Tensor decompositions and their sensitivity

Nick Vannieuwenhoven

FWO

KU LEUVEN

January 30, 2019

Fluorophores are fluorescent molecules with the property that they re-emit light (emission) when they are excited by light.

Application: Fluorescence spectroscopy

The main application of **fluorescence spectroscopy** is determining the **type** and **concentration** of fluorophores in a mixture.

Some of its applications[1] include:

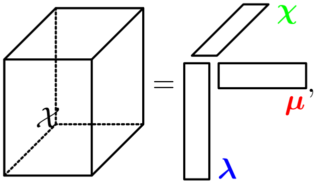
- study of biomolecules (e.g., study of cell dynamics);
- analysis of dissolved organic materials in waste and polluted water (e.g., identifying pollutants);
- food chemistry (e.g., quality assessment).

[1] See resp. the reviews by Weiss (1999) and Moerner and Fromm (2003); Hudson, Baker and Reynolds (2007); Smilde, Bro and Geladi (2005).

The mathematical model

The **intensity** $x_{i,j,k}$ of the light that is emitted at wavelength ω_j when a fluorophore, diluted in water with concentration c_k, is excited at wavelength ω_i is

$$x_{i,j,k} = \lambda_i \mu_j \chi_k$$

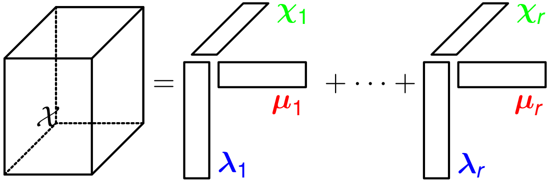


where

- λ_i is the fraction of light absorbed at wavelength ω_i,
- μ_j is the fraction of light emitted at wavelength ω_j, and
- χ_k is a constant proportional to the concentration c_k.

The mathematical model

When r fluorophores occur jointly in different concentrations in several diluted mixtures, the model becomes a **tensor rank decomposition**:

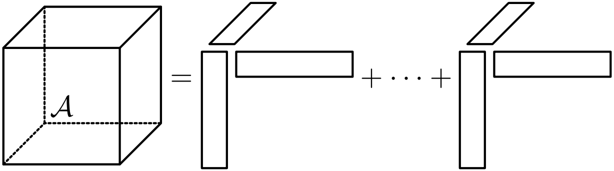


Alternative names: CANDECOMP, PARAFAC, CP decomposition, canonical polyadic decomposition (**CPD**), separated representation . . .

Overview

Hitchcock (1927) introduced the **tensor rank decomposition**:

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^d$$



The **rank** of a tensor is the minimum number of rank-1 tensors of which it is a linear combination.

If the set of rank-1 tensors $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$ is the unique set such that $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r$, then we call $\mathcal{A}$ an ***r*-identifiable** tensor.

Matrices are never r-identifiable, because

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

for every invertible X. For a general choice of X this results in different rank-r factorizations.

Kruskal (1977) gave a famous **sufficient condition** for proving the r-identifiability of a third-order tensor

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i;$$

if the **Kruskal ranks** k_A, k_B, and k_C of respectively $\{\mathbf{a}_i\}$, $\{\mathbf{b}_i\}$, and $\{\mathbf{c}_i\}$ satisfy

$$1 < k_A, k_B, k_C \text{ and } r \leq \frac{1}{2}(k_A + k_B + k_C - 2)$$

then $\mathcal{A}$ is r-identifiable.

The Kruskal rank of a set of vectors $\{\mathbf{x}_i\}$ is the largest number k_X such that every subset of k_X vectors is linearly independent.

For example, a sufficient condition for the tensor

$$\mathcal{A} = \sum_{i=1}^{n} \mathbf{a}_i \otimes \mathbf{b}_i \otimes \mathbf{c}_i, \quad \mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i \in \mathbb{R}^n,$$

to admit only this factorization is that $\{\mathbf{a}_i\}_i$, $\{\mathbf{b}_i\}_i$, and $\{\mathbf{c}_i\}_i$ are all linearly independent sets.

$n_1 \times \cdots \times n_d$ tensors are called **generically r-identifiable** if
the set of rank-r tensors that are not r-identifiable
is contained in a strict subvariety of
the smallest irreducible variety that contains all rank-r tensors.

It is **conjectured**[2] that if $n_1 \geq \cdots \geq n_d \geq 2$,

$$r_{\mathrm{cr}} = \frac{n_1 \cdots n_d}{1 + \sum_{k=1}^{d}(n_k - 1)}, \quad \text{and} \quad r_{\mathrm{ub}} = n_2 \cdots n_d - \sum_{k=2}^{d}(n_k - 1),$$

then the **general rule** is:

if $r \geq r_{\mathrm{cr}}$ or $d = 2$	$\rightarrow$	not generically r-identifiable
if $n_1 > r_{\mathrm{ub}}$ and $r \geq r_{\mathrm{ub}}$	$\rightarrow$	not generically r-identifiable
if none of foregoing and $r < r_{\mathrm{cr}}$	$\rightarrow$	generically r-identifiable

[2] See Bocci, Chiantini, Ottaviani (2014) and Chiantini, Ottaviani, V (2014)

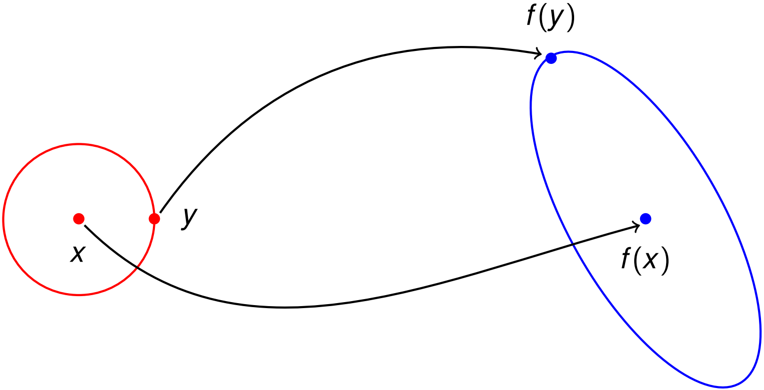
Overview

Sensitivity

In numerical computations, the **sensitivity** of the output of a computation to **perturbations** at the input is very important, because representation and roundoff errors will corrupt any mathematical inputs.

