Principal Component, Independent Component and Parallel Factor Analysis

Lieven De Lathauwer

CNRS ETIS - Cergy-Pontoise
France
delathau@ensea.fr

Overview

- Rank
- Singular Value Decomposition
- Parallel Factor analysis
- Independent Component Analysis

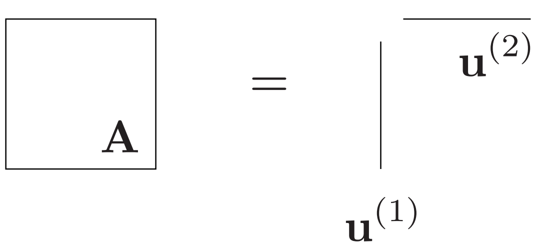
Rank-1 tensor

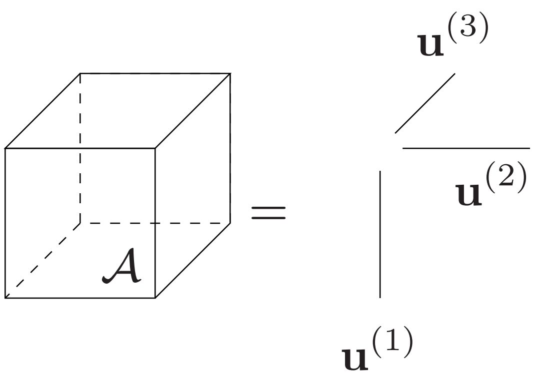
- Rank-1 matrix: outer product of 2 vectors $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$:

$$
\begin{aligned}
a_{i_1 i_2} &= u^{(1)}_{i_1} u^{(2)}_{i_2} \\
\mathbf{A} &= \mathbf{u}^{(1)} \cdot \mathbf{u}^{(2)^T} \equiv \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)}
\end{aligned}
$$

- Rank-1 tensor: outer product of N vectors $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$, ..., $\mathbf{u}^{(N)}$:

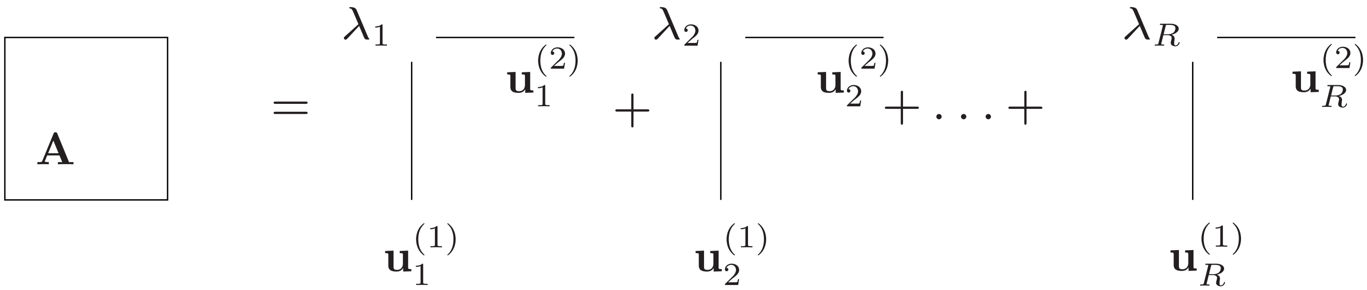
$$
\begin{aligned}
a_{i_1 i_2 \ldots i_N} &= u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N} \\
\mathcal{A} &= \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \ldots \circ \mathbf{u}^{(N)}
\end{aligned}
$$



