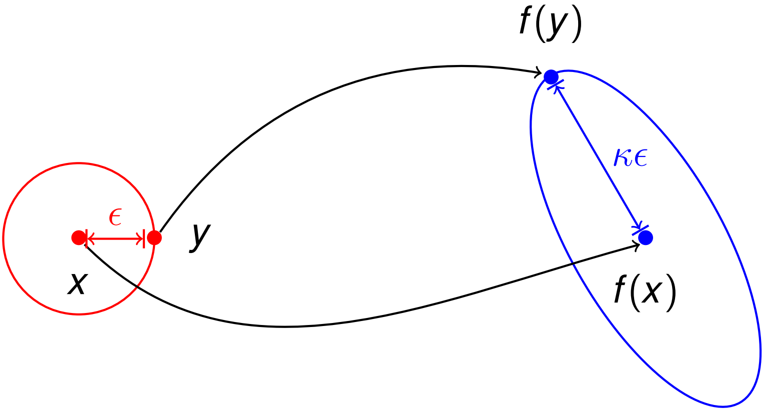
Nick Vannieuwenhoven

KU LEUVEN

with Beltrán, Breiding & Dewaele

Sensitivity of tensor decompositions

I. Sensitivity of tensor decompositions



A numerical experiment ...

Consider the matrix

$$A = \frac{1}{177147}\begin{bmatrix} 88574 & 88574 & 2 \\ 88574 & 88574 & 2 \\ 2 & 2 & 177146 \end{bmatrix}$$

Computing the eigenvalue decomposition $\widehat{V}\widehat{\Lambda}\widehat{V}^{-1}$ of A numerically using Octave, we find $\|A - \widehat{V}\widehat{\Lambda}\widehat{V}^{-1}\|_2 \approx 1.1 \cdot 10^{-15}$.

The eigenvalues are

numerical	exact
0.0000000000000000011..	0
0.9999830649121912	$0.99998306491219156 9... = 1 - 3^{-10}$
1.0000169350878100	$1.00001693508780843 0... = 1 + 3^{-10}$

We found 15 correct digits of the exact solution.

However, when comparing the computed eigenvector corresponding to $\lambda_1 = 1 + 3^{-10}$ to the exact solution, we get

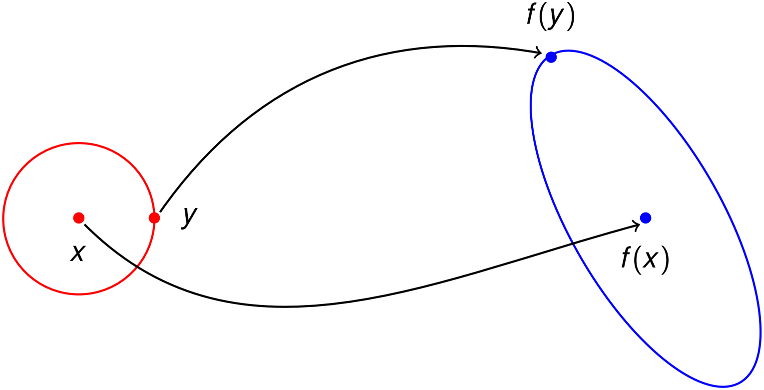
numerical	exact
0.577350269187273	$\frac{1}{\sqrt{3}}$
0.577350269187273	$\frac{1}{\sqrt{3}}$
0.577350269194331	$\frac{1}{\sqrt{3}}$

We recovered only 11 digits correctly, even though the matrix $\widehat{V}\widehat{\Lambda}\widehat{V}^{-1}$ contains at least 15 correct digits of each entry.

It seems that the eigenvalues are **computed more accurately** than the eigenvectors. Why is that?