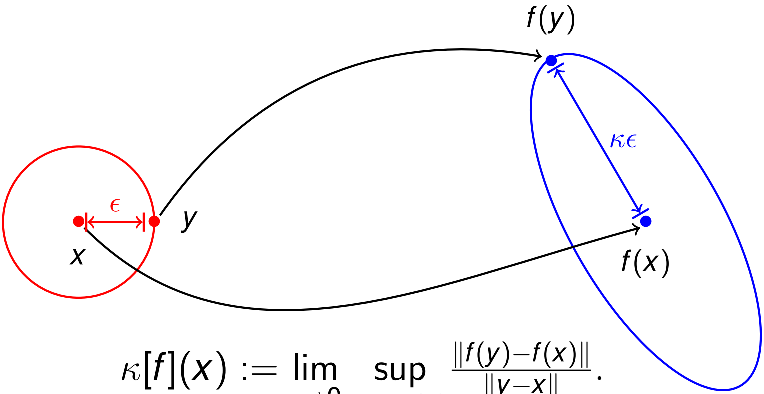
Condition numbers

The **condition number** quantifies the **worst-case sensitivity** of f to perturbations of the input.



Condition numbers

The **condition number** quantifies the **worst-case sensitivity** of f to perturbations of the input.



$$\kappa[f](x) := \lim_{\epsilon \to 0} \sup_{y \in B_\epsilon(x)} \frac{\|f(y) - f(x)\|}{\|y - x\|}.$$

If $f : \mathbb{F}^m \supset X \to Y \subset \mathbb{F}^n$ is a differentiable function, then the condition number is fully determined by the first-order approximation of f.

Indeed, in this case we have

$$f(\mathbf{x} + \mathbf{\Delta}) = f(\mathbf{x}) + J\mathbf{\Delta} + o(\|\mathbf{\Delta}\|),$$

where J is the **Jacobian matrix** containing all first-order partial derivatives. Then,

$$\kappa = \lim_{\epsilon \to 0} \sup_{\|\mathbf{\Delta}\| \le \epsilon} \frac{\|f(\mathbf{x}) + J\mathbf{\Delta} + o(\|\mathbf{\Delta}\|) - f(\mathbf{x})\|}{\|\mathbf{\Delta}\|}$$
$$= \max_{\|\mathbf{\Delta}\|=1} \frac{\|J\mathbf{\Delta}\|}{\|\mathbf{\Delta}\|} = \|J\|_2.$$

The tensor decomposition problem

The condition number of **computing the rank-1 terms in a CPD** was investigated only recently by Breiding and V (2018). I discuss our strategy from Beltrán, Breiding, and V (2019b).

To compute the condition number, we study the **addition map**

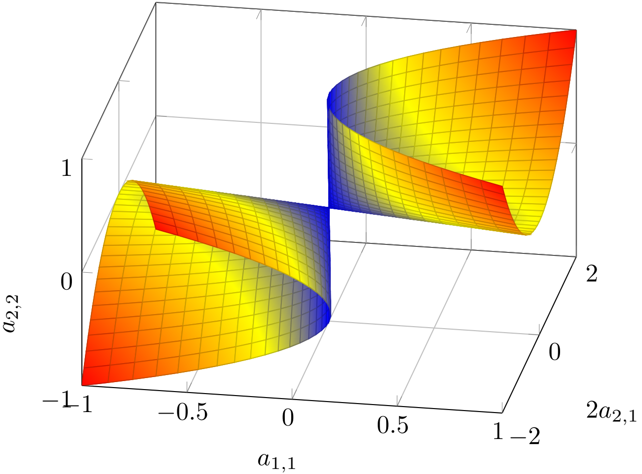
$$\Phi_r : (\mathcal{S} \times \cdots \times \mathcal{S})/\mathfrak{S}_r \to \mathbb{R}^{n_1 \times \cdots \times n_d}$$
$$\{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \mapsto \mathcal{A}_1 + \cdots + \mathcal{A}_r,$$

where $\mathcal{S}$ is the set of rank-1 tensors:

$$\mathcal{S} := \{\mathbf{a}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d \mid \mathbf{a}^k \in \mathbb{R}^{n_k} \setminus \{0\}\};$$

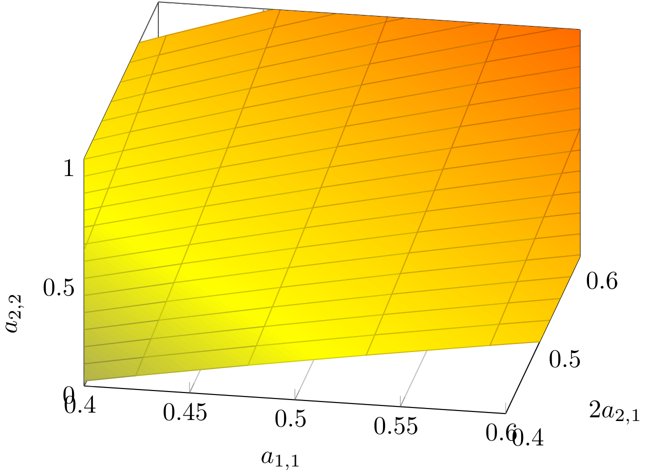
it is a **smooth manifold** called the **Segre manifold**.

$$v_2(\mathbb{R}^2) \subset S^2\mathbb{R}^2 \simeq \mathbb{R}^3.$$



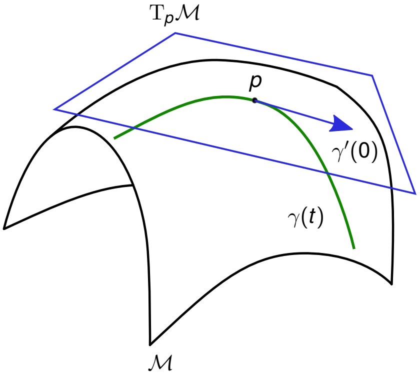
The manifold of rank-1 symmetric matrices $v_2(\mathbb{R}^2) \setminus \{0\}$ is **globally** a nonlinear object . . .

