An algebraic algorithm for blind source separation and tensor decomposition

Martina Iannacito

joint work with Ignat Domanov and Lieven De Lathauwer

New directions in linear algebra
BIRS, Canada, August 31, 2023

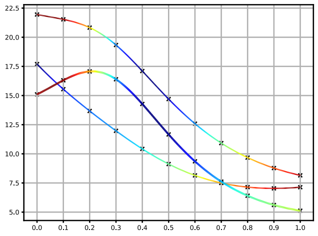
KU LEUVEN kulak

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

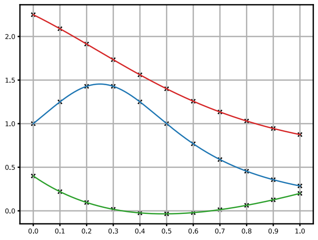
erc

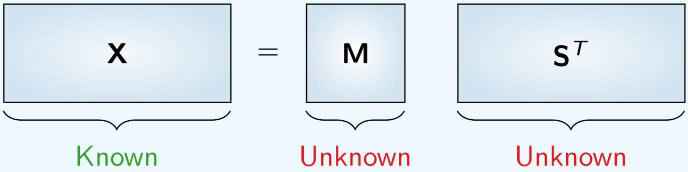
Overview

Blind Source Separation problem



$$= \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} \times$$





Overview

Constraints for uniqueness

Definition: A *deterministic* condition on $\mathbf{X}$ imposes a particular property of $\mathbf{X}$ that is always true.

Definition: A *generic* condition on $\mathbf{X}$ depending on a parameter $\mathbf{z} \in \Omega$ holds almost everywhere, i.e., if the condition doesn't hold for $\mathbf{z} \in \Sigma \subset \Omega$, then $\mu(\Sigma) = 0$ with μ a measure absolute continuous w.r.t. the Lebesgue one.

Deterministic conditions

- Statistical independence → Independent Component Analysis

$$\mathbf{X} = \mathbf{M} \, \underset{\text{Ind}}{\mathbf{S}^T}$$

- Nonnegativity → Nonnegative Matrix Factorization

$$\mathbf{X} = \underset{+}{\mathbf{M}} \, \underset{+}{\mathbf{S}^T}$$

- Sparsity → Sparse Component Analysis

$$\mathbf{X} = \mathbf{M} \, \underset{\text{Max0}}{\mathbf{S}^T}$$

- ...

[Comon and Jutten 2009]

General case

$$\mathbf{X} = \mathbf{M} \quad \mathbf{S}^T$$

$$\underbrace{\mathbf{X}}_{\mathbf{X}} = \underbrace{\mathbf{M} \quad \mathbf{A}}_{\mathbf{M}_1} \quad \underbrace{\mathbf{A}^{-1} \quad \mathbf{S}^T}_{\mathbf{S}_1^T}$$

Uniqueness isn't guaranteed!

Overview

The Blind Source Separation

- Deterministic uniqueness
- Generic uniqueness

The Canonical Polyadic Decomposition

- From the theorem to the algorithm
- The bottleneck

Algorithm improvements

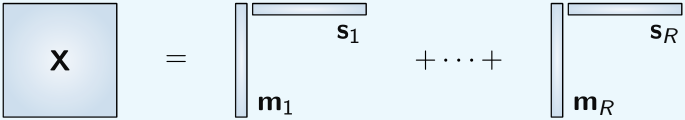
- Exterior algebra
- Numerical results

Constraints for uniqueness

Definition: A *deterministic* condition on $\mathbf{X}$ imposes a particular property of $\mathbf{X}$ that is always true.

Definition: A *generic* condition on $\mathbf{X}$ depending on a parameter $\mathbf{z} \in \Omega$ holds almost everywhere, i.e., if the condition doesn't hold for $\mathbf{z} \in \Sigma \subset \Omega$, then $\mu(\Sigma) = 0$ with μ a measure absolute continuous w.r.t. the Lebesgue one.

