From low rank tensors for grids of multivariate functions to joint distributions of statistical models: selecting the relevant norm.

Alain Franc

INRAE BioGeCo & INRIA Pleiade

July 8, 2021

INRIA Skoltech workshop

Three objects behind a tensor

$$\mathbf{A} \in E_1 \otimes \ldots \otimes E_d \equiv \mathbb{R}^{n_1 \times \ldots n_d}$$

discretization of a multivariate function

joint law of a statistical model

array of data

Preliminaries

CP and TT decomposition

CP decomposition

$$\mathbf{A} = \sum_{\alpha=1}^{r} \mathbf{x}_\alpha \otimes \mathbf{y}_\alpha \otimes \mathbf{z}_\alpha \qquad \in \mathbb{R}^{n \times n \times n}$$

$$r_{\mathrm{cp}}(\mathbf{A}) = \inf \left\{ r \in \mathbb{N} \mid \exists\, \mathbf{A} = \sum_{\alpha=1}^{r} \dots \right\}$$

$3nr$ terms

TT decomposition

$$\mathbf{A}[i,j,k] = \mathbf{u}[i].G[j].\mathbf{v}[k]$$

$$r_{\mathrm{tt}} \ : \ \mathbf{u}_i \in \mathbb{R}^{1 \times r}, \quad G_j \in \mathbb{R}^{r \times r}, \quad \mathbf{v}_k \in \mathbb{R}^{r \times 1}$$

$nr^2 + 2nr$ terms

Link between ranks

Lemma (can be generalized in a strightforward way)

Let

$$\mathbf{A} \in E \otimes F \otimes G, \qquad \text{with} \quad \begin{cases} \dim E & = m \\ \dim F & = n \\ \dim G & = p \end{cases}$$

Then

$$r_{tt}(\mathbf{A}) \leq r_{cp}(\mathbf{A}) \leq r_{tt}^2(\mathbf{A})$$

Proof

- Development on a basis and reorganisation for $r_{cp} \leq r_{tt}^2$
- Development with diagonal G matrix for $r_{tt} \leq r_{cp}$

Example

For example, if $r = 2$

$$\mathbf{A} = \mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1 + \mathbf{x}_2 \otimes \mathbf{y}_2 \otimes \mathbf{z}_2$$

$$\text{with} \qquad \mathbf{x}_1 = \left(x_i^{(1)}\right)_i, \quad \mathbf{y}_1 = \left(y_j^{(1)}\right)_j, \dots$$

Then

$$\begin{aligned} x_i^{(1)} y_j^{(1)} z_k^{(1)} + x_i^{(2)} y_j^{(2)} z_k^{(2)} &= \begin{pmatrix} x_i^{(1)} & x_i^{(2)} \end{pmatrix} \begin{pmatrix} y_j^{(1)} z_k^{(1)} \\ y_j^{(2)} z_k^{(2)} \end{pmatrix} \\ &= \begin{pmatrix} x_i^{(1)} & x_i^{(2)} \end{pmatrix} \begin{pmatrix} y_j^{(1)} & 0 \\ 0 & y_j^{(2)} \end{pmatrix} \begin{pmatrix} z_k^{(1)} \\ z_k^{(2)} \end{pmatrix} \\ &= \mathbf{u}_i . G_j . \mathbf{v}_k \end{aligned}$$

When $r_{\text{cp}} = r_{\text{tt}}$?

TT rank is a tensor property (whatever the basis), and

$$\begin{aligned} u_i.G_j.v_k &= u_i.(PP^{-1}).G_j.(QQ^{-1}).v_k \\ &= (u_i.P)(P^{-1}.G_j.Q)(Q^{-1}.v_k) \\ &= u'_i G'_j v'_k \end{aligned}$$

Simultaneously diagonalizable matrices

- $(G_j)_{1 \leq j \leq n}$ is a set of $r \times r$ diagonalizable matrices

Then, it is equivalent that

- $\forall\, (i,j), \quad G_i G_j = G_j G_i$
- There exists a basis P such that each G_j is diagonal in this basis

Lemma (refined)

$$\text{Let} \qquad \mathbf{A} \in \mathbb{R}^{m \times n \times p} \qquad \text{with} \quad a_{ijk} = u_i.G_j.v_k \,,$$

Lemma

- $r_{\mathrm{tt}} \leq r_{\mathrm{cp}} \leq r_{\mathrm{tt}}^2$
- $r_{\mathrm{cp}} = r_{\mathrm{tt}}$ iff $(G_j)_j$ simultaneously diagonalisable

Take home message: very simple

If a tensor has low CP rank, then it has low TT rank.

