortly for the induction step as well. Consider the perturbed representative $\mathfrak{p}_i^{(k)} + \nabla_i^{(k)}$. Writing $\nabla_i^{(k)}$ with respect to the particular basis K_i of the kernel of T_i as in (B.2), we find that

$$\mathfrak{z}_i^{(k)} = \mathfrak{p}_i^{(k)} + \nabla_i^{(k)} = \big((\gamma_{1,i}^{(k)} + v_{2,i}^{(k)} + \cdots + v_{d,i}^{(k)})\mathbf{a}_i^1, (\gamma_{2,i}^{(k)} - v_{2,i}^{(k)})\mathbf{a}_i^2, \ldots, (\gamma_{d,i}^{(k)} - v_{d,i}^{(k)})\mathbf{a}_i^d\big),$$

where $v_{j,i}^{(k)}$ denotes the $(j-1)$th element of $\mathbf{v}_i^{(k)}$. Consequently, $\mathrm{Seg}(\mathfrak{z}_i^{(k)})$ and $\mathrm{Seg}(\mathfrak{p}_i^{(k)})$ are multiples of each other. The following representative

$$\mathfrak{p}_i^{(k+1)} := \Big(\prod_{j=2}^d (\gamma_{j,i}^{(k)} - v_{j,i}^{(k)})^{-1}\mathbf{a}_i^1, (\gamma_{2,i}^{(k)} - v_{2,i}^{(k)})\mathbf{a}_i^2, \ldots, (\gamma_{d,i}^{(k)} - v_{d,i}^{(k)})\mathbf{a}_i^d\Big) \in \mathcal{D}\mathfrak{p}_i \tag{B.4}$$

will induce the required sequence. Define

$$\mathfrak{z}_i^{(k)} - \mathfrak{p}_i^{(k+1)} = \left(\Big(\gamma_{1,i}^{(k)} + \sum_{j=2}^d v_{j,i}^{(k)} - \prod_{j=2}^d (\gamma_{j,i}^{(k)} - v_{j,i}^{(k)})^{-1}\Big)\mathbf{a}_i^1, 0, \ldots, 0\right) =: \nabla_i^{(k+1)} + \widehat{\Delta}_i^{(k+1)}, \tag{B.5}$$

where $\boldsymbol{\nabla}_i^{(k+1)} = \mathrm{vec}(\nabla_i^{(k+1)}) \in \mathrm{span}(K_i)$ and $\mathrm{vec}(\widehat{\Delta}_i^{(k+1)}) \in \mathrm{span}(K_i)^\perp$. Then,

$$\mathfrak{z}_i^{(k)} = \mathfrak{p}_i^{(k)} + \nabla_i^{(k)} = \mathfrak{p}_i^{(k+1)} + \nabla_i^{(k+1)} + \widehat{\Delta}_i^{(k+1)}.$$

Since this must be inductively true for $l = 1, 2, \ldots, k$, we find that

$$\mathfrak{q}_i = \mathfrak{z}_i^{(1)} = \mathfrak{p}_i^{(k+1)} + \nabla_i^{(k+1)} + \sum_{j=2}^{k+1} \widehat{\Delta}_i^{(j)} = \mathfrak{p}_i^{(k+1)} + \nabla_i^{(k+1)} + \Delta_i^{(k+1)}. \tag{B.6}$$

This proves (H0a).

Appendix B.2. Convergence

It remains to show (H0b), i.e., the sequence of $\boldsymbol{\nabla}_i^{(k)}$'s converges to zero. We demonstrate that

$$\|\boldsymbol{\nabla}_i^{(k)}\| \leq \lambda^{k-1}\|\boldsymbol{\nabla}_i\|^k \quad \text{for some real constant } \lambda > 0; \tag{H1}$$

for the base case $k=1$ this is true by definition. We can assume the truth of the above statement for $l = 1, 2, \ldots, k$. From (B.5) it immediately follows that

$$\|\boldsymbol{\nabla}_i^{(k+1)}\| \leq \|\mathbf{z}_i^{(k)} - \mathbf{p}_i^{(k+1)}\| = \|\mathbf{a}_i^1\| \cdot \left|\gamma_{1,i}^{(k)} + \sum_{j=2}^d v_{j,i}^{(k)} - \prod_{j=2}^d (\gamma_{j,i}^{(k)} - v_{j,i}^{(k)})^{-1}\right|,$$

where $\mathbf{z}_i^{(k)} = \mathrm{vec}(\mathfrak{z}_i^{(k)})$. Our efforts will be focused on showing that the right hand side is bounded by $\lambda^k \|\boldsymbol{\nabla}_i\|^{k+1}$. The boundedness of $\|\boldsymbol{\Delta}_i^{(k+1)}\|$ would then follow from the observation that

$$\|\boldsymbol{\Delta}_i^{(k+1)}\| \leq \sum_{j=2}^{k+1} \|\widehat{\boldsymbol{\Delta}}_i^{(j)}\| \leq \sum_{j=1}^{k} \|\mathbf{z}_i^{(j)} - \mathbf{p}_i^{(j+1)}\| \leq \sum_{j=1}^{k} \lambda^j \|\boldsymbol{\nabla}_i\|^{j+1} \leq \lambda^{-1} \sum_{j=1}^{\infty} (\lambda\|\boldsymbol{\nabla}_i\|)^{j+1} = \lambda \frac{\|\boldsymbol{\nabla}_i\|^2}{1-\lambda\|\boldsymbol{\nabla}_i\|},$$

where $\widehat{\boldsymbol{\Delta}}_i^{(j)} = \mathrm{vec}(\widehat{\Delta}_i^{(j)})$, and provided that $\lambda\|\boldsymbol{\nabla}_i\| < 1$ but this will be satisfied by our assumptions. The boundedness of $\mathbf{p}_i^{(k)}$ is then an immediate consequence of (B.6) and the uniform boundedness of $\mathbf{q}_i = \mathrm{vec}(\mathfrak{q}_i)$, $\boldsymbol{\nabla}_i^{(k+1)}$ and $\boldsymbol{\Delta}_i^{(k+1)}$. In fact, since we will assume (I3), i.e., $\lambda\|\boldsymbol{\nabla}_i\| \leq \frac{1}{2}$, it follows that $\|\boldsymbol{\Delta}_i^{(k+1)}\| \leq 2 \cdot \frac{1}{2}\|\boldsymbol{\nabla}_i\|$, which already proves the bound on $\|\boldsymbol{\Delta}_i\|$.

For bounding $\|\mathbf{z}_i^{(k)} - \mathbf{p}_i^{(k+1)}\|$ from above, we proceed as follows. During our derivations, we will assume some additional convenient constraints on certain quantities; they will be considered more carefully in the next subsection. Consider the coefficient of $\mathbf{a}_i^1$ in (B.5). First, a bound on the $v_{j,i}^{(k)}$'s is obtained as follows. Since $K_i \in \mathbb{F}^{(\Sigma+d)\times(d-1)}$ is a basis for the kernel of T_i it has linearly independent columns. Therefore, the pseudoinverse is $K_i^\dagger = (K_i^H K_i)^{-1} K_i^H$, and as $\boldsymbol{\nabla}_i^{(k)} \in \mathrm{span}(K_i)$ one has $\boldsymbol{\nabla}_i^{(k)} = K_i K_i^\dagger \boldsymbol{\nabla}_i^{(k)}$, because $K_i K_i^\dagger$ is a projector onto the column span of K_i; so $\mathbf{v}_i^{(k)} = K_i^\dagger \boldsymbol{\nabla}_i^{(k)}$. If $\lambda_i(A)$ denotes the ith largest eigenvalue of a Hermitian matrix $A \in \mathbb{F}^{m\times m}$, then it is well-known that $\|K_i^\dagger\|_2 = \big(\lambda_{d-1}(K_i^H K_i)\big)^{-\frac{1}{2}}$. One can verify that $K_i^H K_i = \mathrm{diag}(\|\mathbf{a}_i^2\|^2, \|\mathbf{a}_i^3\|^2, \ldots, \|\mathbf{a}_i^d\|^2) + \|\mathbf{a}_i^1\|^2 \mathbf{1}\mathbf{1}^T$, where $\mathbf{1}$ is a vector of length $d-1$ containing only ones. From [60, Section 2.41] it follows that the eigenvalues of $K_i^H K_i$ satisfy $\lambda_j(K_i^H K_i) \geq \lambda_j\big(\mathrm{diag}(\|\mathbf{a}_i^2\|^2, \|\mathbf{a}_i^3\|^2, \ldots, \|\mathbf{a}_i^d\|^2)\big)$, $j = 1, 2, \ldots, d-1$, so in particular

$$\|K_i^\dagger\|_2 \leq \Big(\min_{2\leq k\leq d} \|\mathbf{a}_i^k\|\Big)^{-1} \leq \Big(\min_{1\leq k\leq d} \|\mathbf{a}_i^k\|\Big)^{-1} =: \chi_i. \tag{B.7}$$

We can now bound the $(j-1)$th element $v_{j,i}^{(k)}$, $j = 2, 3, \ldots, d$, of $\mathbf{v}_i^{(k)}$ as follows:

$$|v_{j,i}^{(k)}| \leq \|\mathbf{v}_i^{(k)}\| \leq \|K_i^\dagger\|_2 \|\boldsymbol{\nabla}_i^{(k)}\| \leq \chi_i \|\boldsymbol{\nabla}_i^{(k)}\| \leq \chi_i \lambda^{k-1} \|\boldsymbol{\nabla}_i\|^k, \tag{B.8}$$

where the last step is due to the induction hypothesis (H1).

For convenience, we define $\epsilon_{j,i}^{(0)} = 0$ and

$$\epsilon_{j,i}^{(k)} := \sum_{\ell=1}^{k} v_{j,i}^{(\ell)}, \quad \text{so that} \quad |\epsilon_{j,i}^{(k)}| \leq \sum_{\ell=1}^{k} |v_{j,i}^{(\ell)}| \leq \chi_i \lambda^{-1} \sum_{\ell=1}^{\infty} (\lambda\|\boldsymbol{\nabla}_i\|)^\ell = \frac{\chi_i \|\boldsymbol{\nabla}_i\|}{1-\lambda\|\boldsymbol{\nabla}_i\|} =: C_i'; \tag{B.9}$$

the penultimate equality holds for $\lambda\|\boldsymbol{\nabla}_i\| < 1$, which will be satisfied by our assumptions. We additionally assume the following bound

$$C_i' \leq \frac{1}{2(d-1)}. \tag{I0}$$

Then, for $j = 2, 3, \ldots, d$ we find from (B.3) and (B.4) that

$$\gamma_{j,i}^{(k+1)} := \gamma_{j,i}^{(k)} - v_{j,i}^{(k)} = \cdots = \gamma_{j,i}^{(1)} - \sum_{\ell=1}^{k} v_{j,i}^{(\ell)} = 1 - \epsilon_{j,i}^{(k)}, \quad j = 2, 3, \ldots, d, \tag{B.10a}$$

where we used that this formula also applies by induction for $\gamma_{j,i}^{(\ell)}$ with $\ell = 1, 2, \ldots, k$. For $j = 1$ we have

$$\gamma_{1,i}^{(k+1)} = \prod_{j=2}^{d} \big(\gamma_{j,i}^{(k)} - v_{j,i}^{(k)}\big)^{-1} = \prod_{j=2}^{d} \big(1 - \epsilon_{j,i}^{(k-1)} - v_{j,i}^{(k)}\big)^{-1} = \prod_{j=2}^{d} \big(1 - \epsilon_{j,i}^{(k)}\big)^{-1}. \tag{B.10b}$$