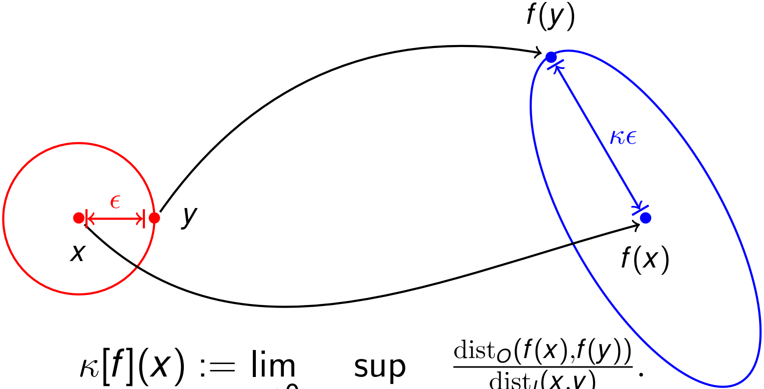
Condition numbers

Rice (1966) defined the **condition number** for maps $f : I \to O$ between general **metric spaces** I, O. It quantifies the **worst-case sensitivity** of f to perturbations of the input.



Condition numbers

Rice (1966) defined the **condition number** for maps $f : I \to O$ between general **metric spaces** I, O. It quantifies the **worst-case sensitivity** of f to perturbations of the input.



$$\kappa[f](x) := \lim_{\epsilon \to 0} \sup_{\mathrm{dist}_I(x,y)=\epsilon} \frac{\mathrm{dist}_O(f(x),f(y))}{\mathrm{dist}_I(x,y)}.$$

The condition number implies the **error bound**:

$$\mathrm{dist}_O(f(x), f(y)) \leq \kappa[f](x) \cdot \mathrm{dist}_I(x, y) + o(\mathrm{dist}_I(x, y))$$

for small $\mathrm{dist}_I(x, y)$.

This is roughly the error one should expect[1] from numerically solving the problem (i.e., computing f).

[1] Armentano (2010) shows that average and worst-case behavior are of the same order.

Assuming the eigenvalues are distinct, our computational problems can be **modeled locally as analytic functions**[2]

$$\lambda_1 : \mathrm{Sym}(\mathbb{R}^{m\times m}) \to \mathbb{R}, \quad \text{resp.} \quad v_1 : \mathrm{Sym}(\mathbb{R}^{m\times m}) \to \mathbb{S}^{m-1}.$$

What we observed above is that

$$0.41 \approx \frac{|\lambda_1(A) - \lambda_1(A+\Delta)|}{\|\Delta\|_2} \ll \frac{\|v_1(A) - v_1(A+\Delta)\|}{\|\Delta\|_2} \approx 5.33 \cdot \mathbf{10^3}$$

where $\|\Delta\|_2 \approx 1.1 \cdot 10^{-15}$ in this case.

[2] See Kato (1995) for a treatise on this subject.

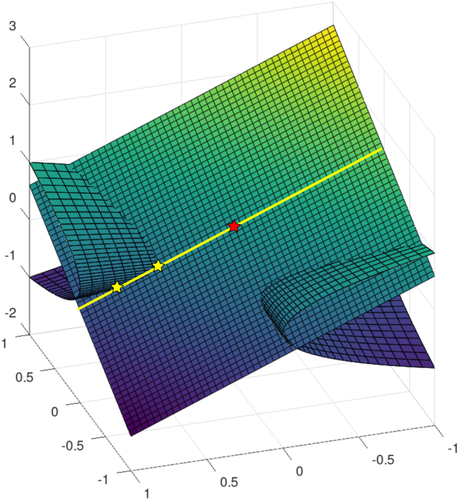
For eigenvalues and eigenvectors, in the foregoing distances, it is known that the condition numbers are respectively[3]

$$\kappa[\lambda_1](A) = 1,$$
$$\kappa[v_1](A) = \frac{1}{\min_{j\neq 1} |\lambda_1(A) - \lambda_j(A)|} = \frac{1}{2 \cdot 3^{-10}} \approx 2.95 \cdot 10^4$$

This largely explains the difference in accuracy between eigenvalues and eigenvectors in the foregoing numerical experiment.

[3] See Chapter 14 of Bürgisser and Cucker (2013) for one approach.

II. Sensitivity of tensor decompositions



Tensor join decompositions

Breiding and V (2018a) studied the condition number of the problem of decomposing a **join decomposition** into its constituent parts:

$$\mathcal{A} = \sum_{i=1}^{r} \mathcal{A}_i \quad \text{where } \mathcal{A}_i \in \mathcal{S}_i$$

and $\mathcal{S}_i \subset \mathbb{R}^N$ is (locally) a differentiable manifold.