A tensor as discretisation of a multivariate function

Three objects behind a function

Multivariate function (here d=2 for simplicity)

Let

$$\mathbb{R}^2 \xrightarrow{f} \mathbb{R}$$

Three implementations

$$\begin{tikzcd}
 & & f \in \mathcal{H} \\
f \arrow[r] & \bullet \arrow[ur] \arrow[r] \arrow[dr] & f(x,y) = x^2 + 2xy + y^2 \\
 & & f(x_i, y_i) = M[i,j]
\end{tikzcd}$$

Cartesian Mesh

1D mesh

$$\mathbf{x} \in \mathbb{R}^n, \qquad \forall\, i, \quad x_i \in \mathbb{R}$$

$$\mathbf{x} = (x_1, \ldots, x_n) \qquad \text{with} \quad x_1 < \ldots < x_n$$

$$\mathscr{D}(f, \mathbf{x}) = (f(x_1), \ldots, f(x_n))$$

$$\mathcal{H} \times \mathcal{M} \xrightarrow{\mathscr{D}} \mathbb{R}^n$$

2D Cartesian Mesh

- Let us consider 2 meshes $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}$ of respective sizes m, n.
- Then $\mathbf{x} \otimes_{\mathrm{m}} \mathbf{y}$ in $\mathbb{R}^2$ can be defined as

$$\mathbf{x} \otimes_{\mathrm{m}} \mathbf{y} = \begin{pmatrix} (x_1, y_1) & (x_1, y_2) & \ldots & (x_1, y_n) \\ (x_2, y_1) & (x_2, y_2) & \ldots & (x_2, y_n) \\ \vdots & \vdots & & \vdots \\ (x_m, y_1) & (x_m, y_2) & \ldots & (x_m, y_n) \end{pmatrix}$$

Discretization of product of functions

Product of functions: two variables

$$
\begin{array}{lrcl}
\bullet & (fg)(x) & = & f(x)g(x) \\
\bullet & (f \otimes g)(x,y) & = & f(x)g(y)
\end{array}
$$

Discretisation of product of functions

$$
\begin{array}{lrcl}
\bullet & \mathscr{D}(fg, \mathbf{x}) & = & \mathscr{D}(f, \mathbf{x}) \odot \mathscr{D}(g, \mathbf{x}) \\
\bullet & \mathscr{D}(f \otimes g, \mathbf{x} \otimes_{\mathrm{m}} \mathbf{y}) & = & \mathscr{D}(f, \mathbf{x}) \otimes \mathscr{D}(g, \mathbf{y})
\end{array}
$$

Consequence

If $u = f \otimes g$ [$u(x,y) = f(x)g(y)$, separation of variables], the matrix $\mathscr{D}(u, \mathbf{x} \otimes_m \mathbf{y})$ of discretization of u on mesh $\mathbf{x} \otimes_m \mathbf{y}$ has rank one too.

Product of d functions

Generalization is straightforward ...

$$\begin{aligned}
\mathscr{D}\left(\prod_\mu f_\mu \, , \, \mathbf{x}\right) &= \bigodot \mathscr{D}(f_\mu, \mathbf{x}) \\
\mathscr{D}\left(\bigotimes_\mu f_\mu \, , \, \otimes_\mu^{(m)} \, \mathbf{x}_\mu\right) &= \bigotimes_\mu \mathscr{D}(f_\mu, \mathbf{x}_\mu)
\end{aligned}$$

An elementary lemma (can be extended to $d > 2$)

Let

$$\begin{array}{ccc} \mathcal{H} \otimes \mathcal{H} & \xrightarrow{\mathscr{D}} & \mathbb{R}^{m \times n} \\ \psi & \longrightarrow & \mathscr{D}(\psi\,, \mathbf{x} \otimes_{\mathrm{m}} \mathbf{y}) \end{array}$$

Let (rank r $CP-$decomposition)

$$\psi = \sum_{\alpha=1}^{r} \mathbf{u}_\alpha \otimes \mathbf{v}_\alpha \qquad \text{with} \quad \mathbf{u}_\alpha, \mathbf{v}_\alpha \in \mathcal{H}$$