With the above observations, the coefficient of $\mathbf{a}_i^1$ in (B.5) can be written as

$$\gamma_{1,i}^{(k)} + \sum_{j=2}^{d} v_{j,i}^{(k)} - \gamma_{1,i}^{(k+1)} = \prod_{j=2}^{d}\big(1-\epsilon_{j,i}^{(k-1)}\big)^{-1} + \sum_{j=2}^{d} v_{j,i}^{(k)} - \prod_{j=2}^{d}\big(1-\epsilon_{j,i}^{(k-1)} - v_{j,i}^{(k)}\big)^{-1}. \tag{B.11}$$

For proceeding, we will need some convoluted series expansions; however, the key idea revolves around expanding $(1-x)^{-1}$ with $x \approx 0$ by a Maclaurin series. Recall that

$$\prod_{j=1}^{n}(1-x_j)^{-1} = \prod_{j=1}^{n}\sum_{\kappa=0}^{\infty} x_j^{\kappa} = \sum_{\kappa=0}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=1}^{n} x_j^{\ell_j}, \tag{B.12}$$

where $\boldsymbol{\ell} = [\,\ell_1\ \ell_2\ \cdots\ \ell_n\,] \in \mathbb{N}^n$ and $\|\boldsymbol{\ell}\|_1 = \sum_{j=1}^{n}|\ell_j|$ is the 1-norm of $\boldsymbol{\ell}$; in the expression it is assumed that $|x_j|<1$ so that all expansions are absolutely convergent. The last term in (B.11) can be rewritten as follows:

$$\begin{aligned} Z := \prod_{j=2}^{d}\big(1-\epsilon_{j,i}^{(k-1)} - v_{j,i}^{(k)}\big)^{-1} &= \prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1}\prod_{j=2}^{d}\frac{1-\epsilon_{j,i}^{(k-1)}}{1-\epsilon_{j,i}^{(k-1)}-v_{j,i}^{(k)}} \\ &= \prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1}\prod_{j=2}^{d}\Big(1-\frac{v_{j,i}^{(k)}}{1-\epsilon_{j,i}^{(k-1)}}\Big)^{-1} \\ &= \prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1}\sum_{\kappa=0}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}\Big(\frac{v_{j,i}^{(k)}}{1-\epsilon_{j,i}^{(k-1)}}\Big)^{\ell_j}, \end{aligned}$$

where the last step was by (B.12), which requires that

$$|v_{j,i}^{(k)}| < |1-\epsilon_{j,i}^{(k-1)}|; \tag{I1}$$

this hypothesis will be investigated later. Let us write $Z = S_0 + S_1 + S_\infty$, where

$$\begin{aligned} S_0 &= \prod_{j=2}^{d}\big(1-\epsilon_{j,i}^{(k-1)}\big)^{-1} = \gamma_{1,i}^{(k)}, \qquad S_1 = \prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1}\cdot\sum_{j=2}^{d}\frac{v_{j,i}^{(k)}}{1-\epsilon_{j,i}^{(k-1)}}, \text{ and} \\ S_\infty &= \prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1}\cdot\sum_{\kappa=2}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}\Big(\frac{v_{j,i}^{(k)}}{1-\epsilon_{j,i}^{(k-1)}}\Big)^{\ell_j}. \end{aligned}$$

We will need to apply (B.12) again to S_1 for obtaining the desired result.

$$\begin{aligned} S_1 &= \sum_{j=2}^{d} v_{j,i}^{(k)}\Big(1+\frac{\epsilon_{j,i}^{(k-1)}}{1-\epsilon_{j,i}^{(k-1)}}\Big)\cdot\prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1} \\ &= \sum_{j=2}^{d} v_{j,i}^{(k)}\Big(1+\frac{\epsilon_{j,i}^{(k-1)}}{1-\epsilon_{j,i}^{(k-1)}}\Big)\cdot\Big(1+\sum_{\kappa=1}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{l=2}^{d}\big(\epsilon_{l,i}^{(k-1)}\big)^{\ell_l}\Big) \\ &= \underbrace{\sum_{j=2}^{d} v_{j,i}^{(k)}}_{S_1'} + \underbrace{\sum_{j=2}^{d} v_{j,i}^{(k)}\Big(\sum_{\kappa=1}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{l=2}^{d}\big(\epsilon_{l,i}^{(k-1)}\big)^{\ell_l}\Big)}_{S_1''} + \underbrace{\sum_{j=2}^{d} v_{j,i}^{(k)}\frac{\epsilon_{j,i}^{(k-1)}}{1-\epsilon_{j,i}^{(k-1)}}\cdot\prod_{\ell=2}^{d}\big(1-\epsilon_{\ell,i}^{(k-1)}\big)^{-1}}_{S_1'''}. \end{aligned}$$

Hence, (B.11) equals $S_0 + S_1' - Z = S_0 + S_1' - (S_0 + S_1' + S_1'' + S_1''' + S_\infty) = -(S_1'' + S_1''' + S_\infty)$. From (B.9), $1-\epsilon_{j,i}^{(k-1)} \geq 1 - |\epsilon_{j,i}^{(k-1)}| \geq 1 - C_i'$, and as $|\epsilon_{j,i}^{(k-1)}| \leq \frac{1}{2}$ by (I0), we get that

$$C_i := (1-C_i')^{-1} \geq \big(1-\epsilon_{j,i}^{(k-1)}\big)^{-1} = \big|1-\epsilon_{j,i}^{(k-1)}\big|^{-1};$$

this bound depends neither on j nor k. Consequently,

$$\Big|\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}\Big(\frac{v_{j,i}^{(k)}}{1-\epsilon_{j,i}^{(k-1)}}\Big)^{\ell_j}\Big| \le C_i^{\kappa}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}|v_{j,i}^{(k)}|^{\ell_j} \le C_i^{\kappa}\|\mathbf{v}_i^{(k)}\otimes\cdots\otimes\mathbf{v}_i^{(k)}\|_1 \le C_i^{\kappa}(d-1)^{\kappa/2}\|\mathbf{v}_i^{(k)}\|^{\kappa}, \tag{B.13}$$

where the second step is because every $\boldsymbol{\ell}$ with $\|\boldsymbol{\ell}\|_1=\kappa$ can be identified with

$$(\underbrace{1,\ldots,1}_{\ell_1},\underbrace{2,\ldots,2}_{\ell_2},\ldots,\underbrace{d-1,\ldots,d-1}_{\ell_{d-1}})$$

whose length is κ. From this identification, it follows that the symmetric part of $\mathbf{v}_i^{(k)}\otimes\cdots\otimes\mathbf{v}_i^{(k)}$ contains precisely all summands of $\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}|v_{j,i}^{(k)}|^{\ell_j}$. As a result, S_∞ can be bounded as follows

$$|S_\infty| \le C_i^{d-1}\sum_{\kappa=2}^{\infty}\big(C_i\sqrt{d-1}\|\mathbf{v}_i^{(k)}\|\big)^{\kappa} \le \frac{C_i^{d+1}(d-1)\chi_i^2\lambda^{2k-2}\|\boldsymbol{\nabla}_i\|^{2k}}{1-C_i\sqrt{d-1}\chi_i\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k}},$$

where (B.8) was used to bound $\|\mathbf{v}_i^{(k)}\|$ in the second inequality; herein, we assumed

$$C_i\sqrt{d-1}\chi_i\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k} \le \frac{1}{2}, \tag{I2}$$

so that the series is convergent. Analogously to the derivation of (B.13), one finds

$$\Big|\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}\big(\epsilon_{j,i}^{(k-1)}\big)^{\ell_j}\Big| \le \|\boldsymbol{\epsilon}_i^{(k-1)}\otimes\cdots\otimes\boldsymbol{\epsilon}_i^{(k-1)}\|_1 \le (d-1)^{\kappa/2}\|\boldsymbol{\epsilon}_i^{(k-1)}\|^{\kappa} \le \big((d-1)C_i'\big)^{\kappa}, \tag{B.14}$$

where $\boldsymbol{\epsilon}_i^{(k)}=\big[\,\epsilon_{2,i}^{(k-1)}\ \cdots\ \epsilon_{d,i}^{(k-1)}\,\big]$. The last step was due to (B.9). A bound for S_1'' is

$$|S_1''| \le \sum_{j=2}^{d}|v_{j,i}^{(k)}|\sum_{\kappa=1}^{\infty}\big((d-1)C_i'\big)^{\kappa} = \frac{(d-1)C_i'}{1-(d-1)C_i'}\sum_{j=2}^{d}|v_{j,i}^{(k)}| = \frac{(d-1)C_i'}{1-(d-1)C_i'}\|\mathbf{v}_i^{(k)}\|_1;$$

the Maclaurin series is convergent because of (I0). A bound for S_1''' is

$$|S_1'''| \le \sum_{j=2}^{d}|v_{j,i}^{(k)}|C_i'C_i\prod_{\ell=2}^{d}|1-\epsilon_{\ell,i,k-1}|^{-1} \le C_i'C_i^d\sum_{j=2}^{d}|v_{j,i}^{(k)}| = C_i'C_i^d\|\mathbf{v}_i^{(k)}\|_1.$$

It follows from the foregoing two bounds that

$$\begin{aligned}|S_1''|+|S_1'''| &\le \Big(\frac{(d-1)C_i'}{1-(d-1)C_i'}+C_i'C_i^d\Big)\|\mathbf{v}_i^{(k)}\|_1 \le C_i'\sqrt{d-1}\Big(\frac{d-1}{1-(d-1)C_i'}+C_i^d\Big)\chi_i\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k}\\ &\le \Big(\frac{d-1}{1-(d-1)C_i'}+C_i^d\Big)\frac{\chi_i^2\sqrt{d-1}}{1-\lambda\|\boldsymbol{\nabla}_i\|}\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k+1},\end{aligned}$$

where we used $\|\mathbf{v}_i^{(k)}\|_1\le\sqrt{d-1}\|\mathbf{v}_i^{(k)}\|$, (B.8), and (B.9). Let $\zeta=\frac{d-1}{1-(d-1)C_i'}+C_i^d$. Combining the above with the bound for S_∞, we get

$$|(B.11)| \le \chi_i^2\sqrt{d-1}\Big(\frac{\zeta}{1-\lambda\|\boldsymbol{\nabla}_i\|}+\frac{C_i^{d+1}\sqrt{d-1}\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k-1}}{1-C_i\sqrt{d-1}\chi_i\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k}}\Big)\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k+1}.$$

Hence, it suffices proving that the right hand side is less than $\|\mathbf{a}_i^1\|^{-1}\lambda^k\|\boldsymbol{\nabla}_i\|^{k+1}$. So, it suffices showing that

$$\|\mathbf{a}_i^1\|\chi_i^2\sqrt{d-1}\Big(\frac{\zeta}{1-\lambda\|\boldsymbol{\nabla}_i\|}+\frac{C_i^{d+1}\sqrt{d-1}\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^{k-1}}{1-C_i\sqrt{d-1}\chi_i\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^k}\Big)\leq\lambda. \tag{B.15}$$

Let us assume additionally that

$$\lambda\|\boldsymbol{\nabla}_i\|\leq\frac{1}{2}. \tag{I3}$$

Exploiting (I0), (I2), and (I3), we find that (B.15) is implied by the inequality

$$\|\mathbf{a}_i^1\|\chi_i^2\sqrt{d-1}\big(4(d-1)+2^{d+1}+2^{d-k+3}\sqrt{d-1}\big)\leq\lambda. \tag{I4}$$

Note that $\lambda>0$ is a free parameter, so it can simply be chosen so as to satisfy the above inequality. For instance, let us choose $\lambda=2^{d+3}(d-1)^{3/2}\cdot\max_{1\leq i\leq r}\big(\chi_i^2\|\mathbf{a}_i^1\|\big)$. If (I0), (I1), (I2), (I3) and (I4) are true, then this proves (H1) and (H0b).