...but zooming in at $(a_{1,1}, 2a_{2,1}, a_{2,2}) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \in v_2(\mathbb{R}^2)$,



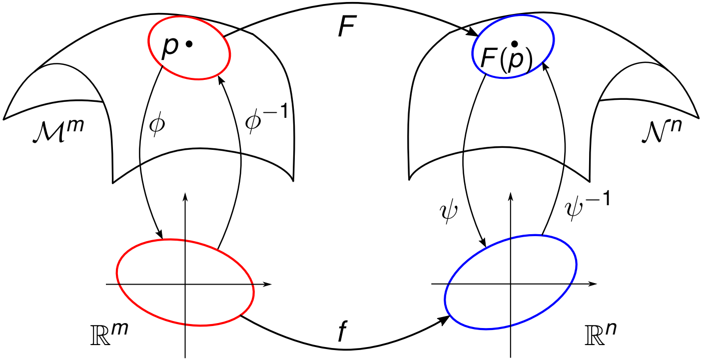
we see that it **locally** looks like a 2-dimensional linear space!
For a manifold, this is true at every point.

A **tangent vector** to an *m*-dimensional embedded submanifold $\mathcal{M} \subset \mathbb{R}^n$ at p is a vector $\mathbf{t}_p \in \mathbb{R}^n$ such that there exists a smooth curve $\gamma(t) \subset \mathcal{M}$, $t \in (-1, 1)$, for which $p = \gamma(0)$ and $\mathbf{t}_p = \gamma'(0)$.



The **tangent space** $\mathrm{T}_p\mathcal{M} \subset \mathbb{R}^n$ is the *m*-dimensional linear subspace spanned by all tangent vectors.

A **smooth map** $F : \mathcal{M}^m \to \mathcal{N}^n$ between smooth manifolds $\mathcal{M}^m$ and $\mathcal{N}^n$ is a generalization of a smooth map between Euclidean domains:



For maps between manifolds, we can apply Rice's (1966) **geometric framework of conditioning**:[3]

Proposition (Rice, 1966)

Let $\mathcal{M} \subset \mathbb{R}^m$ be a manifold of inputs and $\mathcal{N} \subset \mathbb{R}^n$ a manifold of outputs. Then, the condition number of $F : \mathcal{X} \to \mathcal{Y}$ at $x_0 \in \mathcal{X}$ is

$$\kappa[F](x_0) = \|\mathrm{d}_{x_0} F\| = \sup_{\|x\|=1} \|\mathrm{d}_{x_0} F(x)\|,$$

where $\mathrm{d}_{x_0} F : \mathrm{T}_{x_0}\mathcal{M} \to \mathrm{T}_{F(x_0)}\mathcal{N}$ is the ***derivative****.*

[3] See, e.g., Blum, Cucker, Shub, and Smale (1998) or Bürgisser and Cucker (2013) for a more modern treatment.

Recall that we seek the condition number of the **inverse map** of the addition map

$$\Phi_r : (\mathcal{S} \times \cdots \times \mathcal{S})/\mathfrak{S}_r \to \mathbb{R}^{n_1 \times \cdots \times n_d},$$
$$\{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \mapsto \mathcal{A}_1 + \cdots + \mathcal{A}_r.$$

Unfortunately, we cannot apply Rice's theorem because neither the source nor the image of Φ_r is a manifold!

We nevertheless showed that one can **restrict the domain and image** to **open dense subsets** such that the restriction is a **diffeomorphism**: a smooth injective map with smooth inverse.

Let $\mathcal{M}_{r;\mathbf{n}} \subset \mathcal{S}^{\times r}$ be the set of tuples of rank-1 tensors $\mathfrak{a} = (\mathcal{A}_1, \ldots, \mathcal{A}_r)$ in $\mathbb{R}^{n_1 \times \cdots \times n_d}$ that satisfy:

1. $\Phi_r(\mathfrak{a})$ is a **smooth point** of the Zariski closure of the set of rank-r tensors;
2. $\Phi_r(\mathfrak{a})$ is complex ***r*-identifiable**;
3. the **derivative** $\mathrm{d}_{\mathfrak{a}}\Phi_r$ **is injective**;

Definition

The set of ***r*-nice tensors** is

$$\mathcal{N}_{r;\mathbf{n}} := \Phi_r(\mathcal{M}_{r;\mathbf{n}}).$$

Let $\widehat{\mathcal{M}}_{r;\mathbf{n}} := \mathcal{M}_{r;\mathbf{n}}/\mathfrak{S}_r$; Beltrán, Breiding, and V (2019b) proved:

Proposition

Let $\mathbb{R}^{n_1\times\cdots\times n_d}$ be generically r-identifiable. Then,

$$\Phi_r : \widehat{\mathcal{M}}_{r;\mathbf{n}} \to \mathcal{N}_{r;\mathbf{n}}, \{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \to \mathcal{A}_1 + \cdots + \mathcal{A}_r$$

is a diffeomorphism. Moreover, $\mathcal{N}_{r,\mathbf{n}}$ is an open dense subset of the set of tensors of rank bounded by r.

Consequently, the inverse of Φ_r, restricted to the manifold of r-nice tensors, is

$$\tau_{r;\mathbf{n}} : \mathcal{N}_{r;\mathbf{n}} \to \widehat{\mathcal{M}}_{r;\mathbf{n}},\ \mathcal{A}_1 + \cdots + \mathcal{A}_r \to \{\mathcal{A}_1, \ldots, \mathcal{A}_r\},$$

which we call the **tensor rank decomposition map.**

As $\tau_{r;\mathbf{n}}$ is a smooth map between manifolds we can apply Rice's theorem. Since $\tau_{r;\mathbf{n}} \circ \Phi_r = \mathrm{Id}_{\mathcal{N}_{r;\mathbf{n}}}$ we have at $\mathcal{A} \in \mathcal{N}_{r;\mathbf{n}}$ that

$$\mathrm{d}_{\mathcal{A}}\tau_{r;\mathbf{n}} \circ \mathrm{d}_{\mathfrak{a}}\Phi_r = \mathrm{Id}_{\mathrm{T}_{\mathfrak{a}}\mathcal{N}_{r;\mathbf{n}}},$$

so that

$$\kappa[\tau_{r;\mathbf{n}}](\mathcal{A}) = \|\mathrm{d}_{\mathcal{A}}\tau_{r;\mathbf{n}}\|_2 = \|(\mathrm{d}_{\mathfrak{a}}\Phi_r)^{-1}\|_2.$$