Then

$$\forall\, m, n, \quad r_{\mathrm{cp}}\left(\mathscr{D}(\psi\,, \mathbf{x} \otimes_{\mathrm{m}} \mathbf{y})\right) \leq r_{\mathrm{cp}}(\psi)$$

As consequence ...

$$\forall\, m, n, \quad r_{\mathrm{tt}}\left(\mathscr{D}(\psi\,, \mathbf{x} \otimes_{\mathrm{m}} \mathbf{y})\right) \leq r_{\mathrm{cp}}(\psi)$$

Discretization of polynomials (1/2)

Let us define

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{\mathbf{1}_x} & \mathbb{R} \\ x & \longrightarrow & 1 \end{array}$$

Bivariate polynomial

$$P(x,y) = x^2 + 2xy + y^2$$

Then

$$P = x^2 \otimes \mathbf{1}_y + 2\, x \otimes y + \mathbf{1}_x \otimes y^2$$

and

$$r_{\mathrm{cp}}(P) \leq 3$$

and

$$\forall\, m, n \in \mathbb{N}, \quad r_{\mathrm{tt}}(\mathscr{D}(P, \mathbf{x} \otimes_{\mathrm{m}} \mathbf{y})) \leq r_{\mathrm{cp}}(\mathscr{D}(P, \mathbf{x} \otimes_{\mathrm{m}} \mathbf{y})) \leq 3$$

Discretization of polynomials (2/2)

Any polynomial

$$\mathbf{P}(x_1, \ldots, x_d) = \sum_{\mathbf{n}} a_{n_1 \ldots n_d} \, x_1^{n_1} \ldots x_d^{n_d}, \qquad N \text{ terms}$$

Then

$$\mathbf{P} = \sum_{\mathbf{n}} a_{n_1 \ldots n_d} \, x_1^{n_1} \otimes \ldots \otimes x_d^{n_d}, \quad \text{CP rank} = N$$

and

$$\forall \, m_1, \ldots, m_d \in \mathbb{N}, \quad r_{\mathrm{tt}} \left(\mathscr{D}(\mathbf{P}, \mathbf{x}_1 \otimes_{\mathrm{m}} \ldots \otimes_{\mathrm{m}} \mathbf{x}_d) \right) \leq N$$

- m_μ is the size of the Cartesian Grid for mode μ.

a remark

In general, $N = \prod_\mu n_\mu$
$\Longrightarrow$ rank is high if d is significant

Extension to continuous functions

Stone-Weierstrass theorem

- $C(S) = \{f \ : \ S \longrightarrow \mathbb{R}, \quad f \text{ continuous}\}$
- $A \subset C(S)$ s.t. $f, g \in A \Rightarrow fg \in S$
- $x \neq y \quad \Longrightarrow \quad \exists\, f \in A \ : \ f(x) \neq f(y)$

Then

$$\forall\, \epsilon > 0, \quad \forall f \ \in C(S), \quad \exists\, \varphi \in A \ : \ \forall\, x \in S, \quad |f(x) - \varphi(x)| < \epsilon$$

- Any continuous function can be approximated with norm ℓ^∞ as close as wished by a polynomial, i.e. a low rank function
- This is automatially transported to approximation by low rank discretization on Cartesian grids
- And from CP to TT approximation (for working with TT toolbox)
- A well developed theory has been elaborated for ℓ^2 norm as well (development on basis of orthogonal polynomials, leading to Tucker approximations).

Take home message

There is a sound and standard algebraic theory for showing that tensors as discretization of multivariate functions are

- exactly low rank for polynomials
- well approximated by low rank tensors (CP, TT, Tucker) for continuous functions

Some details still deserve attention

- Better understanding of the link between CP et TT decomposition
  - Why CP is numerically unstable sometimes and TT not ?
  - at which boundaries ?
- is it expandable to non Cartesian meshes?

A tensor as a joint law

A tensor as a joint law

Setting

- Let us have a discrete set Λ
- d random variables X_μ with values in Λ

Define

$$\mathbf{T}[i_1, \ldots, i_d] = t_{i_1 \ldots i_d} \propto \mathbb{P}(X_1 = i_1, \ldots, X_d = i_d)$$

- $\mathbf{T}$ is a $d-$modes tensor with elements ≥ 0

Classical examples

- Ising model
- Graphical models
- $\ldots$

Partition function

Definition

In statistical physics, and statistical modeling, one is led to compute

$$Z(\mathbf{T}) = \sum_{i_1} \cdots \sum_{i_d} t_{i_1 \ldots i_d}$$

- requires n^d additions, with $d > 10^3$ often .... see Novikov & al. (2014) for large d.