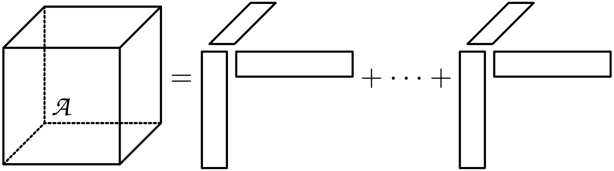
Some examples include:

- low-rank matrix decompositions;
- the tensor rank decomposition, or canonical polyadic decomposition;
- the symmetric tensor rank decomposition, or Waring decomposition; and
- block term decompositions.

Tensor rank decomposition

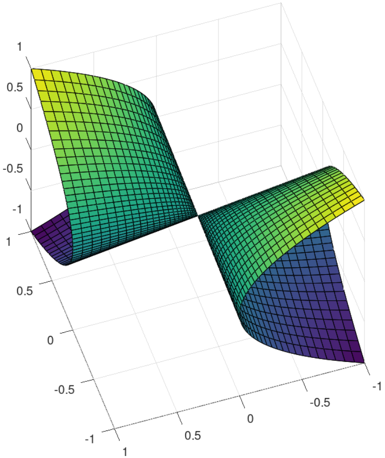
Hitchcock (1927) introduced the **tensor rank decomposition**:

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



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

The tensor rank decomposition has a **beautiful geometric interpretation**, already explained in Anna Seigal's tutorial talks.



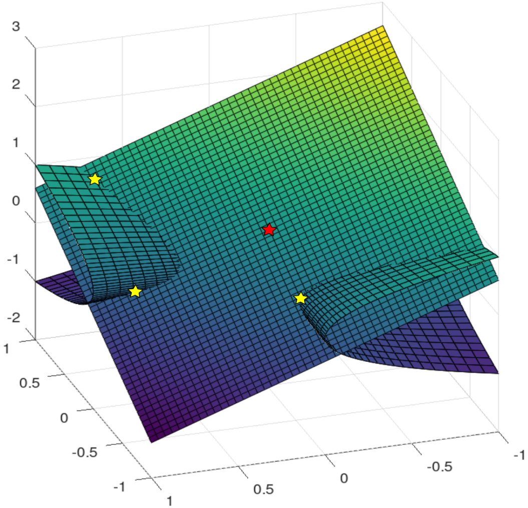
The set of all real rank-1 tensors

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

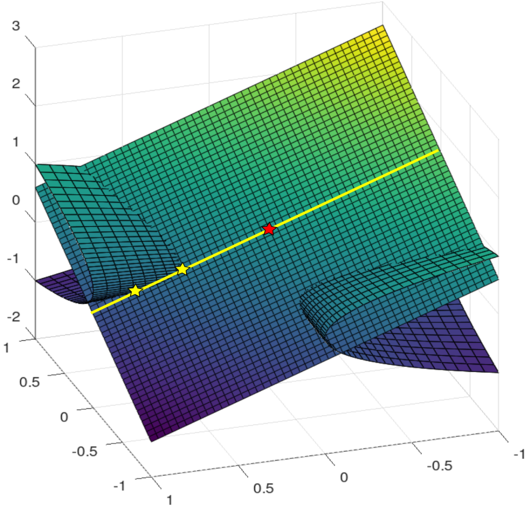
can be equipped with a natural topology and a smooth and metric structure.

It is a smooth submanifold of $\mathbb{R}^{n_1 \times \cdots \times n_d}$ called the **Segre manifold**.

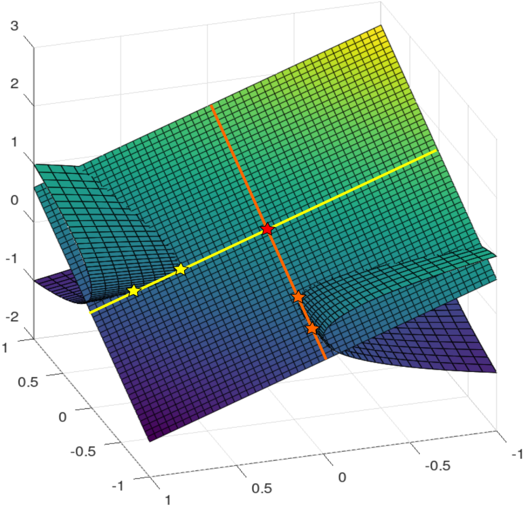
A tensor $\mathcal{A}$ is of rank at most 3, if there exist 3 points on $\mathcal{S}$ such that lies in the span of these points:



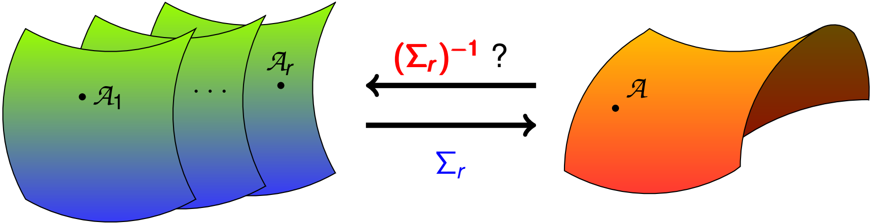
If there is a unique set of rank-1 tensors $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$ of cardinality r such that $\mathcal{A}$ lies in the span of these points, then $\mathcal{A}$ is r**-identifiable.** Otherwise it is not r-identifiable, and this can happen:



If there is a unique set of rank-1 tensors $\{\mathcal{A}_1, \ldots, \mathcal{A}_r\}$ of cardinality r such that $\mathcal{A}$ lies in the span of these points, then $\mathcal{A}$ is **r-identifiable.** Otherwise it is not r-identifiable, and this can happen:



III. Sensitivity of tensor rank decompositions

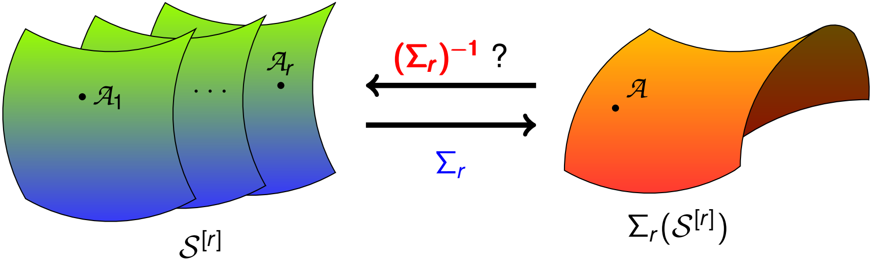


The tensor rank decomposition problem

We want to determine the condition number of the **inverse*** of the **tensor rank addition map**

$$\begin{aligned}\Sigma_r : \quad & \mathcal{S}^{[r]} \to \mathbb{R}^{n_1 \times \cdots \times n_d} \\ & \{\mathcal{A}_1, \ldots, \mathcal{A}_r\} \mapsto \mathcal{A}_1 + \cdots + \mathcal{A}_r,\end{aligned}$$

where $\mathcal{S}$ is the Segre manifold of rank-1 tensors.



Fortunately, essentially this problem was solved by Ulisse Dini in his 1877/1878 *Lezioni d'analisi infinitesimale*.

If $\mathrm{d}_{\{\mathcal{A}_1,\ldots,\mathcal{A}_r\}}\Sigma_r|_{\mathcal{M}_r}$ is invertible, the **inverse function theorem** for manifolds states that there exist open neighborhoods $\mathcal{O}_r$ of $\{\mathcal{A}_1,\ldots,\mathcal{A}_r\}$ and $\mathcal{I}_r$ of $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r$ and a smooth inverse function

$$\tau_r \circ (\Sigma_r|_{\mathcal{O}_r}) = \mathrm{Id}.$$

Consequently,

$$\mathrm{d}_{\mathcal{A}}\tau_r \circ \mathrm{d}_{\{\mathcal{A}_1,\ldots,\mathcal{A}_r\}}(\Sigma_r|_{\mathcal{O}_r}) = \mathrm{Id}.$$

We call τ_r a local **tensor rank decomposition map.**

Since τ_r is a **smooth map** between Riemannian submanifolds of Euclidean spaces we can apply Rice's theorem:

Theorem (Rice, 1966)