The derivative $\mathrm{d}_{\mathfrak{a}}\Phi$ is seen to be the map

$$\begin{aligned}\mathrm{d}_{\mathfrak{a}}\Phi : \mathrm{T}_{\mathcal{A}_1}\mathcal{S} \times \cdots \times \mathrm{T}_{\mathcal{A}_r}\mathcal{S} &\to \mathrm{T}_{\mathcal{A}}\mathbb{R}^{n_1 \times \cdots \times n_d} \\ (\dot{\mathcal{A}}_1, \ldots, \dot{\mathcal{A}}_r) &\mapsto \dot{\mathcal{A}}_1 + \cdots + \dot{\mathcal{A}}_r.\end{aligned}$$

Hence, if U_i is an orthonormal basis of $\mathrm{T}_{\mathcal{A}_i}\mathcal{S} \subset \mathrm{T}_{\mathcal{A}_i}\mathbb{R}^{n_1 \times \cdots \times n_d}$, then the map is represented in coordinates as the matrix

$$U = \begin{bmatrix} U_1 & U_2 & \cdots & U_r \end{bmatrix} \in \mathbb{R}^{n_1 \cdots n_d \times r \dim \mathcal{S}}$$

In summary, the condition number of computing a CPD $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$ of $\mathcal{A}$ equals the inverse of the smallest singular value of $U = [\, U_1 \ U_2 \ \cdots \ U_r \,]$.

The tangent space to the Segre manifold $\mathcal{S}$ at

$$\lambda \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d, \quad \mathbf{a}^k \in \mathbb{S}(\mathbb{R}^{n_k}), \lambda \in \mathbb{R},$$

is given by the span of $\mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^d$ and the vectors

$$\begin{matrix} \mathbf{q}_2^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d, & \ldots, & \mathbf{q}_{n_1}^1 \otimes \mathbf{a}^2 \otimes \cdots \otimes \mathbf{a}^d, \\ & \vdots & \\ \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^{d-1} \otimes \mathbf{q}_2^d, & \ldots, & \mathbf{a}^1 \otimes \cdots \otimes \mathbf{a}^{d-1} \otimes \mathbf{q}_2^{n_d}, \end{matrix}$$

where $\{\mathbf{q}_i^k\}_i$ is an orthonormal basis of $(\mathbf{a}^k)^\perp$. Hence, each U_i has as columns the above vectors.

Some examples

Example 1: Matrices

When $d = 2$, and $A = \sum_{i=1}^{r} \lambda_i \mathbf{a}_i \otimes \mathbf{b}_i$, then

$$U = \begin{bmatrix} \mathbf{a}_i \otimes \mathbf{b}_i & \mathbf{a}_i^\perp \otimes \mathbf{b}_i & \mathbf{a}_i \otimes \mathbf{b}_i^\perp \end{bmatrix}_i .$$

This matrix does not have linearly independent columns because, for example,

$$\mathbf{a}_1 \otimes \mathbf{b}_r \in \langle \mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_1 \otimes \mathbf{b}_1^\perp \rangle \cap \langle \mathbf{a}_r \otimes \mathbf{b}_r, \mathbf{a}_r^\perp \otimes \mathbf{b}_r \rangle$$

Hence, the smallest singular value of U is zero, so that

$$\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) = \infty$$

for all $r > 1$ and $d = 2$.

Example 2: Essentially matrices

When $d \geq 3$ and $\mathcal{A} = \sum_{i=1}^{d} \lambda_i \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$ contains two rank-1 tensors

$$\mathcal{A}_i = \lambda_i \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \mathbf{a}^3 \otimes \cdots \otimes \mathbf{a}^d \quad \text{and} \quad \mathcal{A}_j = \lambda_j \mathbf{a}_j^1 \otimes \mathbf{a}_j^2 \otimes \mathbf{a}^3 \otimes \cdots \otimes \mathbf{a}^d$$

then

$$\mathcal{A}_i + \mathcal{A}_j \in \underbrace{\langle \mathbf{a}_i^1, \mathbf{a}_j^1 \rangle \otimes \langle \mathbf{a}_i^2, \mathbf{a}_j^2 \rangle \otimes \mathbf{a}^3 \otimes \cdots \otimes \mathbf{a}^d}_{\simeq \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R} \otimes \cdots \otimes \mathbb{R} \simeq \mathbb{R}^{2\times 2}}$$

Since matrices have intersecting tangent spaces, i.e., $\mathrm{span}(U_i) \cap \mathrm{span}(U_j) \neq \emptyset$, again we find that

$$\infty = \kappa(\mathcal{A}_i, \mathcal{A}_j) \leq \kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r)$$

Example 3: Odeco tensors

When $d \geq 3$ and $\mathcal{A} = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i^1 \otimes \cdots \otimes \mathbf{u}_i^d$ where all $\{\mathbf{u}_i^k\}_i$ form orthonormal bases, then it can be shown that U has orthonormal columns. Hence,

$$\kappa(\mathcal{A}_1, \ldots, \mathcal{A}_r) = 1$$

Interpretation

If

$$\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d$$
$$\mathcal{B} = \mathcal{B}_1 + \cdots + \mathcal{B}_r = \sum_{i=1}^{r} \mathbf{b}_i^1 \otimes \cdots \otimes \mathbf{b}_i^d$$

are tensors in $\mathbb{R}^{n_1 \times \cdots \times n_d}$, then for $\|\mathcal{A} - \mathcal{B}\|_F \approx 0$ we have the **asymptotically sharp bound**

$$\underbrace{\min_{\pi \in \mathfrak{S}_r} \sqrt{\sum_{i=1}^{r} \|\mathcal{A}_i - \mathcal{B}_{\pi_i}\|_F^2}}_{\text{forward error}} \lesssim \underbrace{\kappa[\tau_{r;\mathbf{n}}](\mathcal{A})}_{\text{condition number}} \cdot \underbrace{\|\mathcal{A} - \mathcal{B}\|_F}_{\text{backward error}}$$

Overview

Expected value

The condition number (of the problem of computing a CPD) is a **local property** by its very definition. Given a CPD, it can be computed quite efficiently.