Problem statement



$$\mathbf{X} = \mathbf{M}(\mathbf{z})\mathbf{S}^T(\mathbf{z}) = \sum_{r=1}^{R} \mathbf{m}_r(\mathbf{z}) \otimes \mathbf{s}(\boldsymbol{\xi}_r)$$

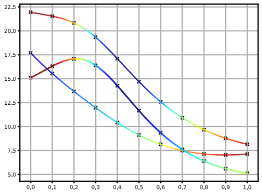
where

- $\mathbf{z} \in \Omega$ a subset of $\mathbb{R}^n$
- $\mathbf{m}_r(\mathbf{z})$ are linearly independent
- each $\mathbf{s}_r(\mathbf{z})$ depends on ℓ independent parameters, entries of $\boldsymbol{\xi}_r \in \mathbb{R}^\ell$
- each $\mathbf{s}_r(\mathbf{z}) = \mathbf{s}_r(\boldsymbol{\xi}_r)$ has the structure

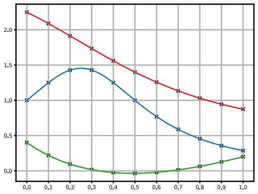
$$\mathbf{s}(\boldsymbol{\xi}_r) = \left[\frac{p_1}{q_1} \circ \mathbf{f}(\boldsymbol{\xi}_r) \quad \ldots \quad \frac{p_N}{q_N} \circ \mathbf{f}(\boldsymbol{\xi}_r) \right]$$

with p_h, q_h polynomials and $\mathbf{f} = [f_1, \ldots, f_\ell]$ a vectorial function.

Example



$$= \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} \times$$



Given the observed mixtures, we assume that

- the mixture matrix $\mathbf{M}$ is constant and full rank;
- the source signals can be modeled by rational functions, i.e., the columns of $\mathbf{S}$ are sampled of

$$s(t) = \frac{a_0 + a_1 t + \cdots + a_p t^p}{b_0 + b_1 t + \cdots + b_q t^q} \qquad \text{with} \qquad a_i, b_i \in \mathbb{R}, \quad t \in [t_b, t_e]$$

⇓

- $\boldsymbol{\xi} = [a_0, \ldots, a_p, b_0, \ldots, b_q]$
- $\ell = p + q + 2$
- $\mathbf{f}$ is the identity

A cookbook recipe for generic uniqueness [Domanov and De Lathauwer 2016]

Let $\mathbf{t}(\mathbf{x}) = \begin{bmatrix} \frac{p_1}{q_1}(\mathbf{x}) & \dots & \frac{p_N}{q_N}(\mathbf{x}) \end{bmatrix}^T$ for $\mathbf{x} \in \Theta = \{\mathbf{x} \in \mathbb{C}^\ell : q_1(\mathbf{x}) \cdots q_N(\mathbf{x}) \neq 0\}$, if

1. $\mathrm{rank}\mathbf{M}(z) = R$ for a generic choice of $\mathbf{z}$
2. each f_h is the ratio of two analytical functions on $\mathbb{C}^\ell$
3. there exists $\boldsymbol{\xi}_0 \in \mathbb{C}^\ell$ s.t. $\det\mathbf{J}(\mathbf{f}, \boldsymbol{\xi}_0) \neq 0$
4. the dimension of the span of $\mathbf{t}(\mathbf{x})$ for $x \in \Theta$ is at least $\hat{N}$
5. $\mathrm{rank}\mathbf{J}(\mathbf{t}, \mathbf{x}) > \hat{\ell}$ for a generic choice of $\mathbf{z}$
6. $R \leq \hat{N} - \hat{\ell}$

then

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{m}_r(\mathbf{z}) \otimes \mathbf{s}(\boldsymbol{\xi}_r)$$

is generically unique.

Remarks for BSS

It is assumed that the columns of $\mathbf{S}$ are values of the rational function

$$\mathbf{t} : \mathbf{x} \to \begin{bmatrix} \frac{p_1}{q_1}(\mathbf{x}) & \dots & \frac{p_N}{q_N}(\mathbf{x}) \end{bmatrix}^T .$$

The columns of $\mathbf{S}$ belong to an algebraic variety $\mathcal{V}$ which is described by a finite system of polynomials $\{P_k\}_{k=1}^K$

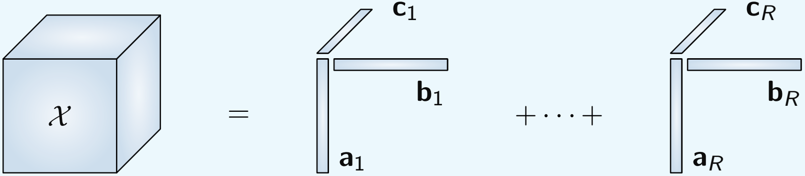
$$\mathcal{V} = \Big\{(z_1, \dots, z_N) \in \mathbb{C}^N : P_k(z_1, \dots, z_N) = 0\Big\}.$$