Appendix B.3. Eliminating assumptions

We show that (B.1) implies (I0), (I1), (I2) and (I3). Assumption (I0) can be eliminated as follows:

$$(I0)\;\Leftrightarrow\;\frac{\|\boldsymbol{\nabla}_i\|}{1-\lambda\|\boldsymbol{\nabla}_i\|}\leq\frac{1}{2(d-1)}\chi_i^{-1}\;\Leftarrow\;\underbrace{\Big(\|\boldsymbol{\nabla}_i\|\leq\frac{1}{4(d-1)}\chi_i^{-1}\Big)\wedge(I3)}_{(I5)},$$

where $\wedge$ denotes conjunction. The last statement also implies (I1), because

$$\begin{aligned}(I1)\;\Leftarrow\;\chi_i\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^k<1-\frac{\chi_i\|\boldsymbol{\nabla}_i\|}{1-\lambda\|\boldsymbol{\nabla}_i\|}\;&\Leftarrow\;\Big(\chi_i\|\boldsymbol{\nabla}_i\|<1-\frac{\chi_i\|\boldsymbol{\nabla}_i\|}{1-\lambda\|\boldsymbol{\nabla}_i\|}\Big)\wedge(I3)\\&\Leftarrow\;\Big(\chi_i\|\boldsymbol{\nabla}_i\|\Big(1+\frac{1}{1-\lambda\|\boldsymbol{\nabla}_i\|}\Big)<1\Big)\wedge(I3)\\&\Leftarrow\;\big(\|\boldsymbol{\nabla}_i\|\leq\tfrac{1}{3}\chi_i^{-1}\big)\wedge(I3).\end{aligned}$$

The elimination of (I2) proceeds as follows:

$$(I2)\;\Leftrightarrow\;\lambda^{k-1}\|\boldsymbol{\nabla}_i\|^k\leq\frac{1}{2\sqrt{d-1}C_i}\chi_i^{-1}\;\Leftarrow\;\Big(\|\boldsymbol{\nabla}_i\|\leq\frac{1}{4\sqrt{d-1}}\chi_i^{-1}\Big)\wedge(I0)\wedge(I3)\Leftarrow(I5)$$

Hence, (I3), or, equivalently, $\|\boldsymbol{\nabla}_i\|\leq\frac{1}{2}\lambda^{-1}$ is the strongest bound, because we have

$$\|\boldsymbol{\nabla}_i\|\leq\frac{1}{2}\lambda^{-1}=\frac{1}{2^{d+2}\sqrt{d-1}}\cdot\frac{1}{4(d-1)}\min_{1\leq j\leq r}\big(\|\mathbf{a}_j^1\|^{-1}\chi_j^{-2}\big)<\frac{1}{4(d-1)}\chi_i^{-1}<\frac{1}{3}\chi_i^{-1},$$

where we exploited $\|\mathbf{a}_j^1\|^{-1}\chi_j^{-1}=\|\mathbf{a}_j^1\|^{-1}\min_{1\leq k\leq d}\|\mathbf{a}_j^k\|\leq\|\mathbf{a}_j^1\|\|\mathbf{a}_j^1\|^{-1}=1$. Thus, (I3) implies (I0), (I1), (I2), (I3), (I4), (I5), (H0a), (H0b), and (H1) for all $i=1,2,\ldots,r$ simultaneously, hereby concluding the main proof.

Appendix B.4. Alternative formulation

It is clear that taking

$$\boldsymbol{\Delta}^T=\begin{bmatrix}\boldsymbol{\Delta}_1^T&\cdots&\boldsymbol{\Delta}_r^T\end{bmatrix}\quad\text{and}\quad\dot{\mathbf{p}}^T=\begin{bmatrix}\mathrm{vec}(\dot{\mathfrak{p}}_1)^T&\cdots&\mathrm{vec}(\dot{\mathfrak{p}}_r)^T\end{bmatrix}$$

satisfies $\dot{\mathbf{p}}\sim\mathbf{p}$ and $\|\boldsymbol{\Delta}\|\leq\|\boldsymbol{\nabla}\|$. By assumption, the columns of K form a basis of the kernel of $T_{\mathbf{p}}$. Hence, the Moore–Penrose pseudoinverse of K is $K^\dagger=(K^HK)^{-1}K^H$. As $KK^\dagger$ is a projector onto the column span of K, the claim $\boldsymbol{\Delta}\in\mathrm{span}(K)^\perp$ is equivalent with $KK^\dagger\boldsymbol{\Delta}=0$. Then, $K^H\boldsymbol{\Delta}=\sum_{i=1}^r K_i^H\boldsymbol{\Delta}_i=0$.

Appendix B.5. Consequences for Terracini's matrix

Consider the definition of the representative $\mathfrak{p}_i^{(k)}$ in (B.3). The limit for $k \to \infty$ can be written as

$$\begin{aligned}\lim_{k\to\infty} \mathfrak{p}_i^{(k)} &= \Big(\mathbf{a}_i^1 \cdot \lim_{k\to\infty} \prod_{j=2}^{d} \big(1-\epsilon_{j,i}^{(k)}\big)^{-1}, \mathbf{a}_i^2 \cdot \lim_{k\to\infty} \big(1-\epsilon_{2,i}^{(k)}\big), \dots, \mathbf{a}_i^d \cdot \lim_{k\to\infty} \big(1-\epsilon_{d,i}^{(k)}\big)\Big) \\ &= (\gamma_{1,i}\mathbf{a}_i^1, \gamma_{2,i}\mathbf{a}_i^2, \dots, \gamma_{d,i}\mathbf{a}_i^d) = \dot{\mathfrak{p}}_i \end{aligned} \tag{B.16}$$

because of (B.10). Note that the series

$$\gamma_{j,i} = 1 - \epsilon_{j,i}^{(\infty)} = 1 - \sum_{k=1}^{\infty} v_{j,i}^{(k)} > 0, \quad j = 2,3,\dots,d,$$

are absolutely convergent because the terms in the infinite sequences can be bounded as in (B.8), and the sum $\chi_i\lambda^{-1}\sum_{k=1}^{\infty}(\lambda\|\boldsymbol{\nabla}\|)^k$ is absolutely convergent by (I3). Since

$$\gamma_{1,i} = (\gamma_{2,i}\gamma_{3,i}\cdots\gamma_{d,i})^{-1},$$

the above $\gamma_{j,i}$ yield the explicit expressions for the coefficients of $\dot{\mathfrak{p}}_i$. Notice that the above furnishes an alternative proof that the sequence of $\mathfrak{p}_i^{(k)}$'s converges. It follows immediately from the definition of Terracini's matrix in (6) that we can write

$$\dot{T}_i = T_i \cdot \operatorname{diag}\left(I_{n_1} \cdot \prod_{\substack{j=1\\ j\neq 1}}^{d} \gamma_{j,i},\ \dots,\ I_{n_d} \cdot \prod_{\substack{j=1\\ j\neq d}}^{d} \gamma_{j,i}\right) = T_i \cdot \operatorname{diag}\big(\gamma_{1,i}^{-1} I_{n_1}, \gamma_{2,i}^{-1} I_{n_2}, \dots, \gamma_{d,i}^{-1} I_{n_d}\big) = T_i \dot{D}_i,$$

so that Terracini's matrix $T_{\dot{\mathbf{p}}}$ in the points $\dot{\mathfrak{p}}_i$ is given by

$$T_{\dot{\mathbf{p}}} = \begin{bmatrix}\dot{T}_1 & \cdots & \dot{T}_r\end{bmatrix} = T_{\mathbf{p}} \cdot \operatorname{diag}(\dot{D}_1, \dot{D}_2, \dots, \dot{D}_r) = T_{\mathbf{p}}\dot{D}.$$

So it suffices demonstrating that $|\gamma_{j,i}^{-1} - 1| \le C\|\boldsymbol{\nabla}\|$. For $j = 2,3,\dots,d$, we have

$$|\gamma_{j,i}^{-1} - 1| = |(1-\epsilon_{j,i}^{(\infty)})^{-1} - 1| \le \sum_{\kappa=1}^{\infty} |\epsilon_{j,i}^{(\infty)}|^{\kappa} \le \sum_{\kappa=1}^{\infty} (2\chi_i\|\boldsymbol{\nabla}_i\|)^{\kappa} \le 4\|\boldsymbol{\nabla}\| \max_{1\le i\le r} \chi_i, \tag{B.17}$$

where in the second equality we used (B.9) and (I3), and where in the last step we used $2\chi_i\|\boldsymbol{\nabla}_i\| \le \frac{1}{2}$ which is true because of (I3). The case of $j = 1$ is due to

$$\begin{aligned}|\gamma_{1,i}^{-1} - 1| &= \Big|-1 + \prod_{j=2}^{d}\big(1-\epsilon_{j,i}^{(\infty)}\big)\Big| = \Big|-1+1+\sum_{\kappa=1}^{d-1}\sum_{\substack{\|\boldsymbol{\ell}\|_1=\kappa,\\ \boldsymbol{\ell}\le 1}} (-1)^{\kappa}\prod_{j=2}^{d}\big(\epsilon_{j,i}^{(\infty)}\big)^{\ell_j}\Big| \le \sum_{\kappa=1}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{j=2}^{d}|\epsilon_{j,i}^{(\infty)}|^{\ell_j} \\ &\le \sum_{\kappa=1}^{\infty}\big((d-1)C_i'\big)^{\kappa} = \frac{(d-1)C_i'}{1-(d-1)C_i'} \le 4(d-1)\|\boldsymbol{\nabla}\| \max_{1\le i\le r}\chi_i; \end{aligned} \tag{B.18}$$

herein, the inequality $\boldsymbol{\ell} \le 1$ is meant componentwise, the second inequality is by (B.14), and the last step used (B.9), (I0) and (I3). Letting $C = 4(d-1)\max_{1\le i\le r}\chi_i$ then concludes the proof.

Appendix C. Proof of the main theorem

Let $\mathcal{N}$ be a connected neighborhood of $\mathfrak{A}$ where all tensors are robustly r-identifiable. Let $\mathfrak{A}' \in \mathcal{N}$ be arbitrary. Then, there exist several $\Delta\mathbf{p}'$ such that $\mathfrak{A}' = f(\mathbf{p} + \Delta\mathbf{p}')$. The first step consists of enforcing a

unique choice. By r-identifiability, $f_{\mathcal{R}}^{-1}(\mathfrak{A}') = [\mathbf{p}+\Delta\mathbf{p}']$, so that the set of rank-1 tensors in the decomposition of $\mathfrak{A}'$ is unique. These rank-1 tensors can be ordered uniquely with respect to the lexicographic total order $\leq$ on $\mathbb{F}^{\Pi}$.[10] Let $\mathfrak{A}' = f(\mathbf{x})$, where $\mathbf{x} = \mathrm{vecr}(\mathfrak{q}_1, \ldots, \mathfrak{q}_r) \in \mathbb{F}^M$ with $\mathfrak{q}_i = (\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^d)$ is chosen so that the corresponding rank-1 tensors $\mathrm{Seg}(\mathfrak{q}_i)$ are sorted. The following set is uniquely defined regardless of $\mathbf{x}$:

$$\mathcal{C} = \left\{ j_{k,i} = \min\left\{ \arg\max_{1\leq \ell \leq n_k} |x_{\ell,i}^{(k)}| \right\} \mid k = 2, 3, \ldots, d,\ i = 1, 2, \ldots, r \right\}, \tag{C.1}$$

where $x_{\ell,i}^{(k)}$ is the ℓth element of $\mathbf{x}_i^k \in \mathbb{F}^{n_k}$. Note that $|x_{j_{k,i},i}^{(k)}| = \|\mathbf{x}_i^k\|_\infty > 0$ and that $|\mathcal{C}| = r(d-1)$. By r-identifiability there is just one choice of VFMs $\mathbf{p} + \Delta\mathbf{p} = \mathrm{vecr}(\mathfrak{q}_1', \ldots, \mathfrak{q}_r')$ with

$$\mathfrak{q}_i' = (\mathbf{b}_i^1, \ldots, \mathbf{b}_i^d),\ \mathrm{Seg}(\mathfrak{q}_1') < \cdots < \mathrm{Seg}(\mathfrak{q}_r'),\ \text{and } \forall j_{k,i} \in \mathcal{C} : b_{j_{k,i},i}^{(k)} = 1 \tag{C.2}$$

such that $\mathfrak{A}' = f(\mathbf{p} + \Delta\mathbf{p})$. The order of the rank-1 tensors is strict, for otherwise the rank of $\mathfrak{A}'$ would be strictly less than r. Given a fixed $\mathbf{p}$ and $\mathfrak{A}'$, $\Delta\mathbf{p}$ is uniquely determined by the foregoing procedure.