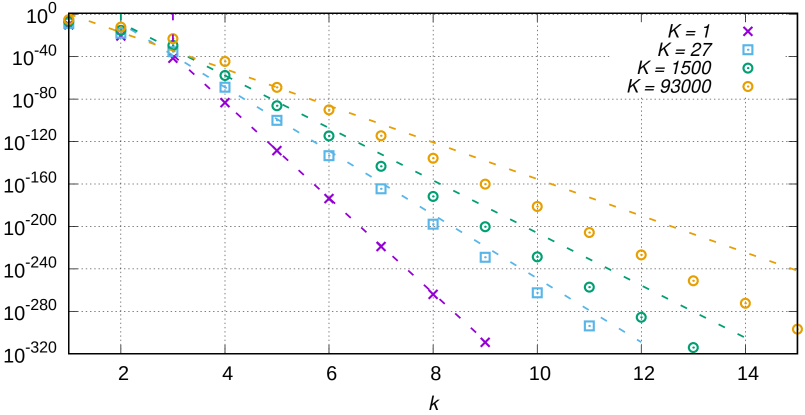
Next, we would like to understand its **global behavior**:

- what is a typical condition number?
- where is the condition number small or large?
- what is the distribution of condition numbers?

The motivation for studying these questions concerns the analysis of the **performance of computational methods** for solving the tensor rank decomposition problem.

For example, the **rate of convergence** and **radius of attraction** of Riemannian Gauss–Newton methods depend on the condition number of the CPD.[4]

Distance to solution



[4] See, e.g., Breiding and V (2018b)

Let the rank r and the size $n_1 \geq \cdots \geq n_d \geq 2$ be fixed.

The crudest global property one could hope to compute is the **expected value**

$$\mathop{\mathbb{E}}_{\mathcal{A}\sim\rho} \kappa(\mathcal{A}) := \int_{\mathcal{N}_{r;\mathbf{n}}} \kappa(\mathcal{A})\rho(\mathcal{A})\,\mathrm{d}\mathcal{A},$$

relative to a density ρ on $\mathcal{N}_{r;\mathbf{n}}$.

We consider the following natural Gaussian density

$$\rho(\mathcal{A}) := \frac{1}{C_{r;\mathbf{n}}} e^{-\frac{\|\mathcal{A}\|^2}{2}}, \quad \text{where } C_{r;\mathbf{n}} := \int_{\mathcal{N}_{r;\mathbf{n}}} e^{-\frac{\|\mathcal{A}\|^2}{2}} \mathrm{d}\mathcal{A}.$$

For **other computational problems** estimates of the average condition number have been computed:[5]

1. $n \times n$ matrices A with i.i.d. $N(0,1)$ entries have

$$\mathbb{E}[\kappa(A)] \leq 20.1n.$$

2. Random homogeneous systems F of n polynomial equations in $n+1$ variables of degrees d_i have

$$\mathbb{E}[\kappa(F)] \leq 5\sqrt{nN}\ln^2(nN) \quad \text{where } N = \sum_{i=1}^{n} \binom{n+d_i}{n}$$

3. $n \times n$ triangular matrices L with i.i.d. $N(0,1)$ entries have

$$\mathbb{E}[\kappa(L)] \geq \frac{2^{n-1}}{n}.$$

[5] See Bürgisser and Cucker (2013) for the precise statements.

Computing the expected value

$$\mathcal{I}_r := \mathop{\mathbb{E}}_{\mathcal{A}\sim\rho} \kappa(\mathcal{A}) = \frac{1}{C_{r;\mathbf{n}}} \int_{\mathcal{N}_{r;\mathbf{n}}} \kappa(\mathcal{A}) e^{-\frac{\|\mathcal{A}\|^2}{2}} \,\mathrm{d}\mathcal{A}$$

is quite hard; we do not have an expression for all ranks.

Nevertheless, Beltrán, Breiding, and V (2019) proved the following **surprising and harsh** result:

Computing the expected value

$$\mathcal{I}_r := \mathop{\mathbb{E}}_{\mathcal{A}\sim\rho} \kappa(\mathcal{A}) = \frac{1}{C_{r;\mathbf{n}}} \int_{\mathcal{N}_{r;\mathbf{n}}} \kappa(\mathcal{A}) e^{-\frac{\|\mathcal{A}\|^2}{2}} \,\mathrm{d}\mathcal{A}$$

is quite hard; we do not have an expression for all ranks.

Nevertheless, Beltrán, Breiding, and V (2019) proved the following **surprising and harsh** result:

Theorem

Assume that $r \geq 2$,

- $n_1 \times \cdots \times n_d$ *tensors are* r*-identifiable, and*
- $(n_1 - 2) \times \cdots \times (n_d - 2)$ *tensors are* $(r-2)$*-identifiable,*

then we have

$$\mathcal{I}_r = \infty.$$

Conjecture (Beltrán, Breiding, V, 2019)

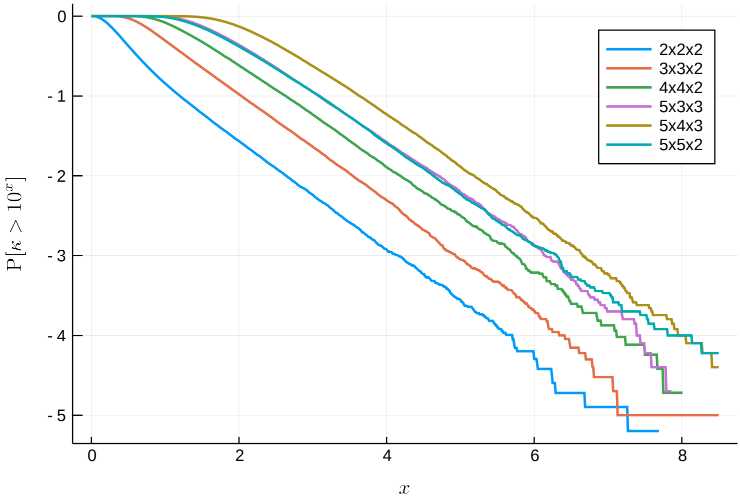
$$\mathcal{I}_r = \infty \quad \text{for all } r \geq 2,$$

In order to test this conjecture numerically, we can consider so-called **perfect tensor spaces** for which there exists an r such that

$$\dim \mathcal{N}_{r;\mathbf{n}} = n_1 n_2 \cdots n_d.$$

For these spaces, $\mathcal{N}_{r;\mathbf{n}}$ is an **open subset** of $\mathbb{R}^{n_1 \times \cdots \times n_d}$ so that we can sample from the Gaussian density ρ simply by drawing an $n_1 \times \cdots \times n_d$ tensor $\mathcal{A}$ with i.i.d. standard normal entries.

We then use **homotopy continuation** to compute a CPD of $\mathcal{A}$ using the HomotopyContinuation.jl package by Breiding and Timme. If it finds a real CPD, we record its condition number.



The asymptotic slope of every curve is about $-\frac{2}{3}$.