Composed with the function $\mathbf{f}$

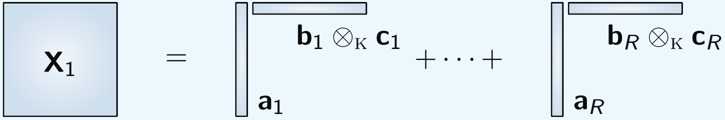
Overview

Link with the CPD

$$\mathcal{X} = \mathbf{a}_1 \otimes \mathbf{b}_1 \otimes \mathbf{c}_1 + \ldots + \mathbf{a}_R \otimes \mathbf{b}_R \otimes \mathbf{c}_R$$



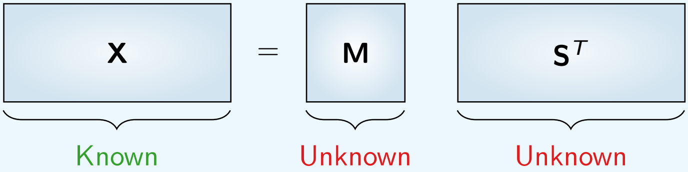
$$\mathbf{X}_1 = \mathbf{a}_1 \otimes (\mathbf{b}_1 \otimes_{\mathrm{K}} \mathbf{c}_1)^T + \ldots + \mathbf{a}_R \otimes (\mathbf{b}_R \otimes_{\mathrm{K}} \mathbf{c}_R)^T$$



$$(\mathbf{b}_r \otimes_{\mathrm{K}} \mathbf{c}_r) \in \mathcal{V} = \left\{ \mathrm{vec}(\mathbf{Z}) : \begin{vmatrix} z_{i_1 j_1} & z_{i_1 j_2} \\ z_{i_2 j_1} & z_{i_2 j_2} \end{vmatrix} = 0 \right\}$$

Overview

Algebraic algorithm outline



1. compute $\mathbf{M}^{-1}$ from $\mathbf{X}$;
2. compute $\mathbf{S}$ as $\mathbf{M}^{-1}\mathbf{X}$ transposed

Equivalent condition I

$\mathbf{a}$ is a column of $\mathbf{M}^{-1}$ if and only if $\mathbf{X}^T\mathbf{a}$ is equal to a column of $\mathbf{S}$

⇓

$$\mathbf{X}^T\mathbf{a} = (\mathbf{x}_1^T\mathbf{a}, \ldots, \mathbf{x}_N^T\mathbf{a}) = (z_1, \ldots, z_N) \in \mathcal{V}$$

⇕

$$P_k(\mathbf{x}_1^T\mathbf{a}, \ldots, \mathbf{x}_N^T\mathbf{a}) = 0 \text{ for } k = 1, \ldots, K$$

⇕

$$P_k^{\otimes}(\mathbf{x}_1^T, \ldots, \mathbf{x}_N^T)(\mathbf{a} \otimes \cdots \otimes \mathbf{a}) = 0 \text{ for } k = 1, \ldots, K$$

where $P_k^{\otimes}(\mathbf{x}_1^T, \ldots, \mathbf{x}_N^T)$ is the vector obtained by formal substitution of $(z_1, \ldots, z_N)$ by $\mathbf{x}_1^T, \ldots, \mathbf{x}_N^T$ and the scalar multiplication by the tensor product.

Equivalent condition II

$\mathbf{a}$ is a column of $\mathbf{M}^{-1}$ if and only if $\mathbf{X}^T\mathbf{a}$ is equal to a column of $\mathbf{S}$

⇓

$$P_k^{\otimes}(\mathbf{x}_1^T, \dots, \mathbf{x}_N^T)(\mathbf{a} \otimes \cdots \otimes \mathbf{a}) = 0 \text{ for } k = 1, \dots, K$$

⇕

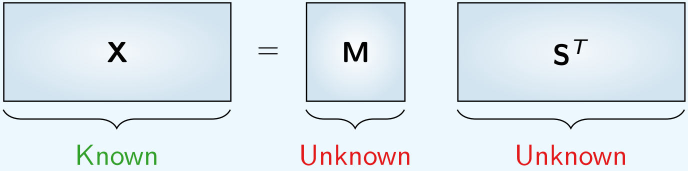
$$\mathbf{Q}\text{vec}(\mathbf{a}^{\otimes p}) = \begin{bmatrix} P_1^{\otimes}(\mathbf{x}_1^T, \dots, \mathbf{x}_N^T) \\ \vdots \\ P_K^{\otimes}(\mathbf{x}_1^T, \dots, \mathbf{x}_N^T) \end{bmatrix} \text{vec}(\mathbf{a} \otimes \cdots \otimes \mathbf{a}) = 0$$

⇕

The columns of $\mathbf{M}^{-1}$ belong to the intersection of $\mathbf{Q}$ kernel and the subspace of vectorized order p symmetric tensors.