Next, we write $\mathrm{dist}(\mathbf{p}, f_{\mathcal{R}}^{-1}(\mathfrak{A}'))$ as a norm. Let $\mathfrak{A}' = \mathfrak{A} + \Delta\mathfrak{A} = f(\mathbf{p} + \Delta\mathbf{p})$. By r-identifiability, its inverse is $f_{\mathcal{R}}^{-1}(\mathfrak{A}') = [\mathbf{p} + \Delta\mathbf{p}]$. Let $\mathcal{T}$ be as in (A.1). Choose any

$$\widehat{S} \in \mathcal{Z} = \arg\min_{S\in\mathcal{T}} \|\mathbf{p} - S(\mathbf{p} + \Delta\mathbf{p})\|; \tag{C.3}$$

note that $\widehat{S}$ depends on $\mathfrak{A}'$, but for brevity such dependencies are not made explicit in the notation. We can write $\widehat{S}(\mathbf{p} + \Delta\mathbf{p}) = \mathbf{p} + \widetilde{\Delta\mathbf{p}}$, so that for every $\mathfrak{A}' \in \mathcal{N}$ the following holds:

$$\mathfrak{A}' = f(\mathbf{p} + \Delta\mathbf{p}) = f(\mathbf{p} + \widetilde{\Delta\mathbf{p}}).$$

By definition, $\|\widetilde{\Delta\mathbf{p}}\| = \mathrm{dist}(\mathbf{p}, f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A}))$, so we can just analyze this norm. It is important to note that while $\widetilde{\Delta\mathbf{p}}$ depends on the choice of $\widehat{S}$, the norm of this vector is independent of this choice.

In the remainder of the proof, most considered quantities, such as $\widetilde{\Delta\mathbf{p}}$, depend on $\Delta\mathfrak{A}$, which is arbitrary, and the choice of $\widehat{S}$ (which, in turn, depends on $\Delta\mathfrak{A}$). To be precise, one could indicate these dependencies in the notation by writing, e.g., $\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A}, \widehat{S}(\Delta\mathfrak{A}))$, but this will be avoided wherever no confusion can arise.

Appendix C.1. Continuity

For convenience, let

$$\mathcal{G}_\epsilon = \{\Delta\mathfrak{A} \in \mathbb{F}_0^{\Pi} \mid \|\Delta\mathfrak{A}\| \leq \epsilon \text{ and } \mathfrak{A} + \Delta\mathfrak{A} \in \sigma_r^0(\mathcal{S}_{\mathbb{F}})\}.$$

We prove that

$$\lim_{\epsilon\to 0} \max_{\Delta\mathfrak{A}\in\mathcal{G}_\epsilon} \|\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A})\| \to 0 \tag{C.4}$$

by showing that $\mathcal{N}$ is locally diffeomorphic to $\mathbb{F}^N$ and choosing a particularly suitable chart (U, ϕ) such that the Euclidean distance on the chart provides an upper bound on the foregoing expression.

Let $\mathcal{C}$ be defined as in (C.1) but replacing $x_{\ell,i}^{(k)}$ by $a_{\ell,i}^{(k)}$, where the latter is the ℓth element of $\mathbf{a}_i^k$. Let $C = [a_{j_{k,i},i}^{(k)}]_{k=2,i=1}^{d,r} \in \mathbb{F}^{(d-1)\times r}$, where the $j_{k,i} \in \mathcal{C}$. Let $\widehat{U}_{\mathcal{C},C}$, $f_{\mathcal{C},C}$, $\pi_{\mathcal{C}}$, $\pi_{\mathcal{C},C}^{-1}$, and $\eta_{\mathcal{C},C}$ be defined as in Section 3, substituting the general $\mathcal{C}$ and C for the aforementioned choices. Then by definition $\mathfrak{A}$ lies in the image of $f_{\mathcal{C},C}$; in fact the construction is such that $\mathfrak{A} = f_{\mathcal{C},C}(\overline{\mathbf{p}})$ where $\overline{\mathbf{p}} = \pi_{\mathcal{C}}(\eta_{\mathcal{C},C}(\mathbf{p})) = \pi_{\mathcal{C}}(\mathbf{p})$. I claim that $f_{\mathcal{C},C}$ is a local diffeomorphism. By the definition of robust r-identifiability, $\mathcal{N}$ is an open neighborhood of a smooth $\mathbb{F}$-manifold of dimension N. Hence, there exists a chart (U, ϕ) where $U \subset \mathcal{N}$ is an open

[10] For $\mathbb{F} = \mathbb{C} \simeq \mathbb{R}^2$, one should consider the lexicographic order after the identification $\mathbb{C}^{\Pi} = \mathbb{R}^{2\Pi}$.

neighborhood of $\mathfrak{A}$ and a local diffeomorphism ϕ between U and an open neighborhood $V = \phi(U) \subset \mathbb{F}^N$. Restrict $\mathcal{N}$ to the open neighborhood $U \subset \mathcal{N}$. Let $\mathfrak{A} = \phi^{-1}(\mathbf{q}) \in \mathcal{N}$. Since ϕ is a local diffeomorphism, $\phi \circ \phi^{-1} = \mathrm{Id}_{\mathbb{F}^N}$. Hence, the Jacobian matrix at $\mathbf{q} \in V$ of the composition $\phi \circ \phi^{-1}$, namely $J_{\phi}(\phi^{-1}(\mathbf{q}))J_{\phi^{-1}}(\mathbf{q})$, is the $N \times N$ identity matrix. Since $J_{\phi^{-1}}(\mathbf{q}) \in \mathbb{F}^{\Pi \times N}$ and $J_{\phi}(\phi^{-1}(\mathbf{q})) \in \mathbb{F}^{N \times \Pi}$ with $\Pi \geq N$, it follows that both matrices are of maximal rank N. By definition, $J_{\phi^{-1}}(\mathbf{q})$ is contained in the tangent space $\mathrm{T}_{\mathfrak{A}}\sigma_r(\mathcal{S}_{\mathbb{F}})$. Let $\overline{J}_{\overline{\mathbf{p}}} \in \mathbb{F}^{\Pi \times N}$ denote the Jacobian matrix of $f_{\mathcal{C},C}$ at $\overline{\mathbf{p}}$. By the assumption on robust r-identifiability and (10), $\mathrm{span}(\overline{J}_{\overline{\mathbf{p}}}) = \mathrm{span}(T_{\mathbf{p}}) = \mathrm{T}_{\mathfrak{A}}\sigma_r(\mathcal{S}_{\mathbb{F}})$. Since $\dim \mathrm{span}\big(J_{\phi^{-1}}(\mathbf{q})\big) = N = \dim \mathrm{span}(\overline{J}_{\overline{\mathbf{p}}})$ it follows that $J_{\phi^{-1}}(\mathbf{q})Z = \overline{J}_{\overline{\mathbf{p}}}$ for an invertible matrix $Z \in \mathbb{F}^{N \times N}$. Hence, the Jacobian matrix of the composition $g : \phi \circ f_{\mathcal{C},C}$ at $\overline{\mathbf{p}}$ is given by $J_{\phi}(f_{\mathcal{C},C}(\overline{\mathbf{p}}))\overline{J}_{\overline{\mathbf{p}}} = J_{\phi}(\phi^{-1}(\mathbf{q}))J_{\phi^{-1}}(\mathbf{q})Z = Z$, where the first equality is because $\mathfrak{A} = f_{\mathcal{C},C}(\overline{\mathbf{p}}) = \phi^{-1}(\mathbf{q})$. Now, the smooth function $g : \mathbb{F}^N \to \mathbb{F}^N$ has nonsingular Jacobian matrix Z, hence by the Inverse Function Theorem [24, Theorem 1A.1] it has a smooth inverse function g^{-1}. Thus, g is a local diffeomorphism between a neighborhood of $\overline{\mathbf{p}} \in \mathbb{F}^N$ and a neighborhood of $\mathbf{q} \in \mathbb{F}^N$. Letting now $f_{\mathcal{C},C}^{-1}$ be the composition $g^{-1} \circ \phi$, which is a composition of local diffeomorphisms, it follows that $f_{\mathcal{C},C}$ is a local diffeomorphism between an open neighborhood of $\overline{\mathbf{p}} \in \mathbb{F}^N$ and an open neighborhood of $\mathfrak{A} \in \mathcal{N}$. Further restrict $\mathcal{N}$ to the neighborhood where this local diffeomorphism is defined.

For an arbitrary choice of $\mathfrak{A}' = f(\mathbf{p} + \Delta\mathbf{p}) \in \mathcal{N}$ there exists an $\overline{S} \in \mathcal{T}$ such that

$$\mathfrak{A}' = \mathfrak{A} + \Delta\mathfrak{A} = f(\mathbf{p} + \Delta\mathbf{p}) = f_{\mathcal{C},C}(\pi_{\mathcal{C}}(\eta_{\mathcal{C},C}(\mathbf{p} + \Delta\mathbf{p}))) =: f_{\mathcal{C},C}\big(\pi_{\mathcal{C}}\big(\overline{S}(\mathbf{p} + \Delta\mathbf{p})\big)\big).$$

Let $\overline{\Delta\mathbf{p}} := \pi_{\mathcal{C}}(\overline{S}(\mathbf{p} + \Delta\mathbf{p})) - \overline{\mathbf{p}}$. As $f_{\mathcal{C},C}^{-1}$ is a local diffeomorphism, it follows that $\Delta\mathfrak{A} \to 0$ implies $\overline{\Delta\mathbf{p}} \to 0$. Both $\mathbf{p} = \pi_{\mathcal{C},C}^{-1}(\overline{\mathbf{p}})$ and $\overline{S}(\mathbf{p} + \Delta\mathbf{p}) = \pi_{\mathcal{C},C}^{-1}(\overline{\mathbf{p}} + \overline{\Delta\mathbf{p}})$ live in $\widehat{U}_{\mathcal{C},C}$, so that their values at all positions $j_{k,i} \in \mathcal{C}$ agree. As $\pi_{\mathfrak{A}}^{-1}(\mathbf{x})$ is analytic, $\overline{\Delta\mathbf{p}} \to 0$ implies that $(\overline{S}(\mathbf{p} + \Delta\mathbf{p}) - \mathbf{p}) \to 0$ as well. By definition, $\|\widetilde{\Delta\mathbf{p}}\| = \min_{S \in \mathcal{T}} \|\mathbf{p} - S(\mathbf{p} + \Delta\mathbf{p})\| \leq \|\mathbf{p} - \overline{S}(\mathbf{p} + \Delta\mathbf{p})\|$, (C.4) follows.

Appendix C.2. Sandwiching

It remains to demonstrate that the next equalities hold:

$$\lim_{\epsilon \to 0} \max_{\Delta\mathfrak{A} \in \mathcal{G}_\epsilon} \frac{\|\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A})\|}{\|\Delta\mathfrak{A}\|} = \lim_{\epsilon \to 0} \max_{\Delta\mathfrak{A} \in \mathcal{G}_\epsilon} \frac{\|\mathbf{x}(\Delta\mathfrak{A})\|}{\|T_{\mathbf{p}}\mathbf{x}(\Delta\mathfrak{A})\|} = \big(\varsigma_N(T_{\mathbf{p}})\big)^{-1},$$

where $\mathbf{x}(\Delta\mathfrak{A}) \in \mathbb{F}^M$. The key consists of exploiting the Iterated Scaling Lemma for transforming any $\widetilde{\Delta\mathbf{p}}$ into a new vector $\mathbf{x}(\Delta\mathfrak{A})$ that is orthogonal to $T_{\mathbf{p}}$'s kernel.