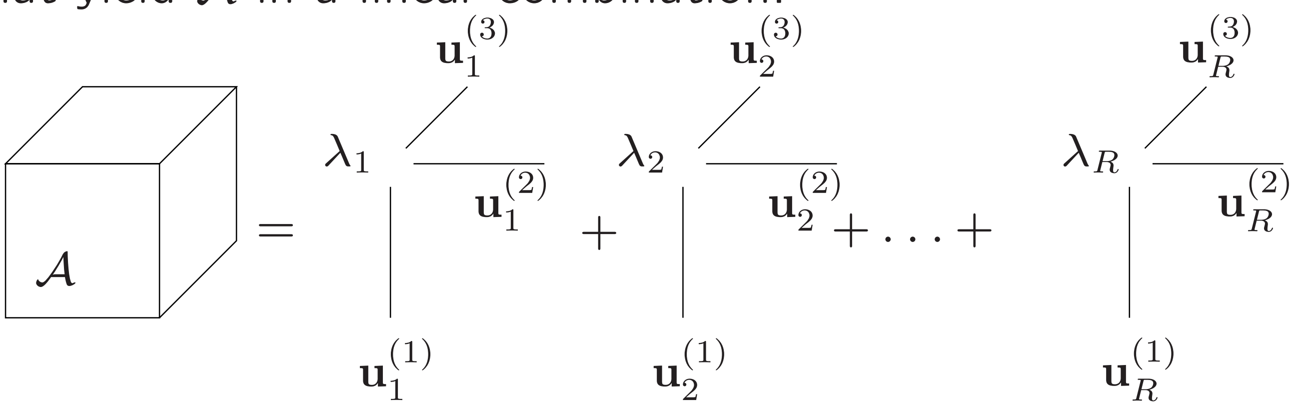


Rank of a tensor

- The rank R of a matrix $\mathbf{A}$ is minimal number of rank-1 matrices that yield $\mathbf{A}$ in a linear combination.



- The rank R of an Nth-order tensor $\mathcal{A}$ is the minimal number of rank-1 tensors that yield $\mathcal{A}$ in a linear combination.

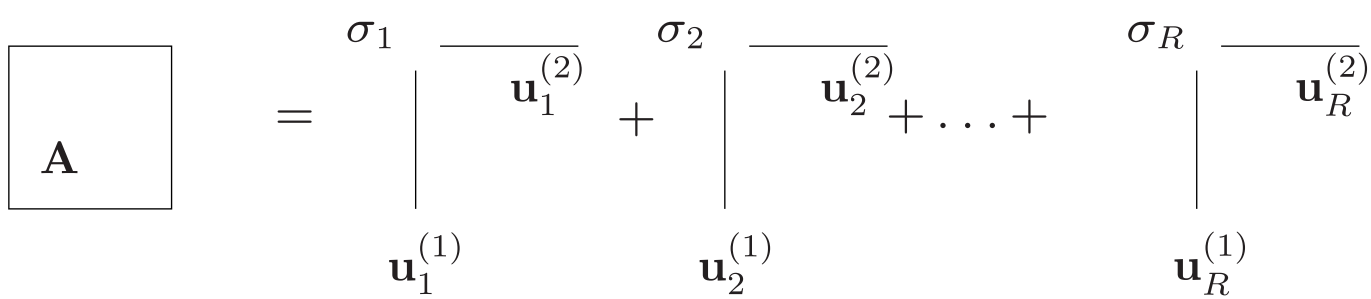


Matrix Singular Value Decomposition

- Definition:

$$\mathbf{A} = \mathbf{U}^{(1)} \cdot \mathbf{\Sigma} \cdot \mathbf{U}^{(2)^T}$$

$\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ orthogonal, $\mathbf{\Sigma}$ diagonal

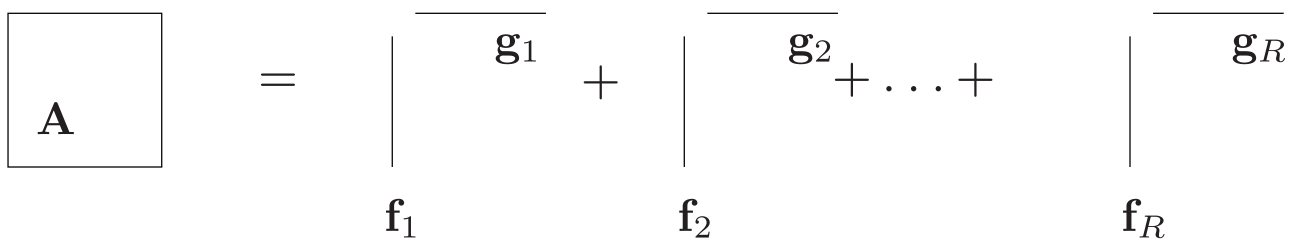


- Best rank-r approximation $\prec$ truncation SVD

SVD and Factor Analysis

- Decompose a data matrix in rank-1 terms
  E.g. independent component analysis, telecommunications, biomedical applications, chemometrics, data analysis, ...