The foregoing experiments support the conjecture. Indeed,

$$P(x) = \int_0^x p(y)\,\mathrm{d}y = \mathrm{P}[\kappa \le x] \approx 1 - cx^{-\frac{2}{3}},$$

where p is the **probability density function**, P its **cumulative distribution function**, and $c > 0$ is some constant.

From this we find, empirically,

$$\int_1^\infty \kappa \cdot p(\kappa)\,\mathrm{d}\kappa = \frac{2}{3}\int_1^\infty \kappa^{-\frac{2}{3}}\mathrm{d}\kappa = \infty$$

Overview

Numerical stability

In the previous section, we showed that the expected condition number at the **input** of r-nice tensors $\mathcal{N}_{r;\mathbf{n}}$ is

$$\mathop{\mathbb{E}}_{\mathcal{A}\sim\rho} \kappa(\mathcal{A}) = \infty$$

in many cases.

What if we would choose a different density ρ? For example, by sampling the **output** of the decomposition problem randomly?

We could decide to choose the so-called **factor matrices**

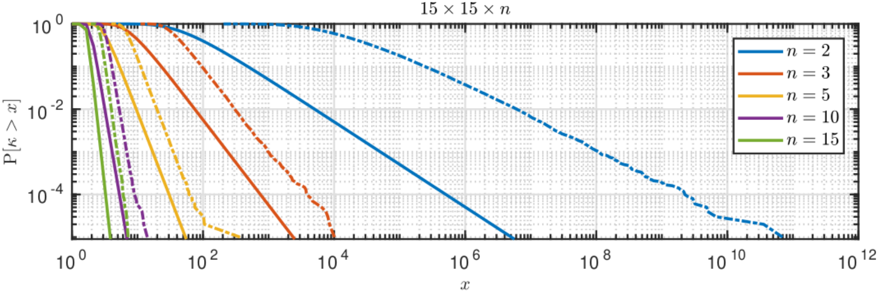
$$A_k := \begin{bmatrix} \mathbf{a}_1^k & \mathbf{a}_2^k & \cdots & \mathbf{a}_r^k \end{bmatrix} \in \mathbb{R}^{n_k \times r}$$

randomly by sampling each entry i.i.d. $N(0, 1)$. This yields a corresponding tensor

$$\mathcal{A} = \sum_{i=1}^{r} \mathbf{a}_i^1 \otimes \cdots \otimes \mathbf{a}_i^d;$$

with probability 1 it is an r-nice tensor if the space is generically r-identifiable.

Something very interesting is going on here ($r = 15$):



(a) A, B, and C i.i.d. standard normal entries.

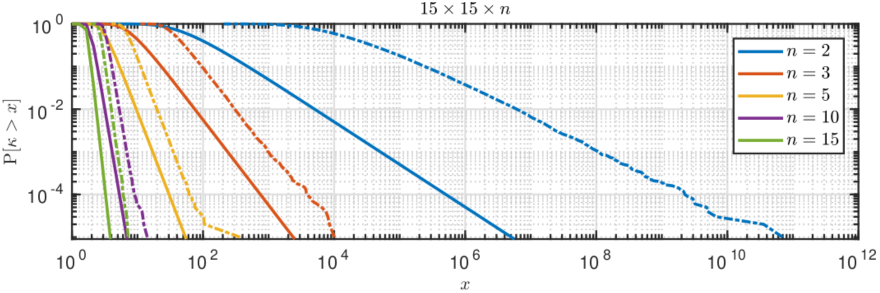
This picture was an eye-opener for us!

The **only practical class of direct algorithms** for computing exact CPDs performs the following steps:

1. **Orthogonally project** the input $n_1 \times n_2 \times n_3$ tensor $\mathcal{A}$ to an $n_1 \times n_2 \times 2$ tensor $P\mathcal{A}$.
2. Recover the first two factor matrices A_1 and A_2 from $P\mathcal{A}$, e.g., from a **generalized eigendecomposition.**
3. Recover the last factor matrix A_3, e.g., by solving a **linear least squares problem.**

Beltrán, Breiding and V (2019b) called them **pencil-based algorithms.**

Look again!



(a) A, B, and C i.i.d. standard normal entries.

With some effort we were able to prove the following result:

Theorem (Beltrán, Breiding, V, 2019b)

Let $n_1, n_2 \geq r \geq 2$ and $n_3 \geq r + 2$. For every pencil-based algorithm, there exists an **open set** of the rank-r tensors in $\mathbb{R}^{n_1 \times n_2 \times n_3}$ for which it is **arbitrarily numerically forward unstable**.

Overview

Conclusions

The presented results contribute to a body of work indicating that computing CPDs is a very **challenging problem** in general. The new state of the art is that

- tensor **rank is NP complete** (Håstad, 1990);
- **open boundary tensors exist** and there is an open set of ill-posed inputs for approximation by a low-rank CPD (de Silva and Lim, 2008);
- the **average condition number is infinite** for most spaces (Beltrán, Breiding, and V, 2019); and
- almost all practical **direct algorithms for CPD are numerically unstable** (Beltrán, Breiding, and V, 2019b).

Further reading

- Breiding and V (2018), *The condition number of join decompositions*, SIMAX.
- Breiding and V (2018c), *On the average condition number of tensor rank decompositions*, arXiv:1801.01673, 2018.
- Beltrán, Breiding, and V (2019), *The average condition number of tensor rank decomposition is infinte*, In preparation.
- Beltrán, Breiding, and V (2019b), *Pencil-based algorithms for tensor rank decomposition are not stable*, arXiv:1807.04159.

Vielen Dank für Ihre Aufmerksamkeit!

References

Introduction

- Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 1927.
- Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Lin. Alg. Appl., 1977.

Generic identifiability

- Angelini, *On complex and real identifiability of tensors*, Rev. Mat. Uni. Parma, 2017.
- Angelini, Bocci, Chiantini, *Real identifiability vs. complex identifiability*, Linear Multilinear Algebra, 2017.
- Ballico, *On the weak non-defectivity of veronese embeddings of projective spaces* , Centr. Eur. J. Math., 2005.
- Bocci, Chiantini, and Ottaviani, *Refined methods for the identifiability of tensors*, Ann. Mat. Pura Appl., 2013.
- Chiantini and Ottaviani, *On generic identifiability of 3-tensors of small rank*, SIAM J. Matrix Anal. Appl., 2013.
- Chiantini, Ottaviani, and Vannieuwenhoven, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIMAX, 2014.
- Chiantini, Ottaviani, and Vannieuwenhoven, *On generic identifiability of symmetric tensors of subgeneric rank*, Trans. Amer. Math. Soc., 2017.
- Chiantini, Ottaviani, and Vannieuwenhoven, *Effective criteria for specific identifiability of tensors and forms*, SIAM J. Matrix Anal. Appl., 2017.
- Galuppi and Mella, *Identifiability of homogeneous polynomials and Cremona Transformations*, J. Reine Angew. Math., 2017.
- Qi, Comon, and Lim, *Semialgebraic geometry of nonnegative tensor rank*, SIMAX, 2016.