Since Terracini's matrix $T_{\mathbf{p}}$ has a kernel $\mathcal{K} = \mathrm{span}(K)$ of dimension exactly equal to $r(d-1)$, we can factorize $\widetilde{\Delta\mathbf{p}}$ as $\widetilde{\Delta\mathbf{p}} = \boldsymbol{\Delta} + \boldsymbol{\nabla}$, where $\boldsymbol{\Delta} \in \mathcal{K}^{\perp}$ and $\boldsymbol{\nabla} \in \mathcal{K}$. Note that $\langle \boldsymbol{\nabla}, \boldsymbol{\Delta} \rangle = 0$, so that $\|\boldsymbol{\nabla}\| \leq \|\widetilde{\Delta\mathbf{p}}\|$ and $\|\boldsymbol{\Delta}\| \leq \|\widetilde{\Delta\mathbf{p}}\|$. From Lemma 22 it follows that for small $\|\Delta\mathfrak{A}\| > 0$, and, hence, small $\|\widetilde{\Delta\mathbf{p}}\| > 0$, we have

$$\mathbf{p} + \boldsymbol{\nabla} = \dot{\mathbf{p}} + \boldsymbol{\Delta}' \quad \text{with } \dot{\mathbf{p}} \sim \mathbf{p},\ \boldsymbol{\Delta}' \in \mathcal{K}^{\perp}, \text{ and } \|\boldsymbol{\Delta}'\| \leq \|\boldsymbol{\nabla}\|.$$

Let $\dot{\boldsymbol{\Delta}} = \boldsymbol{\Delta}' + \boldsymbol{\Delta} \in \mathcal{K}^{\perp}$. As particular consequences, we have that $f(\mathbf{p}) = f(\dot{\mathbf{p}})$ and $f(\mathbf{p} + \widetilde{\Delta\mathbf{p}}) = f(\dot{\mathbf{p}} + \dot{\boldsymbol{\Delta}})$. Observe that f is an analytic multivariate polynomial that is homogeneous of degree d in M variables, so that it has a finite, convergent Taylor series expansion at every point. Since $T_{\dot{\mathbf{p}}} = T_{\mathbf{p}} + T_{\mathbf{p}}\dot{E}$ by Lemma 22, the expansion about the VFMs $\dot{\mathbf{p}}$ is

$$f(\mathbf{p} + \widetilde{\Delta\mathbf{p}}) = f(\dot{\mathbf{p}} + \dot{\boldsymbol{\Delta}}) = f(\dot{\mathbf{p}}) + T_{\dot{\mathbf{p}}}\dot{\boldsymbol{\Delta}} + \mathcal{O}(\|\dot{\boldsymbol{\Delta}}\|^2) = f(\mathbf{p}) + T_{\mathbf{p}}\dot{\boldsymbol{\Delta}} + T_{\mathbf{p}}\dot{E}\dot{\boldsymbol{\Delta}} + \mathcal{O}(\|\dot{\boldsymbol{\Delta}}\|^2).$$

The following bound is then obtained:

$$\big|\|f(\mathbf{p} + \widetilde{\Delta\mathbf{p}}) - f(\mathbf{p})\| - \|T_{\mathbf{p}}\dot{\boldsymbol{\Delta}}\|\big| \leq \|T_{\mathbf{p}}\dot{E}\dot{\boldsymbol{\Delta}}\|_2 + \mathcal{O}(\|\dot{\boldsymbol{\Delta}}\|^2) = \mathcal{O}(\|\dot{\boldsymbol{\Delta}}\|\|\widetilde{\Delta\mathbf{p}}\|),$$

where the last step invoked $\|T_{\mathbf{p}}\dot{E}\dot{\boldsymbol{\Delta}}\|_2 \leq \|T_{\mathbf{p}}\|_2\|\dot{E}\|_2\|\dot{\boldsymbol{\Delta}}\|$ and $\|\dot{E}\|_2 \leq \|\boldsymbol{\nabla}\| \leq \|\widetilde{\Delta\mathbf{p}}\|$. It follows from the Iterated Scaling Lemma that if $f(\mathbf{p} + \widetilde{\Delta\mathbf{p}}) \neq f(\mathbf{p})$ then $\|\dot{\boldsymbol{\Delta}}\| \neq 0$. Indeed, the contrapositive $\|\dot{\boldsymbol{\Delta}}\| = 0$ yields $f(\mathbf{p} + \widetilde{\Delta\mathbf{p}}) = f(\dot{\mathbf{p}}) = f(\mathbf{p})$. Thus, if $\|\Delta\mathfrak{A}\| > 0$ then $\|\dot{\boldsymbol{\Delta}}\| > 0$, so that

$$\frac{\|\Delta\mathfrak{A}\|}{\|\dot{\boldsymbol{\Delta}}\|} = \frac{\|T_{\mathbf{p}}\dot{\boldsymbol{\Delta}}\|}{\|\dot{\boldsymbol{\Delta}}\|} + \varepsilon(\Delta\mathfrak{A}) \text{ with } \lim_{\Delta\mathfrak{A} \to 0} |\varepsilon(\Delta\mathfrak{A})| \to 0; \tag{C.5}$$

the limit is zero because of (C.4). It remains to relate this expression to $\|\widetilde{\Delta\mathbf{p}}\|/\|\Delta\mathfrak{A}\|$.

An upper bound on $\|\widetilde{\Delta\mathbf{p}}\|$ in terms of $\dot{\mathbf{\Delta}}$ is obtained by observing that $\widehat{S}(\mathbf{p}+\Delta\mathbf{p}) = \mathbf{p}+\widetilde{\Delta\mathbf{p}} = \dot{\mathbf{p}}+\dot{\mathbf{\Delta}} =: S\mathbf{p}+\dot{\mathbf{\Delta}}$ for some invertible $S \in \mathcal{T}$. It follows that

$$\|\widetilde{\Delta\mathbf{p}}\| = \min_{D\in\mathcal{T}} \|D(\mathbf{p}+\Delta\mathbf{p})-\mathbf{p}\| \le \|S^{-1}\widehat{S}(\mathbf{p}+\Delta\mathbf{p})-\mathbf{p}\| = \|S^{-1}\dot{\mathbf{\Delta}}\| \le \|S^{-1}\|_2\|\dot{\mathbf{\Delta}}\|,$$

where the first inequality arises because $\mathcal{T}$ is a multiplicative group. As $S^{-1} \in \mathcal{T}$ so $S \in \mathcal{T}$, we can write $S = PD$ where $P \in \mathcal{P}$ and $D \in \mathcal{B}$, so that $\|S^{-1}\|_2 = \|D^{-1}\|_2$. It follows from the Iterated Scaling Lemma that the diagonal entries of D are the $\gamma_{j,i}$'s in (B.16). Then, it follows immediately from (B.17) and (B.18) that the largest diagonal entry in absolute value of D^{-1} is smaller than $1+4(d-1)\|\widetilde{\Delta\mathbf{p}}\|\cdot\max_i \chi_i =: 1+\delta$, where χ_i is as in (B.7). It follows from Appendix C.1 that $\delta \to 0$ as $\epsilon \to 0$. A lower bound on $\|\widetilde{\Delta\mathbf{p}}\|$ in terms of $\dot{\mathbf{\Delta}}$ is obtained as follows. From the triangle inequality, $\|\dot{\mathbf{\Delta}}\| \le \|\mathbf{\Delta}'\| + \|\mathbf{\Delta}\|$ so that

$$\begin{aligned}\|\dot{\mathbf{\Delta}}\|^2 &\le \|\mathbf{\Delta}'\|^2 + 2\|\mathbf{\Delta}'\|\|\mathbf{\Delta}\| + \|\mathbf{\Delta}\|^2 \\ &= \|\mathbf{\Delta}'\|(\|\mathbf{\Delta}'\| + 2\|\mathbf{\Delta}\|) + \|\mathbf{\Delta}\|^2 \le \|\boldsymbol{\nabla}\|^2(4\lambda^2\|\boldsymbol{\nabla}\|^2 + 4\lambda\|\mathbf{\Delta}\|) + \|\mathbf{\Delta}\|^2,\end{aligned}$$

where the last step is due to Lemma 22. Note that λ is a constant. Provided that $\|\Delta\mathfrak{A}\|$ and, hence, $\|\widetilde{\Delta\mathbf{p}}\|$, are sufficiently small, i.e., so that the inequality

$$4\lambda^2\|\boldsymbol{\nabla}\|^2 + 4\lambda\|\mathbf{\Delta}\| \le 4\lambda^2\|\widetilde{\Delta\mathbf{p}}\|^2 + 4\lambda\|\widetilde{\Delta\mathbf{p}}\| \le 1$$

is satisfied, then the foregoing implies the lower bound

$$\|\dot{\mathbf{\Delta}}\|^2 \le \|\boldsymbol{\nabla}\|^2 + \|\mathbf{\Delta}\|^2 = \|\boldsymbol{\nabla}\|^2 + 2\cdot\mathrm{Re}\langle\boldsymbol{\nabla},\mathbf{\Delta}\rangle + \|\mathbf{\Delta}\|^2 = \|\widetilde{\Delta\mathbf{p}}\|^2,$$

where the first equality is because of the orthogonality in the Hermitian inner product of $\boldsymbol{\nabla}$ and $\mathbf{\Delta}$. Thus, by combining the lower and upper bound and dividing by $\|\Delta\mathfrak{A}\|$, we can conclude that for every non-zero $\Delta\mathfrak{A}$ of sufficiently small norm the following relations hold true

$$\frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A},\widehat{S}(\Delta\mathfrak{A}))\|}{\|\Delta\mathfrak{A}\|} \le \frac{\|\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A})\|}{\|\Delta\mathfrak{A}\|} \le (1+\delta(\Delta\mathfrak{A}))\frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A},\widehat{S}(\Delta\mathfrak{A}))\|}{\|\Delta\mathfrak{A}\|}, \tag{C.6}$$

provided that $\epsilon \ge \|\Delta\mathfrak{A}\|$ is sufficiently small and where the dependence on $\Delta\mathfrak{A}$ and $\widehat{S}(\Delta\mathfrak{A})$ was explicitly indicated. Let

$$\Delta\mathfrak{A}' \in \operatorname*{arg\,max}_{\Delta\mathfrak{A}\in\mathcal{G}_\epsilon} \max_{\widehat{S}(\Delta\mathfrak{A})\in\mathcal{Z}} \frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A},\widehat{S}(\Delta\mathfrak{A}))\|}{\|\Delta\mathfrak{A}\|},\ \Delta\mathfrak{A}'' \in \operatorname*{arg\,max}_{\Delta\mathfrak{A}\in\mathcal{G}_\epsilon} \frac{\|\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A})\|}{\|\Delta\mathfrak{A}\|},\ \widehat{S}' \in \operatorname*{arg\,max}_{\widehat{S}\in\mathcal{Z}} \frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A}',\widehat{S})\|}{\|\Delta\mathfrak{A}'\|},$$

where $\mathcal{Z}$ is as in (C.3). It follows that

$$\begin{aligned}\frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A}',\widehat{S}')\|}{\|\Delta\mathfrak{A}'\|} \le \frac{\|\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A}')\|}{\|\Delta\mathfrak{A}'\|} \le \frac{\|\widetilde{\Delta\mathbf{p}}(\Delta\mathfrak{A}'')\|}{\|\Delta\mathfrak{A}''\|} &\le (1+\delta(\Delta\mathfrak{A}''))\frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A}'',\bullet)\|}{\|\Delta\mathfrak{A}''\|} \\ &\le \frac{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A}',\widehat{S}')\|}{\|\Delta\mathfrak{A}'\|}\cdot\max_{\Delta\mathfrak{A}\in\mathcal{G}_\epsilon}(1+\delta(\Delta\mathfrak{A})),\end{aligned} \tag{C.7}$$