Let $\mathcal{I} \subset \mathbb{R}^m$ be a manifold of inputs and $\mathcal{O} \subset \mathbb{R}^n$ a manifold of outputs. Then, the condition number of $f : \mathcal{I} \to \mathcal{O}$ at $x \in \mathcal{I}$ is

$$\kappa[f](x) = \|\mathrm{d}_x f\|_2 = \sup_{\|x\|=1} \|\mathrm{d}_x f(x)\|,$$

*where $\mathrm{d}_x f : \mathrm{T}_x\mathcal{I} \to \mathrm{T}_{f(x)}\mathcal{O}$ is the **derivative**.*

Putting Dini and Rice together:

$$\kappa[\tau_r](\mathcal{A}) = \|\mathrm{d}_{\mathcal{A}}\tau_r\|_2 = \|(\mathrm{d}_{\{\mathcal{A}_1,\ldots,\mathcal{A}_r\}}\Sigma_r|_{\mathcal{O}_r})^{-1}\|_2 = \frac{1}{\sigma_{r\dim\mathcal{S}}\left(\mathrm{d}_{\{\mathcal{A}_1,\ldots,\mathcal{A}_r\}}\Sigma_r|_{\mathcal{O}_r}\right)}.$$

In other words, it suffices to understand the **derivative of** Σ_r:

$$\left(\mathrm{d}_{\{\mathcal{A}_1,\ldots,\mathcal{A}_r\}}\Sigma_r|_{\mathcal{O}_r}\right)(\dot{\mathcal{A}}_1,\ldots,\dot{\mathcal{A}}_r) = \dot{\mathcal{A}}_1 + \cdots + \dot{\mathcal{A}}_r.$$

Representing this derivative in coordinates yields the so-called **Terracini matrix**:[4]

$$T_{\mathcal{A}_1,\ldots,\mathcal{A}_r} = \begin{bmatrix} Q_1 & Q_2 & \cdots & Q_r \end{bmatrix},$$

where Q_i is a matrix whose columns contain an orthonormal **basis of the tangent space** $\mathrm{T}_{\mathcal{A}_i}\mathcal{S}$.

[4] For more details see Breiding and V (2018a).

In conclusion, Breiding and V (2018a) proved the following

Theorem

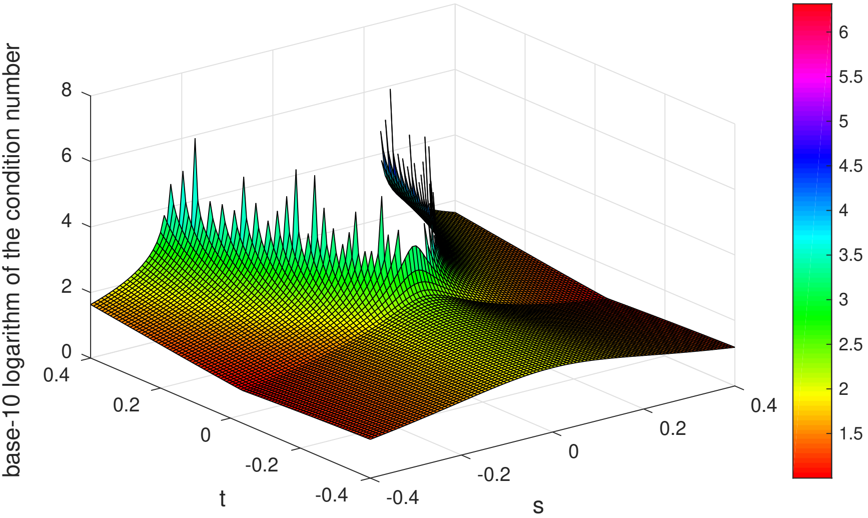
The condition number of computing the tensor rank decomposition of $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_r$ is

$$\kappa[\tau_r](\mathcal{A}) = \frac{1}{\sigma_{r \dim \mathcal{S}}(T_{\mathcal{A}_1, \ldots, \mathcal{A}_r})}$$