Algebraic algorithm outline

[Domanov and De Lathauwer 2013]

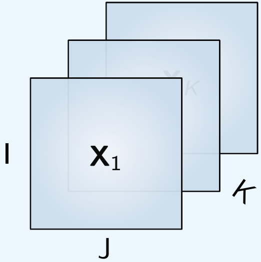


1. compute $\mathbf{M}^{-1}$ from $\mathbf{X}$;
   1.1 compute $\mathbf{Q}$;
   1.2 compute null($\mathbf{Q}$) intersected with the space of vectorized symmetric tensors;
2. compute $\mathbf{S}$ as $\mathbf{M}^{-1}\mathbf{X}$ transposed.

Overview

$\mathbf{Q}$ for CPD

Let $\mathcal{X}$ be an order-3 tensor of dimension $(I \times J \times K)$, then



Definition: $\mathbf{Q}$ is a $C_I^2 C_J^2 \times C_{K+1}^2$ matrix whose k-th column can be written as

$$\mathbf{Q}(\cdot, k) = \text{vec}\big(\mathcal{C}_2(\mathbf{X}_{k_1} + \mathbf{X}_{k_2}) - \mathcal{C}_2(\mathbf{X}_{k_1}) - \mathcal{C}_2(\mathbf{X}_{k_2})\big)$$

where

- (k_1, k_2) is the k-th element of $\mathcal{Q}_K^2 = \{(k_1, k_2) : 1 \leq k_1 \leq k_2 \leq K\}$;
- C_N^2 is the binomial of N over 2;
- $\mathcal{C}_2(\mathbf{X}_h)$ is the matrix with the determinants of every 2×2 minors of $\mathbf{X}_h$.

Overview

Example

Let $\mathcal{X}$ be an order-3 tensor of dimension $(I \times 3 \times 2)$ such that

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix}$$

The matrix $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix}$ associated with $\mathcal{X}$ slices is such that

- $\mathbf{q}_1 = \text{vec}(\mathcal{C}_2(\mathbf{X}_1 + \mathbf{X}_1) - \mathcal{C}_2(\mathbf{X}_1) - \mathcal{C}_2(\mathbf{X}_1)) = 2\text{vec}(\mathcal{C}_2(\mathbf{X}_1))$
- $\mathbf{q}_2 = \text{vec}(\mathcal{C}_2(\mathbf{X}_1 + \mathbf{X}_2) - \mathcal{C}_2(\mathbf{X}_1) - \mathcal{C}_2(\mathbf{X}_2))$
- $\mathbf{q}_3 = \text{vec}(\mathcal{C}_2(\mathbf{X}_2 + \mathbf{X}_2) - \mathcal{C}_2(\mathbf{X}_2) - \mathcal{C}_2(\mathbf{X}_2)) = 2\text{vec}(\mathcal{C}_2(\mathbf{X}_2))$

The compound matrices are

- $\mathcal{C}_2(\mathbf{X}_1) = \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \boldsymbol{\alpha}_3 \end{bmatrix}$
- $\mathcal{C}_2(\mathbf{X}_2) = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \boldsymbol{\beta}_3 \end{bmatrix}$
- $\mathcal{C}_2(\mathbf{X}_1 + \mathbf{X}_2) = \begin{bmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \boldsymbol{\gamma}_3 \end{bmatrix}$

$$\Rightarrow \quad \mathbf{Q} = \begin{bmatrix} 2\boldsymbol{\alpha}_1 & \boldsymbol{\gamma}_1 - (\boldsymbol{\alpha}_1 + \boldsymbol{\beta}_1) & 2\boldsymbol{\beta}_1 \\ 2\boldsymbol{\alpha}_2 & \boldsymbol{\gamma}_2 - (\boldsymbol{\alpha}_2 + \boldsymbol{\beta}_2) & 2\boldsymbol{\beta}_2 \\ 2\boldsymbol{\alpha}_3 & \boldsymbol{\gamma}_3 - (\boldsymbol{\alpha}_3 + \boldsymbol{\beta}_3) & 2\boldsymbol{\beta}_3 \end{bmatrix}$$