where the first step is by the first inequality in (C.6), the second inequality is by optimality of $\Delta\mathfrak{A}''$, the third inequality is because of the second inequality in (C.6) with the "$\bullet$" indicating that the inequality is valid for all $S(\Delta\mathfrak{A}'') \in \mathcal{Z}$ (so in particular the maximum), and the last step is because of the optimality of $\Delta\mathfrak{A}'$. For continuing, the explicit notation indicating the dependence on $\Delta\mathfrak{A}$ and $\widehat{S}$ is discarded again. Let n be sufficiently large, and then consider the sequences

$$c_n = \min_{\Delta\mathfrak{A}_n\in\mathcal{G}_{1/n}} \min_{\widehat{S}(\Delta\mathfrak{A}_n)\in\mathcal{Z}} \frac{\|\Delta\mathfrak{A}_n\|}{\|\dot{\mathbf{\Delta}}_n\|} = \frac{\|\Delta\mathfrak{A}'_n\|}{\|\dot{\mathbf{\Delta}}'_n\|} \text{ and } s_n = \min_{\Delta\mathfrak{A}_n\in\mathcal{G}_{1/n}} \min_{\widehat{S}(\Delta\mathfrak{A}_n)\in\mathcal{Z}} \frac{\|T_{\mathbf{p}}\dot{\mathbf{\Delta}}_n\|}{\|\dot{\mathbf{\Delta}}_n\|},$$

where $\dot{\mathbf{\Delta}}_n' := \dot{\mathbf{\Delta}}(\Delta\mathfrak{A}_n', \widehat{S}'(\Delta\mathfrak{A}_n'))$. Then it follows from the first part of (C.5) that

$$c_n \geq s_n + \min_{\Delta\mathfrak{A}_n \in \mathcal{G}_{1/n}} \min_{\widehat{S}(\Delta\mathfrak{A}_n)\in\mathcal{Z}} \varepsilon(\Delta\mathfrak{A}_n) \quad \text{and} \quad s_n \geq c_n + \min_{\Delta\mathfrak{A}_n \in \mathcal{G}_{1/n}} \min_{\widehat{S}(\Delta\mathfrak{A}_n)\in\mathcal{Z}} (-\varepsilon(\Delta\mathfrak{A}_n)).$$

From the above and the second part of (C.5), it follows there exist $\vartheta_n \geq 0$ so that we can bound

$$s_n - \vartheta_n \leq c_n \leq s_n + \vartheta_n; \text{ hence,} (s_n - \vartheta_n)^{-1} \geq c_n^{-1} \geq (s_n + \vartheta_n)^{-1}, \tag{C.8}$$

for sufficiently large n so that $s_n > \vartheta_n$. Note that $\vartheta_n \to 0$ as $n \to \infty$ because of (C.5), (C.4) and Lemma 22. Since $\dot{\mathbf{\Delta}}_n \in \mathcal{K}^\perp$, it follows immediately from the Courant–Fisher min-max characterization of the least singular value [30, Theorem 4.2.11] that

$$0 < \varsigma_N(T_{\mathbf{p}}) \leq \frac{\|T_{\mathbf{p}}\dot{\mathbf{\Delta}}_n\|}{\|\dot{\mathbf{\Delta}}_n\|} \leq \varsigma_1(T_{\mathbf{p}}),$$

where $N = r(\Sigma + 1)$, $\varsigma_i(T_{\mathbf{p}})$ denotes the ith singular value of $T_{\mathbf{p}}$, and the first inequality is due to the assumption on the rank of $T_{\mathbf{p}}$. I claim that the sequence of s_n converges:

$$\lim_{n\to\infty} s_n = \lim_{\epsilon\to 0} \min_{\Delta\mathfrak{A}\in\mathcal{G}_\epsilon} \min_{\widehat{S}(\Delta\mathfrak{A})\in\mathcal{Z}} \frac{\|T_{\mathbf{p}}\dot{\mathbf{\Delta}}(\Delta\mathfrak{A}, \widehat{S}(\Delta\mathfrak{A}))\|}{\|\dot{\mathbf{\Delta}}(\Delta\mathfrak{A}, \widehat{S}(\Delta\mathfrak{A})\|} = \varsigma_N(T_{\mathbf{p}}). \tag{C.9}$$

This would conclude the proof because then all of the following limits are well-defined:

$$\lim_{n\to\infty} (s_n - \vartheta_n)^{-1} = \big(\varsigma_N(T_{\mathbf{p}})\big)^{-1} \quad \text{and} \quad \lim_{n\to\infty} (s_n + \vartheta_n)^{-1} = \big(\varsigma_N(T_{\mathbf{p}})\big)^{-1},$$

sandwiching c_n^{-1} in (C.8), so that by considering (C.7) we find

$$\lim_{n\to\infty} c_n^{-1} \leq \lim_{n\to\infty} \max_{\Delta\mathfrak{A}_n\in\mathcal{G}_{1/n}} \frac{\|\widetilde{\Delta\mathbf{p}}_n\|}{\|\Delta\mathfrak{A}_n\|} \leq \lim_{n\to\infty} c_n^{-1} \cdot \lim_{n\to\infty} \max_{\Delta\mathfrak{A}_n\in\mathcal{G}_{1/n}} (1 + \delta_n),$$

and as $\delta_n \to 0$, it follows that

$$\kappa_A = \lim_{\epsilon\to 0} \max_{\Delta\mathfrak{A}\in\mathcal{G}_\epsilon} \frac{\operatorname{dist}(\mathbf{p}, f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A}))}{\|\Delta\mathfrak{A}\|} = \lim_{n\to\infty} \max_{\Delta\mathfrak{A}_n\in\mathcal{G}_{1/n}} \frac{\|\widetilde{\Delta\mathbf{p}}_n\|}{\|\Delta\mathfrak{A}_n\|} = \big(\varsigma_N(T_{\mathbf{p}})\big)^{-1},$$

which would conclude the proof.

Appendix C.3. The limit

For proving (C.9), we consider the specific perturbation $\Delta\mathbf{p}_n' = n^{-1}\mathbf{w}$, where $\mathbf{w}$ is the right singular vector of $T_{\mathbf{p}}$ corresponding to the singular value $\varsigma_N(T_{\mathbf{p}})$. The perturbed tensor is $\mathfrak{A}_n = f(\mathbf{p}+\Delta\mathbf{p}_n') = f(\mathbf{p}+n^{-1}\mathbf{w}) = \mathfrak{A} + \Delta\mathfrak{A}_n$. If n_0 is sufficiently large, then $\mathfrak{A}_n \subset \mathcal{N}$ for all $n > n_0$. Clearly, $\|\Delta\mathfrak{A}_n\| \to 0$ as $n \to \infty$. From a Taylor series expansion, it is also clear that for large n, $\Delta\mathfrak{A}_n \neq 0$, as the norms of the higher-order terms will be dominated by $n^{-1} = \|n^{-1}\mathbf{w}\|$. Let $\Delta\mathbf{p}_n$ be the specific normalization of $\mathbf{p} + \Delta\mathbf{p}_n \in f_{\mathcal{R}}^{-1}(\mathfrak{A} + \Delta\mathfrak{A}_n)$ that satisfies the properties imposed in (C.2). Let $\widehat{S}_n \in \arg\min_{S\in\mathcal{T}} \|\mathbf{p} - S(\mathbf{p} + \Delta\mathbf{p}_n)\|$ be arbitrary. Then, $\mathbf{p}+\widetilde{\Delta\mathbf{p}}_n = \widehat{S}_n(\mathbf{p}+\Delta\mathbf{p}_n)$. By definition of $\Delta\mathbf{p}_n$ there exists some $T_n \in \mathcal{T}$ such that $\mathbf{p}+\Delta\mathbf{p}_n = T_n(\mathbf{p}+\Delta\mathbf{p}_n')$. Hence, $\mathbf{p} + \widetilde{\Delta\mathbf{p}}_n = \widehat{S}_n T_n(\mathbf{p} + n^{-1}\mathbf{w})$, so that

$$\begin{aligned} \|\widetilde{\Delta\mathbf{p}}_n\| = \min_{S\in\mathcal{T}} \|\mathbf{p} - S(\mathbf{p} + \Delta\mathbf{p}_n)\| &= \min_{S\in\mathcal{T}} \|\mathbf{p} - ST_n(\mathbf{p} + n^{-1}\mathbf{w})\| \\ &= \min_{S\in\mathcal{T}} \|\mathbf{p} - S(\mathbf{p} + n^{-1}\mathbf{w})\| \leq \|\mathbf{p} - I(\mathbf{p} + n^{-1}\mathbf{w})\| = n^{-1}; \end{aligned} \tag{C.10}$$

the third equality is because $\mathcal{T}$ is a group.

Let n be sufficiently large. Then, $\widehat{S}_n T_n = I + E_n$ where E_n is a diagonal matrix whose diagonal entries are bounded by $\mathcal{O}(n^{-1})$. Indeed, the proof of Proposition 9 shows that (C.10) can only be true if for all $k = 2, \dots, d$ and all $i = 1, \dots, r$ simultaneously we have that

$$\|\mathbf{a}_i^k - \theta_{k,\pi_i}(n)(\mathbf{a}_{\pi_i}^k + n^{-1}\mathbf{w}_{\pi_i}^k)\| \le n^{-1}, \quad \theta_{k,\pi_i}(n) \in \mathbb{F}_0, \tag{C.11}$$

where $\mathbf{w} = \mathrm{vecr}(\mathfrak{w}_1, \dots, \mathfrak{w}_r)$, $\mathfrak{w}_i = (\mathbf{w}_i^1, \dots, \mathbf{w}_i^d)$, and π is a permutation of $\{1, \dots, r\}$. Since the above equation is valid for all all n and as $\mathbf{a}_i^k$ are constant vectors, it follows that there exist subsequences such that all $\theta_{k,\pi_i}(n)$ converge: $\theta_{k,\pi_i}(n) \to \widehat{\theta}_{k,\pi_i}$. In the limit of this subsequence, the inequality can only be satisfied if $\mathbf{a}_i^k = \widehat{\theta}_{k,\pi_i}\mathbf{a}_{\pi_i}^k$ for all $k = 2, 3, \dots, d$ and $i = 1, 2, \dots, r$. Suppose that π is not the identity, then there exists a $j \neq \pi_j$ for which it holds that

$$\begin{aligned}\mathbf{a}_j^1 \otimes \cdots \otimes \mathbf{a}_j^d + \mathbf{a}_{\pi_j}^1 \otimes \cdots \otimes \mathbf{a}_{\pi_j}^d &= \mathbf{a}_j^1 \otimes \cdots \otimes \mathbf{a}_j^d + \mathbf{a}_{\pi_j}^1 \otimes (\widehat{\theta}_{2,\pi_j}^{-1}\mathbf{a}_j^2) \otimes \cdots \otimes (\widehat{\theta}_{d,\pi_j}^{-1}\mathbf{a}_j^d) \\ &= \big(\mathbf{a}_j^1 + \widehat{\theta}_{2,\pi_j}^{-1} \cdots \widehat{\theta}_{d,\pi_j}^{-1}\mathbf{a}_{\pi_j}^1\big) \otimes \mathbf{a}_j^2 \otimes \cdots \otimes \mathbf{a}_j^d,\end{aligned}$$

showing that $\mathfrak{A}$ really has a decomposition of length at most $r-1$, which is a contradiction. Thus, on every convergent subsequence, for large n, π is the identity. Hence, for large n, (C.11) simplifies to

$$\|(1 - \theta_{k,i}(n))\mathbf{a}_i^k - n^{-1}\theta_{k,i}(n)\mathbf{w}_i^k\|^2 \le n^{-2}, \tag{C.12}$$

which should hold for all $i = 1, 2, \dots, r$ and $k = 2, 3, \dots, d$. It is shown next that this inequality may be satisfied only if

$$1 - 3n^{-1}\|\mathbf{a}_i^k\|^{-1} \le \theta_{k,i}(n) \le 1 + 3n^{-1}\|\mathbf{a}_i^k\|^{-1}. \tag{C.13}$$