$$\mathbf{A} = \mathbf{F} \cdot \mathbf{G}^T$$



- Decomposition in rank-1 terms is not unique

$$\begin{aligned} \mathbf{A} &= (\mathbf{F}\mathbf{M}) \cdot (\mathbf{M}^{-1}\mathbf{G}^T) \\ &= \tilde{\mathbf{F}} \cdot \tilde{\mathbf{G}}^T \end{aligned}$$

- Exploitation of prior knowledge
- SVD made unique by adding orthogonality constraints

$$\mathbf{A} = \mathbf{U}^{(1)} \cdot \mathbf{\Sigma} \cdot \mathbf{U}^{(2)^T}$$

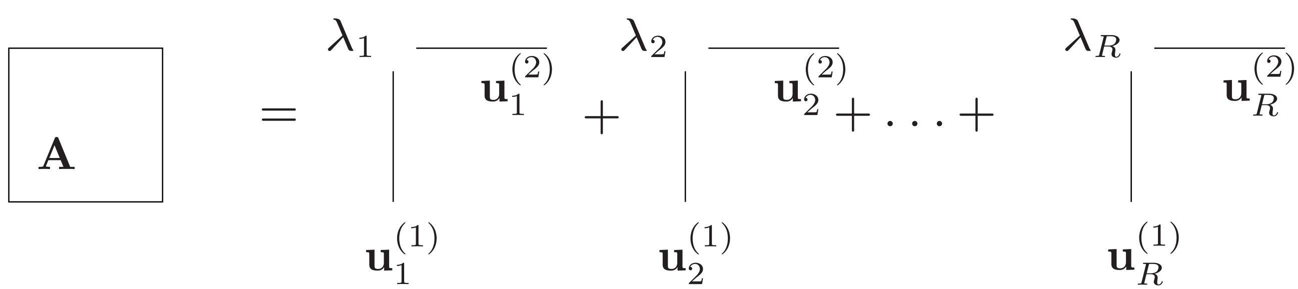
  $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ orthogonal, $\mathbf{\Sigma}$ diagonal

- Problems: interpretability
  reification

Example: emission-excitation fluorescence in chemometrics

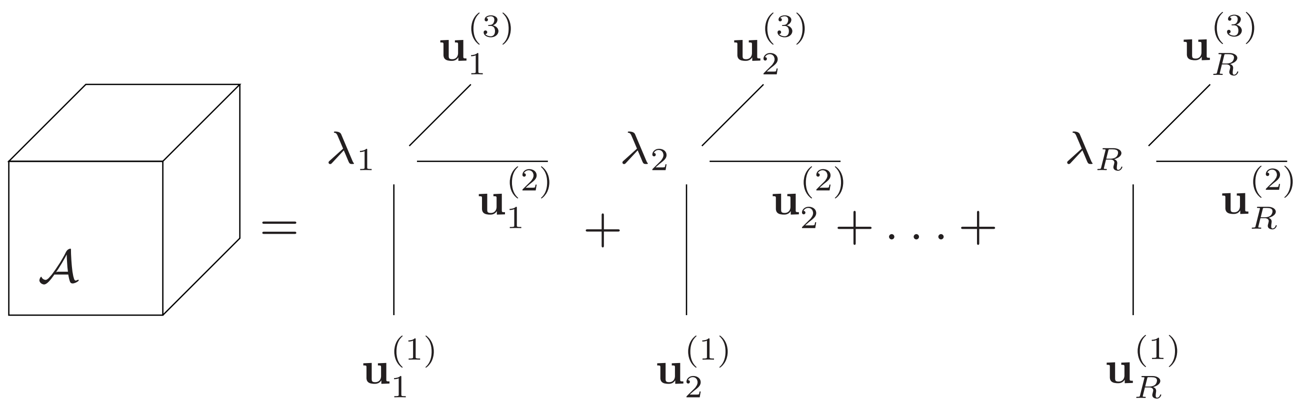
Matrix approach

row vector $\sim$ excitation spectrum
column vector $\sim$ emission spectrum
coefficients $\sim$ concentrations