Overview

Exterior algebra and $\mathcal{C}_2$

Definition: The exterior product $\wedge : \mathbb{R}^I \times \mathbb{R}^J \to \mathbb{R}^{I \times J}$ is defined as

$$\mathbf{a} \wedge \mathbf{b} = \mathbf{a}\mathbf{b}^T - \mathbf{b}\mathbf{a}^T.$$

Example

Let $\mathbf{a}^T = \begin{bmatrix} a & b & c \end{bmatrix}$ and $\mathbf{b}^T = \begin{bmatrix} d & e & f \end{bmatrix}$ be two vectors, then

$$\mathbf{a} \wedge \mathbf{b} = \begin{bmatrix} 0 & ae - bd & af - cd \\ -ae + bd & 0 & bf - ce \\ -af + cd & -bf + ce & 0 \end{bmatrix}$$

The matrix $\mathbf{a} \wedge \mathbf{b}$ has the properties:

- the diagonal entries are zeros;
- it is skew-symmetric;
- the elements highlighted are the entries of $\mathcal{C}_2([\mathbf{a} \quad \mathbf{b}])$;

Algebraic algorithm optimization steps

Property: $\left\langle \text{vec}(\mathcal{C}_2([\mathbf{a} \quad \mathbf{b}])), \text{vec}(\mathcal{C}_2([\mathbf{a} \quad \mathbf{b}])) \right\rangle = 2\left\langle \text{vec}(\mathbf{a} \wedge \mathbf{b}), \text{vec}(\mathbf{a} \wedge \mathbf{b}) \right\rangle$

$$= 4||\mathbf{a}||^2||\mathbf{b}||^2 - \langle \mathbf{a}, \mathbf{b} \rangle^2$$

Idea: Using the exterior product to improve the algebraic algorithm.

- $\mathbf{Q}$ is not explicitly constructed
- $\text{null}(\mathbf{Q}) = \text{null}(\mathbf{Q}^T\mathbf{Q})$
- $\langle \mathbf{q}_k, \mathbf{q}_h \rangle = \langle \text{vec}(\mathcal{C}_2(\mathbf{Y})), \text{vec}(\mathcal{C}_2(\mathbf{Z})) \rangle$ with $\mathbf{Y}, \mathbf{Z}$ slices or a linear combination of slices of the input tensor $\mathcal{X}$

  - Pre-compute $\mathbf{A}_{kh} = \mathbf{X}_k^T \mathbf{X}_h$ for $h \leq k$
  - define 6 functions f_i that depends on the $\mathbf{A}_{kh}$
  - use f_i and $\mathbf{A}_{kh}$ to compute the entries of $\mathbf{Q}^T\mathbf{Q}$

Overview

Preliminary numerical results (work in progress)

Given an input tensor $\mathcal{X}$ of dimensions $(I \times J \times K)$ such that its rank is $I = (J-1)(K-1)$

dimensions	**rank**	**CPU time**
$20 \times 5 \times 6$	$R = 20$	5.16sec
$24 \times 5 \times 7$	$R = 24$	24.42sec
$25 \times 6 \times 6$	$R = 25$	21.82sec
$28 \times 5 \times 8$	$R = 28$	283.22sec
$30 \times 6 \times 7$	$R = 30$	$\approx$ 7min

The link between rank and dimensions is meant to satisfy [Domanov and De Lathauwer 2016] theorem.

Conclusive remarks

- conditions that guarantee generic uniqueness;
- from the theorem structure to a CPD algorithm;
- bottleneck due to compound matrices;
- optimization based on the exterior product.

Thank you for the attention!
Questions?

References I

Comon, P. and C. Jutten (2009). *Handbook of blind source separation: Independent component analysis and applications*. Academic press.

Domanov, I. and L. De Lathauwer (July 2013). “On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix”. In: *SIAM J. Matrix Anal. Appl.* 34.3, pp. 855–875.

— (June 2016). “Generic Uniqueness of a Structured Matrix Factorization and Applications in Blind Source Separation”. In: *IEEE J. Sel. Topics Signal Process.* 10.4, pp. 701–711.