In the remainder of the proof, the dependence of the $\theta_{k,i}$'s on n will no longer be indicated. Note that (C.12) is a quadratic equation in $\theta_{k,i}$ with positive coefficient for $\theta_{k,i}^2$. So it suffices proving that inequality (C.12) is satisfied for $\theta_{k,i} = 1$ while it is no longer satisfied for $\theta_{k,i} = 1 \pm 3n^{-1}\|\mathbf{a}_i^k\|^{-1}$ to conclude that any $\theta_{k,i}$ satisfying (C.12) must be contained in the interval (C.13). For $\theta_{k,i} = 1$ one has $\|(1-1)\mathbf{a}_i^k - n^{-1}\mathbf{w}_i^k\| = n^{-1}\|\mathbf{w}_i^k\| \le n^{-1}$, so (C.12) is satisfied. Pugging $\theta_{k,i} = 1 \pm 3n^{-1}\|\mathbf{a}_i^k\|^{-1}$ into (C.12), we get from the triangle inequality

$$\begin{aligned}\|(1 - \theta_{k,i})\mathbf{a}_i^k - n^{-1}\theta_{k,i}\mathbf{w}_i^k\| &\ge 3n^{-1}\|\mathbf{a}_i^k\|^{-1}\|\mathbf{a}_i^k\| - n^{-1}|1 \pm 3n^{-1}\|\mathbf{a}_i^k\|^{-1}| \cdot \|\mathbf{w}_i^k\| \\ &\ge n^{-1}(3 - |1 \pm 3n^{-1}\|\mathbf{a}_i^k\|^{-1}|) > n^{-1},\end{aligned}$$

where the last inequality is valid for $n > 3\|\mathbf{a}_i^k\|^{-1}$. Hence, $|\theta_{k,i} - 1| \le 3n^{-1}\|\mathbf{a}_i^k\|^{-1}$ for $k = 2, \dots, d$. For $\theta_{1,i} = (\theta_{2,i} \cdots \theta_{d,i})^{-1}$ we get

$$\begin{aligned}|\theta_{1,i} - 1| = \left|-1 + \prod_{k=2}^{d}(1 \pm 3n^{-1}\|\mathbf{a}_i^k\|^{-1})^{-1}\right| &\le -1 + 1 + \sum_{\kappa=1}^{\infty}\sum_{\|\boldsymbol{\ell}\|_1=\kappa}\prod_{k=2}^{d}\Big|3n^{-1}\|\mathbf{a}_i^k\|^{-1}\Big|^{\ell_k} \\ \le \sum_{\kappa=1}^{\infty}(3n^{-1}\chi_i)^{\kappa}\|\mathbf{1} \otimes \cdots \otimes \mathbf{1}\|_1 &\le \sum_{\kappa=1}^{\infty}(3(d-1)n^{-1}\chi_i)^{\kappa} \le 6(d-1)n^{-1}\chi_i,\end{aligned}$$

where χ_i is as in (B.7); in the third step the same argument surrounding the derivation of (B.13) was employed and in the last step we assumed $n \ge 6(d-1)\chi_i$. Let $\mu_{k,i}$ be defined as

$$\theta_{k,i} =: 1 - \mu_{k,i}, \qquad k = 2, 3, \dots, d,\ i = 1, 2, \dots, r \tag{C.14a}$$

$$\theta_{1,i} = \theta_{2,i}^{-1} \cdots \theta_{d,i}^{-1} =: 1 - \mu_{1,i}, \qquad i = 1, 2, \dots, r. \tag{C.14b}$$

By the foregoing derivation $|\mu_{k,i}| = \mathcal{O}(n^{-1})$. We can then write E_n explicitly as

$$E_n = \mathrm{diag}(E_{n,1}, \dots, E_{n,r}) \quad \text{with } E_{n,i} = \mathrm{diag}(-\mu_{1,i}I_{n_1}, -\mu_{2,i}I_{n_2}, \dots, -\mu_{d,i}I_{n_d}).$$

We now consider the following factorization

$$\begin{aligned}\mathbf{p}+\widetilde{\Delta\mathbf{p}}_n = \widehat{S}_n(\mathbf{p}+\Delta\mathbf{p}_n) &= \widehat{S}_n T_n(\mathbf{p}+\Delta\mathbf{p}'_n) \\ &= (I+E_n)(\mathbf{p}+\Delta\mathbf{p}'_n) = \mathbf{p}+(\boldsymbol{\nabla}_{n,1}+\boldsymbol{\nabla}_{n,2})+(\boldsymbol{\Delta}_{n,1}+\boldsymbol{\Delta}_{n,2}+\boldsymbol{\Delta}\mathbf{p}'_n)\end{aligned} \tag{C.15}$$

where $E_n\mathbf{p}=\boldsymbol{\nabla}_{n,1}+\boldsymbol{\Delta}_{n,1}$ and $E_n\Delta\mathbf{p}'_n=\boldsymbol{\nabla}_{n,2}+\boldsymbol{\Delta}_{n,2}$ with $\boldsymbol{\nabla}_{n,i}\in\mathcal{K}$ and $\boldsymbol{\Delta}_{n,i}\in\mathcal{K}^\perp$. One sees immediately that $\|\boldsymbol{\nabla}_{n,1}\|\le\|E_n\|_2\|\mathbf{p}\|=\mathcal{O}(n^{-1})$, $\|\boldsymbol{\nabla}_{n,2}\|\le\|E_n\|_2\|\Delta\mathbf{p}'_n\|=\mathcal{O}(n^{-2})$, and $\|\boldsymbol{\Delta}_{n,2}\|\le\|E_n\|_2\|\Delta\mathbf{p}'_n\|=\mathcal{O}(n^{-2})$. The key difficulty lies in proving that $\|\boldsymbol{\Delta}_{n,1}\|=\mathcal{O}(n^{-2})$. By definition,

$$\boldsymbol{\Delta}_{n,1}=(I-KK^\dagger)E_n\mathbf{p}=(I-K\,\mathrm{diag}(K_1^HK_1,\dots,K_r^HK_r)^{-1}K^H)E_n\mathbf{p}$$

so it suffices to prove that $\|(I-K_iK_i^\dagger)\mathbf{q}_i\|=\mathcal{O}(n^{-2})$, where $\mathbf{q}_i=\mathbf{q}_i^{(1)}=E_{n,i}\,\mathrm{vec}(\mathfrak{p}_i)$ with $\mu_{k,i}$ as in (C.14). By definition and some Maclaurin series, we know that

$$-\mu_{1,i}=-1+\prod_{k=2}^{d}(1-\mu_{k,i})^{-1}=\sum_{k=2}^{d}\mu_{k,i}+\sum_{\kappa=2}\sum_{\|\ell\|_1=\kappa}\prod_{k=2}^{d}\mu_{k,i}^{\ell_k}=:\sum_{k=2}^{d}\mu_{k,i}+M_i; \tag{C.16}$$

it can be shown in the usual way that $|M_i|=\mathcal{O}(n^{-2})$. Let $\alpha_{k,i}=\|\mathbf{a}_i^k\|$. Then, we find $\mathbf{q}_i^{(2)}=K_i^H\mathbf{q}_i^{(1)}$. The entries of $\mathbf{q}_i^{(2)}$ are $q_{k-1,i}^{(2)}=-\mu_{1,i}\alpha_{1,i}^2+\mu_{k,i}\alpha_{k,i}^2$, $k=2,\dots,d$. For the next step $\mathbf{q}_i^{(3)}=(K_i^HK_i)^{-1}\mathbf{q}_i^{(2)}$, we recall from the Sherman–Morrison formula that

$$(K_i^HK_i)^{-1}=(\mathrm{diag}(\alpha_{2,i}^2,\dots,\alpha_{d,i}^2)+\alpha_{1,i}^2\mathbf{1}\mathbf{1}^T)^{-1}=\mathrm{diag}(\alpha_{2,i}^{-2},\dots,\alpha_{d,i}^{-2})-\frac{\alpha_{1,i}^2\mathbf{h}_i\mathbf{h}_i^T}{1+\alpha_{1,i}^2\sum_{j=2}^{d}\alpha_{j,i}^{-2}},$$

where $\mathbf{h}_i^T=[\,\alpha_{2,i}^{-2}\ \cdots\ \alpha_{d,i}^{-2}\,]$. Let $\widehat{\alpha}_i=\alpha_{1,i}^2\sum_{j=2}^{d}\alpha_{j,i}^{-2}$. Then, $\mathbf{h}_i^T\mathbf{q}_i^{(2)}=-\mu_{1,i}\widehat{\alpha}_i+\sum_{k=2}^{d}\mu_{k,i}$, so that the elements of $\mathbf{q}_i^{(3)}$ are given by

$$\begin{aligned}q_{k-1,i}^{(3)} &= -\mu_{1,i}\alpha_{1,i}^2\alpha_{k,i}^{-2}+\mu_{k,i}-\frac{\alpha_{1,i}^2\alpha_{k,i}^{-2}(-\mu_{1,i}\widehat{\alpha}_i+\sum_{k=2}^{d}\mu_{k,i})}{1+\widehat{\alpha}_i} \\ &= -\mu_{1,i}\alpha_{1,i}^2\alpha_{k,i}^{-2}+\mu_{k,i}-\frac{\alpha_{1,i}^2\alpha_{k,i}^{-2}(1+\widehat{\alpha}_i)\sum_{k=2}^{d}\mu_{k,i}+\alpha_{1,i}^2\alpha_{k,i}^{-2}\widehat{\alpha}_iM_i}{1+\widehat{\alpha}_i} \\ &= \mu_{k,i}+\alpha_{1,i}^2\alpha_{k,i}^{-2}M_i-\alpha_{1,i}^2\alpha_{k,i}^{-2}\widehat{\alpha}_i(1+\widehat{\alpha}_i)^{-1}M_i \\ &= \mu_{k,i}+M_i\alpha_{1,i}^2\alpha_{k,i}^{-2}(1-\widehat{\alpha}_i(1+\widehat{\alpha}_i)^{-1})=:\mu_{k,i}+M_{i,k},\end{aligned}$$

where in the second and third step (C.16) was used. Now $\mathbf{q}_i^{(4)}=\mathbf{q}_i^{(1)}-K_i\mathbf{q}_i^{(3)}=(I-K_iK_i^\dagger)\mathbf{q}_i$ where

$$\mathbf{q}_i^{(4)}=\begin{bmatrix}((-\mu_{1,i}-\sum_{k=2}^{d}(\mu_{k,i}+M_{i,k}))\mathbf{a}_i^1\\(-\mu_{2,i}+\mu_{2,i}+M_{i,2})\mathbf{a}_i^2\\\vdots\\(-\mu_{d,i}+\mu_{d,i}+M_{i,d})\mathbf{a}_i^d\end{bmatrix}=\begin{bmatrix}(M_i-\sum_{k=2}^{d}M_{i,k})\mathbf{a}_i^1\\M_{i,2}\mathbf{a}_i^2\\\vdots\\M_{i,d}\mathbf{a}_i^d\end{bmatrix},$$

from which $\|\boldsymbol{\Delta}_{n,1}\|=\mathcal{O}(n^{-2})$ follows.