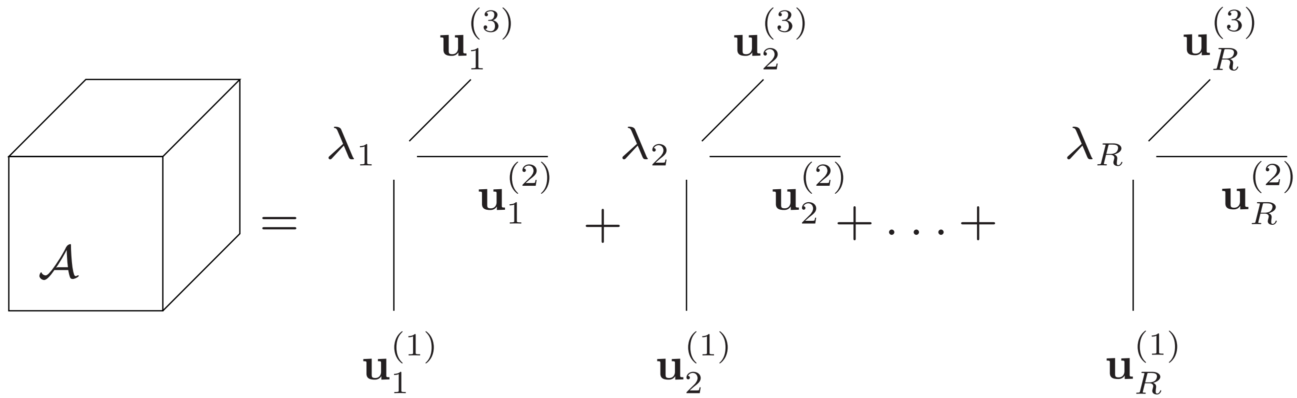
Tensor approach

row vector $\sim$ excitation spectrum
column vector $\sim$ emission spectrum
coefficients $\sim$ concentrations



CANDECOMP/PARAFAC

Canonical Decomposition / Parallel Factor Decomposition of a tensor $\mathcal{A}$ is its decomposition in a minimal sum of rank-1 tensors



Matrix formulation:

$$\mathbf{A}_{I_1 I_2 \times I_3} = (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \cdot \mathbf{\Lambda} \cdot \mathbf{U}^{(3)^T}$$

[*Hitchcock '27*], [*Harshman '70*], [*Carroll and Chang '70*]

Orthogonality constraints:

[*Comon '94*], [*Kolda '01*], [*Moravitz and Van Loan '06*]

Uniqueness (1)

The k-rank of a matrix **A** is the maximal number such that any set of k columns of **A** is linearly independent.

Deterministic bound: For $\mathcal{A} \in \mathbb{C}^{I \times J \times K}$ uniqueness if

$$k(\mathbf{U}^{(1)}) + k(\mathbf{U}^{(2)}) + k(\mathbf{U}^{(3)}) \geqslant 2R + 2$$

[*Kruskal '77*], [*Sidiropoulos '00*], [*Stegeman and Sidiropoulos '06*]

Generic bound:

$$\min(I, R) + \min(J, R) + \min(K, R) \geqslant 2R + 2$$

If $K \geqslant R$:

$$R \leqslant \min(I, R) + \min(J, R) - 2 \leqslant I + J - 2$$

Uniqueness (2)

Theorem 1. For $\mathcal{A} \in \mathbb{C}^{I \times J \times K}$ uniqueness if

$$\min(I, R) + \min(J, R) + \min(K, R) \geqslant 2(R + 1)$$

[*Kruskal '77*], [*Sidiropoulos '00*], [*Stegeman and Sidiropoulos '06*]

Theorem 2. For $\mathcal{A} \in \mathbb{C}^{I \times J \times K}$, with $K \geqslant R$, uniqueness if

$$2R(R - 1) \leqslant I(I - 1)J(J - 1)$$

[*De Lathauwer '06*]