Alexander Novikov & al. (2014) - Putting MRFs on a Tensor Train. *Proceedings of the 31 st International Conference on Machine Learning*, Beijing, China.

When does it work?

When $\mathbf{T}$ is written in TT format

$$t_{i_1 \dots i_d} = G_1(i_1) \times \dots \times G_\mu(i_\mu) \times \dots \times G_d(i_d)$$

with $G_\mu(i_\mu) \in \mathbb{R}^{r_{\mu-1} \times r_\mu}$. Then, if $B_\mu = \sum_{i_\mu} G_\mu(i_\mu)$

$$Z(\mathbf{T}) = B_1 B_2 \dots B_d$$

(Novikov & al., 2014)

When $\mathbf{T}$ has low CP-rank

$$\mathbf{T} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} + \mathbf{x}' \otimes \mathbf{y}' \otimes \mathbf{z}'$$

Observation

- All terms in $\mathbf{T}$ are ≥ 0 Then

$$Z(\mathbf{T}) = \|\mathbf{T}\|_1 \qquad (\text{norm } \ell^1)$$

Variational approach

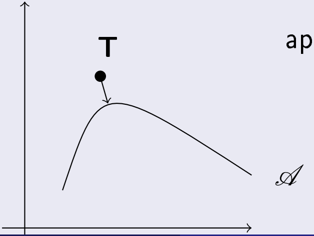
General approach

- a quantity $f(\mathbf{T})$ on a tensor (like $Z(\mathbf{T})$) is difficult to compute in general, but easy in a (closed) subset $\mathscr{A} \subset \mathbb{R}^{n_1 \times \ldots \times n_d}$
- Then, one computes $\widehat{\mathbf{T}} \in \mathscr{A}$ such that, for a selected distance δ

$$\delta(\mathbf{T}, \widehat{\mathbf{T}}) \quad \text{is minimal}$$

- and approximates

$$f(\mathbf{T}) \approx f(\widehat{\mathbf{T}})$$



approaching $Z(\mathbf{T})$ with ℓ^1 norm

Low rank approximation of matrices for ℓ^1 norm

Setting the problem

- given $A \in \mathbb{R}^{m \times n}$
- find $\widehat{A}$ with $\text{rank}(\widehat{A}) = r$
- such that $\|A - \widehat{A}\|_1$ minimal

State of the art

- Difficulty: it has been shown to be NP-hard (Gillis & Vavasis, 2015)
- Available algorithm with "provable approximation guarantees" (Song, Woodruff & Zhong, 2018)

What about tensors?

- Let $\mathbf{A} \in E \otimes F \otimes G$
- Let A_E be its first matricization

$$F \otimes G \xrightarrow{A_E} E$$

- Then $\|A_E\|_1 = \|\mathbf{A}\|_1$

Proposed heuristics

for all matricizations A of $\mathbf{A}$ **do**

  computes the best rank one approximation A_{best} of A with norm ℓ^1

  computes $\delta(A) = \|A - A_{best}\|$

select A such that $\delta(A)$ is minimal

computes $Z(A_{best})$ (easy)

return $Z(A_{best}) \approx Z(\mathbf{A})$

Take home message

- The theory behind best low rank approximation of a tensor as joint distribution is not as mature than links between multilinear algebra, PDE, functional analysis
- A difficulty is that a wider diversity of norms is relevant
  - Kullback-Leibler for mutual information
  - ℓ^1 for partition function
  - ...

  which do not rely on Euclidean geometry

To be taken into account: different levels of complexity

rank		**Kolmogorov complexity**
high	**Randomness**	high
medium	**Data structure**	medium
	Statistical modeling	
low	**Discretized functions**	low

Acknowledgements

Thanks to following colleagues for helpful discussions

- Mohamed Anwar Abouabdallah, PhD Student
- Olivier Coulaud, INRIA Hiepacs
- Martina Iannacito, PhD student
- Nathalie Peyrard, INRAE MIAT, Toulouse

and Olivier Beaumont for the proposition to join the team!