Sensitivity

- Blum, Cucker, Shub, and Smale, *Complexity and Real Computation*, 1998.
- Beltrán, Breiding, and Vannieuwenhoven, *The average condition number of tensor rank decomposition is infinite*, 2019. (In preparation)
- Beltrán, Breiding, and Vannieuwenhoven, *Pencil-based algorithms for tensor rank decomposition are not stable*, arXiv:1807.04159, 2019b.
- Breiding and Vannieuwenhoven, *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl., 2018.
- Breiding and Vannieuwenhoven, *Convergence analysis of Riemannian Gauss-Newton methods and its connection with the geometric condition number*, Appl. Math. Letters, 2018b.
- Breiding and Vannieuwenhoven, *On the average condition number of tensor rank decompositions*, arXiv:201801.01673, 2018c.
- Bürgisser and Cucker, *Condition: The Geometry of Numerical Algorithms*, Springer, 2013.
- Lee, *Introduction to Smooth Manifolds*, 2013.
- Rice, *A theory of condition*, SIAM J. Numer. Anal., 1966.
- Vannieuwenhoven, *A condition number for the tensor rank decomposition*, Linear Algebra Appl., 2017.

Detailed integral computation

The idea is to transform this integral to a simpler domain by exploiting the local diffeomorphism between $\mathcal{N}_r$ and $\mathcal{M}_r \subset \mathcal{S} \times \cdots \times \mathcal{S}$ and the fact that the Segre manifold can be parameterized via the 2^d-to-1 covering map

$$\psi : \overbrace{\mathbb{R}_0 \times \mathbb{S}^{n_1-1} \times \cdots \times \mathbb{S}^{n_d-1}}^{D} \to \mathcal{S}$$
$$(\lambda, \mathbf{u}_1, \ldots, \mathbf{u}_d) \mapsto \lambda \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d.$$

For rank $r = 2$, we get from applying the co-area formula

$$\mathcal{I}_2 = \int_{\mathcal{N}_2} \kappa(\mathcal{A}) e^{-\frac{\|\mathcal{A}\|^2}{2}} \mathrm{d}\mathcal{A}$$
$$\simeq \int_{D^{\times 2}} \mathrm{Jac}[\Phi_2 \circ (\psi \times \psi)](\mathfrak{a}, \mathfrak{b}) \cdot \kappa(\psi(\mathfrak{a}), \psi(\mathfrak{b})) e^{-\frac{\|\psi(\mathfrak{a})+\psi(\mathfrak{b})\|^2}{2}} \mathrm{d}\mathfrak{a}\,\mathrm{d}\mathfrak{b},$$

where $\simeq$ indicates equality up multiplication by a constant, and $\mathrm{Jac}[\phi](A)$ is the Jacobian determinant of ϕ at A.

Let $\mathfrak{a} = (\lambda, \mathbf{u}_1, \ldots, \mathbf{u}_d)$, $\mathfrak{b} = (\mu, \mathbf{v}_1, \ldots, \mathbf{v}_d)$, and consider

$$J_{\mathfrak{a}} = \begin{bmatrix} \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d & \lambda \dot{U}_2 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d & \cdots & \lambda \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d-1} \otimes \dot{U}_d \end{bmatrix}$$
$$U_{\mathfrak{a}} = \begin{bmatrix} \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_d & \dot{U}_2 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d & \cdots & \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_{d-1} \otimes \dot{U}_d \end{bmatrix},$$

where $\dot{U}_k \in \mathbb{R}^{n_k \times (n_k - 1)}$ contains as columns an orthonormal basis of $\mathbf{u}_k^\perp$.

It follows that the Jacobian determinant is

$$\mathrm{Jac}[\Phi_2 \circ (\psi \times \psi)](\mathfrak{a}, \mathfrak{b}) = \det\left(\begin{bmatrix} J_{\mathfrak{a}} & J_{\mathfrak{b}} \end{bmatrix}^T \begin{bmatrix} J_{\mathfrak{a}} & J_{\mathfrak{b}} \end{bmatrix} \right)^{\frac{1}{2}}$$
$$= |\lambda|^{\Sigma - 1} |\mu|^{\Sigma - 1} \det\left(\begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix}^T \begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix} \right)^{\frac{1}{2}}$$

From Breiding and V (2018), we also know

$$\kappa(\psi(\mathfrak{a}), \psi(\mathfrak{b})) = \left\| \begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix}^{\dagger} \right\|_2.$$

Consequently,

$$\mathrm{Jac}[\Phi_2 \circ (\psi \times \psi)](\mathfrak{a}, \mathfrak{b})\, \kappa(\psi(\mathfrak{a}), \psi(\mathfrak{b})) = \varsigma_1 \varsigma_2 \cdots \varsigma_{2\Sigma - 1},$$

where ς_i is the ith largest singular value of $\begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix}$.

Analyzing the spectrum of $\begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix}$ is a challenge, but it is feasible for rank 2. The key idea is the following observation.

Let $\mathbf{u} \in \mathbb{S}^n$ and $\mathbf{v} := \frac{\mathbf{u}+\epsilon\mathbf{x}}{\sqrt{1+\epsilon^2}} \in \mathbb{S}^n$ for some $\mathbf{x} \in \mathbb{S}^n$ with $\mathbf{x} \perp \mathbf{u}$. Then,

$$\det\left(\begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix}\right) = \det\left(\frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\right)$$

The matrix product simplifies to

$$\frac{1}{\sqrt{2}} \left[\left(1 + \tfrac{1}{\sqrt{1+\epsilon^2}}\right)\mathbf{u} + \tfrac{\epsilon}{\sqrt{1+\epsilon^2}}\mathbf{x}, \quad \left(1 - \tfrac{1}{\sqrt{1+\epsilon^2}}\right)\mathbf{u} - \tfrac{\epsilon}{\sqrt{1+\epsilon^2}}\mathbf{x} \right],$$

which has **orthogonal columns**!