Continuing from (C.15), we can apply the Iterated Scaling Lemma to $\mathbf{p}+\boldsymbol{\nabla}_n$, where $\boldsymbol{\nabla}_n=\boldsymbol{\nabla}_{n,1}+\boldsymbol{\nabla}_{n,2}$ and $\|\boldsymbol{\nabla}_n\|=\mathcal{O}(n^{-1})$, thus obtaining

$$\mathbf{p}+\widetilde{\Delta\mathbf{p}}_n=\dot{\mathbf{p}}_n+\boldsymbol{\Delta}'_n+\boldsymbol{\Delta}_{n,1}+\boldsymbol{\Delta}_{n,2}+\Delta\mathbf{p}'_n=\dot{\mathbf{p}}_n+n^{-1}\mathbf{w}+n^{-2}\mathbf{x}_n,$$

where $\boldsymbol{\Delta}'_n\in\mathcal{K}^\perp$ is of norm $\mathcal{O}(n^{-2})$ for sufficiently large n, and $\mathbf{x}_n$ is a vector whose norm can be uniformly bounded by a constant C. By definition, $\dot{\boldsymbol{\Delta}}_n=n^{-1}\mathbf{w}+n^{-2}\mathbf{x}_n$ and $T_{\mathbf{p}}\mathbf{w}=\varsigma_N\mathbf{w}$. From

$$\frac{\varsigma_N-n^{-1}\|T_{\mathbf{p}}\|_2\|\mathbf{x}_n\|}{1+n^{-1}\|\mathbf{x}_n\|}\le\frac{n^{-1}\|T_{\mathbf{p}}(\mathbf{w}+n^{-1}\mathbf{x}_n)\|}{n^{-1}\|\mathbf{w}+n^{-1}\mathbf{x}_n\|}\le\frac{\varsigma_N+n^{-1}\|T_{\mathbf{p}}\|_2\|\mathbf{x}_n\|}{1-n^{-1}\|\mathbf{x}_n\|},$$

it follows that $\|T_{\mathbf{p}}\dot{\boldsymbol{\Delta}}(\Delta\mathfrak{A}_n,\bullet)\|/\|\dot{\boldsymbol{\Delta}}(\Delta\mathfrak{A}_n,\bullet))\|$, where "$\bullet$" indicates any $\widehat{S}(\Delta\mathfrak{A}_n)\in\mathcal{Z}$, tends to $\varsigma_N(T_{\mathbf{p}})$ as $n\to\infty$, concluding the proof.

References

[1] Abo, H., Ottaviani, G., Peterson, C., 2009. Induction for secant varieties of Segre varieties. Trans. Amer. Math. Soc. 361, 767–792.

[2] Acar, E., Dunlavy, D., Kolda, T., Mørup, M., 2011. Scalable tensor factorizations for incomplete data. Chemometr. Intell. Lab. 106 (1), 41–56.

[3] Allman, E. S., Matias, C., Rhodes, J. A., 2009. Identifiability of parameters in latent structure models with many observed variables. Ann. Stat. 37 (6A), 3099–3132.

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

[5] Appellof, C., Davidson, E., 1981. Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents. Anal. Chem. 53 (13), 2053–2056.

[6] Bernardi, A., Brachat, J., Comon, P., Mourrain, B., 2013. General tensor decomposition, moment matrices and applications. J. Symbolic Comput. 52, 51–71.

[7] Bhaskara, A., Charikar, M., Vijayaraghavan, A., 2014. Uniqueness of tensor decompositions with applications to polynomial identifiability. JMLR: Workshop and Conference Proceedings 35, 1–37.

[8] Blum, L., Cucker, F., Shub, M., Smale, S., 1998. Complexity and real computation. Springer-Verlag, New York.

[9] Bocci, C., Chiantini, L., Ottaviani, G., 2014. Refined methods for the identifiability of tensors. Ann. Mat. Pura Appl. (4) 193 (6), 1691–1702.

[10] Bochnak, J., Coste, M., Roy, M., 1998. Real Algebraic Geometry. Springer–Verlag.

[11] Boralevi, A., Draisma, J., Horobeţ, E., Robeva, E., 2017. Orthogonal and unitary tensor decomposition from an algebraic perspective. Isr. J. Math.To appear.

[12] Breiding, P., Vannieuwenhoven, N., 2016. The condition number of join decompositions. arXiv:1611.08117.

[13] Bürgisser, P., Cucker, F., 2013. Condition: The Geometry of Numerical Algorithms. Vol. 349 of Grundlehren der mathematischen Wissenschaften. Springer-Verlag.

[14] Chen, J., Saad, Y., 2009. On the tensor SVD and the optimal low rank orthogonal approximation of tensors. SIAM J. Matrix Anal. Appl. 30 (4), 1709–1734.

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

[16] Chiantini, L., Ottaviani, G., Vannieuwenhoven, N., 2014. An algorithm for generic and low-rank specific identifiability of complex tensors. SIAM J. Matrix Anal. Appl. 35 (4), 1265–1287.

[17] Chiantini, L., Ottaviani, G., Vannieuwenhoven, N., 2017. Effective criteria for specific identifiability of tensors and forms. SIAM J. Matrix Anal. Appl.Accepted.

[18] Comon, P., 1994. Independent component analysis, a new concept? Signal Proc. 36 (3), 287–314.

[19] De Lathauwer, L., 2006. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. SIAM J. Matrix Anal. Appl. 28 (3), 642–666.

[20] de Silva, V., Lim, L.-H., 2008. Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. 30 (3), 1084–1127.

[21] Demmel, J. W., 1987. The geometry of ill-conditioning. J. Complexity 3 (2), 201–229.

[22] Domanov, I., De Lathauwer, L., 2014. Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition. SIAM J. Matrix Anal. Appl. 35 (2), 636–660.

[23] Domanov, I., De Lathauwer, L., 2015. Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL. SIAM J. Matrix Anal. Appl. 36 (4), 1567–1589.

[24] Dontchev, A. L., Rockafellar, R. T., 2009. Implicit Function and Solution Mappings: A View from Variational Analysis. Springer Monographs in Mathematics. Springer.

[25] Griffiths, P., Harris, J., 1978. Principles of Algebraic Geometry. John Wiley & Sons, Inc.

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

[27] Hayashi, C., Hayashi, F., 1982. A new algorithm to solve PARAFAC-model. Behaviormetrika 11, 49–60.

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

[29] Hitchcock, F., 1927. The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. 6 (1), 164–189.

[30] Horn, R., Johnson, C., 1986. Matrix Analysis. Cambridge University Press, New York, NY, USA.

[31] Kolda, T. G., 2001. Orthogonal tensor decompositions. SIAM J. Matrix Anal. Appl. 23 (1), 243–255.

[32] Koldovský, Z., Tichavský, P., Phan, A.-H., 2011. Stability analysis and fast damped Gauss–Newton algorithm for INDSCAL tensor decomposition. In: Proc. IEEE Workshop Statist. Signal Process. pp. 581–584.

[33] Krantz, S. G., Parks, H. R., 2002. The Implicit Function Theorem: History, Theory, and Applications. Birkhäuser Boston, New York.

[34] Kruskal, J. B., 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Appl. 18 (2), 95–138.

[35] Landsberg, J., 2012. Tensors: Geometry and Applications. Vol. 128 of Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island.

[36] Leurgans, S. E., Ross, R. T., Abel, R. B., 1993. A decomposition for three-way arrays. SIAM J. Matrix Anal. Appl. 14 (4), 1064–1083.

[37] Liu, X., Sidiropoulos, N. D., 2001. Cramér–Rao lower bounds for low-rank decomposition of multidimensional arrays. IEEE Trans. Signal Process. 49 (9), 2074–2086.

[38] Nocedal, J., Wright, S. J., 2006. Numerical Optimization, 2nd Edition. Springer Series in Operations Research and Financial Engineering. Springer.
[39] Oseledets, I., Savost'yanov, D., 2006. Minimization methods for approximating tensors and their comparison. Comp. Math. Math. Phys+ 46 (10), 1641–1650.
[40] Paatero, P., 1997. A weighted non-negative least squares algorithm for three-way PARAFAC factor analysis. Chemometr. Intell. Lab. 38 (2), 223–242.
[41] Phan, A.-H., Tichavský, P., Cichocki, A., 2013. Fast alternating LS algorithms for high order CANDECOMP/PARAFAC tensor factorizations. IEEE Trans. Signal Process. 61 (19), 4834–4846.
[42] Phan, A.-H., Tichavský, P., Cichocki, A., 2013. Low complexity damped Gauss–Newton algorithms for CANDECOMP/-PARAFAC. SIAM J. Matrix Anal. Appl. 34 (1), 126–147.
[43] Qi, Y., Comon, P., Lim, L.-H., 2016. Semialgebraic Geometry of Nonnegative Tensor Rank. SIAM J. Matrix Anal. Appl. 37 (4), 1556–1580.
[44] Robeva, E., 2016. Orthogonal decomposition of symmetric tensors. SIAM J. Matrix Anal. Appl. 37 (1), 86–102.
[45] Rotman, J. J., 1994. An Introduction to the Theory of Groups, 4th Edition. Vol. 148 of Graduate Texts in Mathematics. Springer–Verlag, New York, USA.
[46] Sorber, L., Van Barel, M., De Lathauwer, L., 2013. Optimization-based algorithms for tensor decompositions: canonical polyadic decomposition, decomposition in rank-$(l_r, l_r, 1)$ terms, and a new generalization. SIAM J. Optim. 23 (2), 695–720.
[47] Spivak, M., 1965. Calcalus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calcalus. Addison-Wesley.
[48] Stewart, G. W., Sun, J.-G., 1990. Matrix Perturbation Theory. Vol. 33 of Computer Science and Scientific Computing. Academic Press.
[49] Strassen, V., 1983. Rank and optimal computation of generic tensors. Linear Algebra Appl. 52–53, 645–685.
[50] Terracini, A., 1911. Sulla V_k per cui la varietà degli S_h $h+1$-secanti ha dimensione minore dell'ordinario. Rend. Circ. Mat. Palermo (2) 31, 392–396.
[51] Tichavský, P., Koldovský, Z., 2011. Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources. IEEE Trans. Signal Process. 59 (3), 1037–1047.
[52] Tichavský, P., Phan, A.-H., Koldovský, Z., 2013. Cramér–Rao-induced bounds for CANDECOMP/PARAFAC tensor decomposition. IEEE Trans. Signal Process. 61 (8), 1986–1997.
[53] Tomasi, G., 2006. Practical and computational aspects in chemometric data analysis. Ph.D. thesis, The Royal Veterinary and Agricultural University, Frederiksberg, Denmark.
[54] Tomasi, G., Bro, R., 2005. PARAFAC and missing values. Chemometr. Intell. Lab. 75, 163–180.
[55] Vannieuwenhoven, N., 2016. A condition number for the tensor rank decomposition. arXiv:1604.00052, 45 pp.
[56] Vannieuwenhoven, N., Nicaise, J., Vandebril, R., Meerbergen, K., 2014. On generic nonexistence of the Schmidt–Eckart–Young decomposition for complex tensors. SIAM J. Matrix Anal. Appl. 35 (3), 886–903.
[57] Vannieuwenhoven, N., Vandebril, R., Meerbergen, K., 2012. A new truncation strategy for the higher-order singular value decomposition. SIAM J. Sci. Comput. 34 (2), A1027–A1052.
[58] Vannieuwenhoven, N., Vandebril, R., Meerbergen, K., 2015. A randomized algorithm for testing nonsingularity of structured matrices with an application to asserting nondefectivity of Segre varieties. IMA J. Numer. Anal. 35 (1), 289–324.
[59] Vervliet, N., Debals, O., Sorber, L., Van Barel, M., De Lathauwer, L., 2014. Tensorlab v3.0. `http://www.tensorlab.net` (last accessed 18 January 2017).
[60] Wilkinson, J. H., 1965. The Algebraic Eigenvalue Problem. Oxford University Press, Amen House, London E.C. 4, United Kingdom.
[61] Zhang, T., Golub, G. H., 2001. Rank-one approximation to high order tensors. SIAM J. Matrix Anal. Appl. 23 (2), 534–550.