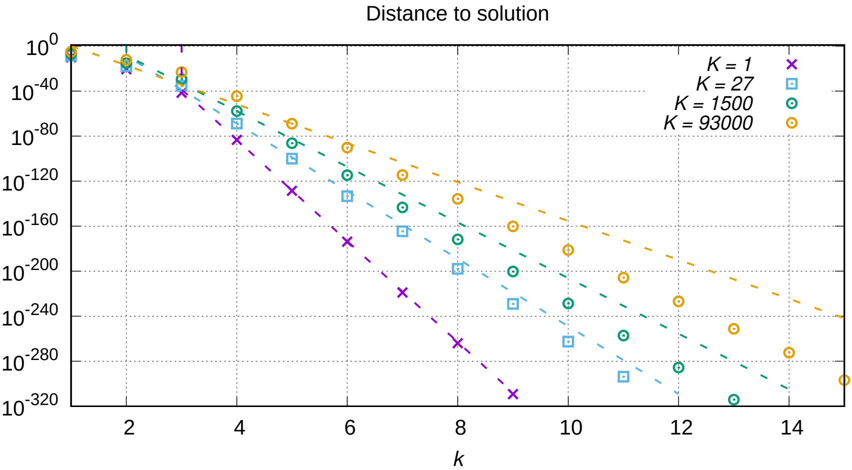
and ∞ otherwise, by convention.

IV. A gallery of properties

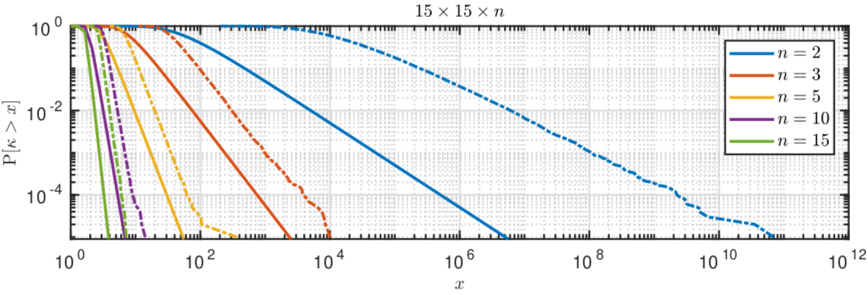
The condition number of rank-r tensors **blows up** to ∞ near the border with tensors of higher rank (but border rank r) (Breiding and V, 2018a):



The condition number controls the **rate of convergence in Riemannian optimization** methods (Breiding and V, 2018b):

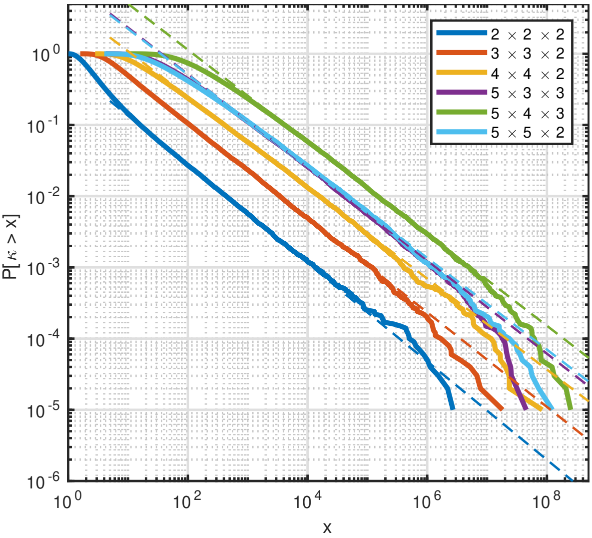


The condition number explains the **numerical instability of pencil-based algorithms** (Beltrán, Breiding and V, 2019a):

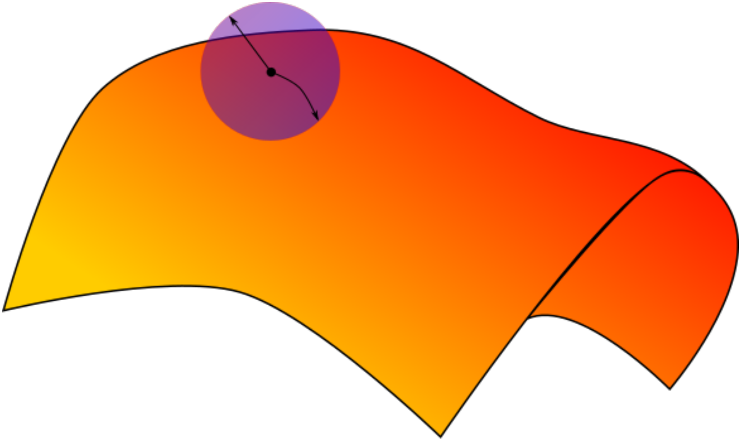


(a) A, B, and C i.i.d. standard normal entries.