Analyzing the spectrum of $\begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix}$ is a challenge, but it is feasible for rank 2. The key idea is the following observation.

Let $\mathbf{u} \in \mathbb{S}^n$ and $\mathbf{v} := \frac{\mathbf{u}+\epsilon\mathbf{x}}{\sqrt{1+\epsilon^2}} \in \mathbb{S}^n$ for some $\mathbf{x} \in \mathbb{S}^n$ with $\mathbf{x} \perp \mathbf{u}$. Then,

$$\det\left(\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix}\right) = \det\left(\frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\right) \approx \underbrace{\sqrt{2}}_{\varsigma_1}\cdot\underbrace{\left(\epsilon - \frac{3}{8}\epsilon^3 + \cdots\right)}_{\varsigma_2}$$

The matrix product simplifies to

$$\frac{1}{\sqrt{2}}\left[\left(1+\tfrac{1}{\sqrt{1+\epsilon^2}}\right)\mathbf{u} + \tfrac{\epsilon}{\sqrt{1+\epsilon^2}}\mathbf{x},\quad \left(1-\tfrac{1}{\sqrt{1+\epsilon^2}}\right)\mathbf{u} - \tfrac{\epsilon}{\sqrt{1+\epsilon^2}}\mathbf{x}\right],$$

which has **orthogonal columns**!

The whole idea generalizes to $\begin{bmatrix} U_{\mathfrak{a}} & U_{\mathfrak{b}} \end{bmatrix}$. We can show that

$$\varsigma_1 \approx \cdots \approx \varsigma_\Sigma \approx \sqrt{2} \quad \text{and} \quad \varsigma_{\Sigma+1} \geq \cdots \geq \varsigma_{2\Sigma-1} \geq C \cdot \epsilon$$

provided that

$$\frac{9}{10}\|\mathbf{u}_1 - \mathbf{v}_1\| \leq \|\mathbf{u}_i - \mathbf{v}_i\| \leq \|\mathbf{u}_1 - \mathbf{v}_1\| \quad i = 1, 2, \ldots, d,$$

and where $\mathfrak{a} = (\lambda, \mathbf{u}_1, \ldots, \mathbf{u}_d)$ and $\mathfrak{b} = (\mu, \mathbf{v}_1, \ldots, \mathbf{v}_d)$.

Let $D(\epsilon)$ be the open neighborhood of $(\mathbb{S}^{n_1-1} \times \cdots \times \mathbb{S}^{n_d-1})^{\times 2}$ where $\|\mathbf{u}_1 - \mathbf{v}_1\| = \epsilon$ and the above conditions hold.

Putting all of the foregoing together, we get

$$\mathcal{I}_2 \geq C' \int_{(\mathfrak{a},\mathfrak{b})\in\mathbb{R}^2\times D(\epsilon)} \|\mathbf{u}_1 - \mathbf{v}_1\|^{\Sigma-1} |\lambda|^{\Sigma-1} |\mu|^{\Sigma-1} e^{-\frac{\|\psi(\mathfrak{a})+\psi(\mathfrak{b})\|^2}{2}} \mathrm{d}\mathfrak{a}\mathrm{d}\mathfrak{b}$$

With some effort, the integral over (λ, μ) against the weight function can be shown to satisfy

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |\lambda|^{\Sigma-1} |\mu|^{\Sigma-1} e^{-\frac{\|\lambda \mathbf{u}_1\otimes\cdots\otimes\mathbf{u}_d + \mu\mathbf{v}_1\otimes\cdots\otimes\mathbf{v}_d\|^2}{2}} \mathrm{d}\lambda\mathrm{d}\mu \geq \frac{C''}{\|\mathbf{u}_1 - \mathbf{v}_1\|^{2\Sigma-1}}.$$

Hence,

$$\mathcal{I}_2 \geq C'C'' \int_{(\mathfrak{a},\mathfrak{b})\in D(\epsilon)} \frac{1}{\|\mathbf{u}_1 - \mathbf{v}_1\|^{\Sigma-1}} \mathrm{d}\mathfrak{a}\,\mathrm{d}\mathfrak{b}.$$

After some more work integrating out the spherical bands, we are left with

$$\mathcal{I}_2 \geq C'C''C''' \int_{\mathbf{u}_1 \in \mathbb{S}^{n_1-1}} \int_{\mathbf{v}_1 \in \mathbb{S}^{n_1-1}, \|\mathbf{u}_1 - \mathbf{v}_1\| \leq \epsilon} \frac{1}{\|\mathbf{u}_1 - \mathbf{v}_1\|^{n_1}} \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{v}_1.$$

The inner integral, after switching to polar coordinates, integrates to

$$\int_0^\epsilon \frac{t^{n_1-2}}{t^{n_1}} \mathrm{d}t$$

After some more work integrating out the spherical bands, we are left with

$$\mathcal{I}_2 \geq C'C''C''' \int_{\mathbf{u}_1 \in \mathbb{S}^{n_1-1}} \int_{\mathbf{v}_1 \in \mathbb{S}^{n_1-1}, \|\mathbf{u}_1 - \mathbf{v}_1\| \leq \epsilon} \frac{1}{\|\mathbf{u}_1 - \mathbf{v}_1\|^{n_1}} \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{v}_1.$$

The inner integral, after switching to polar coordinates, integrates to

$$\int_0^\epsilon \frac{t^{n_1-2}}{t^{n_1}} \mathrm{d}t = \infty.$$

After some more work integrating out the spherical bands, we are left with

$$\mathcal{I}_2 \geq C'C''C''' \int_{\mathbf{u}_1 \in \mathbb{S}^{n_1-1}} \int_{\mathbf{v}_1 \in \mathbb{S}^{n_1-1}, \|\mathbf{u}_1 - \mathbf{v}_1\| \leq \epsilon} \frac{1}{\|\mathbf{u}_1 - \mathbf{v}_1\|^{n_1}} \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{v}_1.$$

The inner integral, after switching to polar coordinates, integrates to

$$\int_0^\epsilon \frac{t^{n_1-2}}{t^{n_1}} \mathrm{d}t = \infty.$$

Consequently,

$$\mathcal{I}_2 = \infty \, \mathbf{!}$$