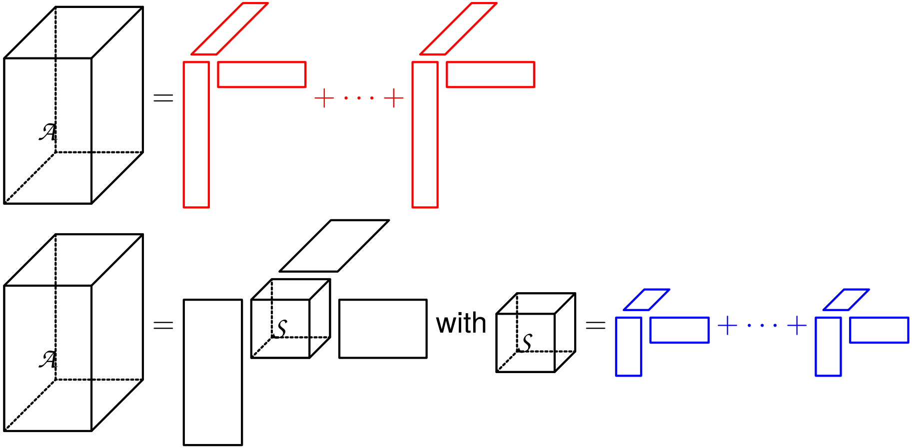
The **expected value** of the condition number for random rank-r tensors **is** ∞ (Beltrán, Breiding and V, 2019b):



The condition numbers of rank-r **approximation** and **decomposition** are equal at a rank-r tensor $\mathcal{A}$. **General perturbations do not increase the condition number** (Breiding and V, 2021):



The condition number is **invariant under orthogonal Tucker compression** (Dewaele, Breiding and V, 2021):



Then $\kappa[\tau_r](\mathcal{A}) = \kappa[\tau_r](\mathcal{S})$.

V. Conclusions

The condition number of tensor join decomposition helps us understand the sensitivity of the output to input perturbations and complexity of solving join decomposition problems.

If you want to know more about identifiability and sensitivity, check out the IPAM seminar series talks:

- Luca Chiantini, May 10
- Paul Breiding, May 12
- NV, May 12

Thanks for your attention!

References

- Armentano, *Stochastic perturbations and smooth condition numbers*, J. Complexity, 2010.
- Beltrán, Breiding, Vannieuwenhoven, *Pencil-based algorithms for tensor rank decomposition are not stable*, SIAM J. Matrix Anal. Appl., 2019a.
- ——, *The average condition number of most tensor rank decomposition problems is infinite*, arXiv:1903.05527, 2019b.
- Bocci, Chiantini, Ottaviani, *Refined methods for the identifiability of tensors*, Ann. Mat. Pura Appl., 2013.
- Breiding, Vannieuwenhoven, *The condition number of join decompositions*, SIAM J. Matrix Anal. Appl., 2018a.
- ——, *Convergence analysis of Riemannian Gauss-Newton methods and its connection with the geometric condition number*, Appl. Math. Letters, 2018b.
- ——, *The condition number of Riemannian approximation problems*, SIAM J. Optim., 2021.
- Bürgisser, Cucker, *Condition: The Geometry of Numerical Algorithms*, Springer, 2013.

- Dewaele, Breiding, Vannieuwenhoven, *work in progress*, 2021.
- Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 1927.
- Chiantini, Ottaviani, Vannieuwenhoven, *An algorithm for generic and low-rank specific identifiability of complex tensors*, SIAM J. Matrix Anal. Appl., 2014.
- ——, *Effective criteria for specific identifiability of tensors and forms*, SIAM J. Matrix Anal. Appl., 2017.
- Kato, *Perturbation Theory for Linear Operators*, Springer, 1995.
- Rice, *A theory of condition*, SIAM J. Numer. Anal., 1966.

$\mathbb{R}^{n_1 \times \cdots \times n_d}$ is called **generically complex r-identifiable** if the set of rank-r complex tensors that are not r-identifiable is contained in a strict closed subvariety.

It is **conjectured**[5] 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**, modulo a few exceptions, is:

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

[5] Bocci, Chiantini, Ottaviani (2014); Chiantini, Ottaviani, V (2014, 2017)