(Compare to $R \leqslant \min(I, R) + \min(J, R) - 2 \leqslant I + J - 2$)

Computation

No greedy algorithm

[*Kofidis and Regalia '02*]

Classical approach: direct minimization of

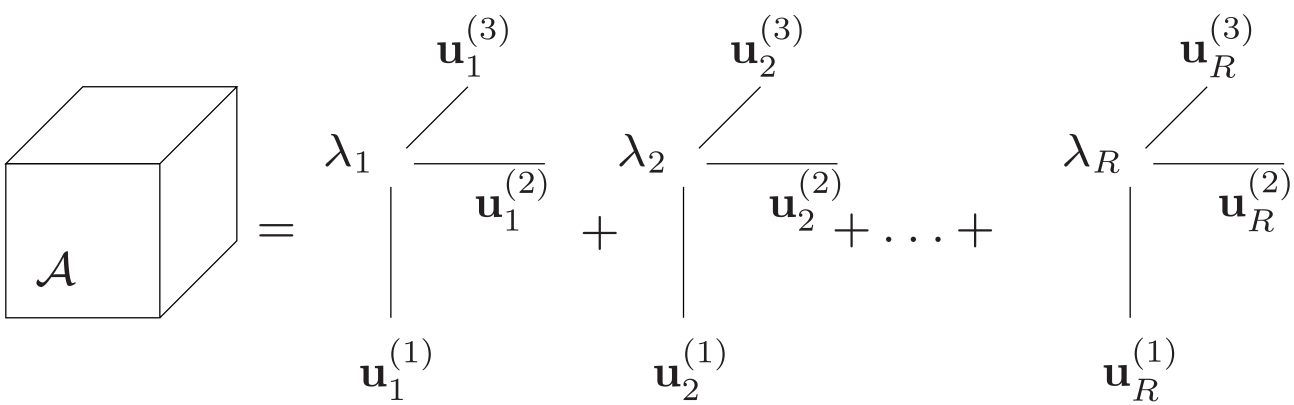
$$f(\mathbf{U},\mathbf{V},\mathbf{W}) = \|\mathcal{A} - \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \mathbf{u}_r^{(3)}\|^2$$

e.g. Alternating Least Squares

Simultaneous matrix diagonalization:

$$\begin{aligned}
\mathbf{A}_1 &= \mathbf{U}^{(1)} \cdot \mathbf{D}_1 \cdot {\mathbf{U}^{(2)}}^T \\
\mathbf{A}_2 &= \mathbf{U}^{(1)} \cdot \mathbf{D}_2 \cdot {\mathbf{U}^{(2)}}^T \\
&\vdots \\
\mathbf{A}_K &= \mathbf{U}^{(1)} \cdot \mathbf{D}_K \cdot {\mathbf{U}^{(2)}}^T
\end{aligned}$$

note: $\mathbf{A}_1 \cdot \mathbf{A}_2^{-1} = \mathbf{U}^{(1)} \cdot (\mathbf{D}_1 \cdot \mathbf{D}_2^{-1}) \cdot {\mathbf{U}^{(1)}}^{-1}$

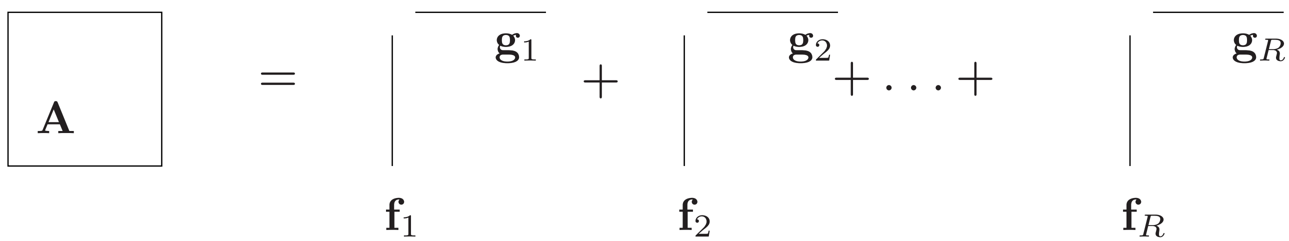


[*Leurgans et al '93*], [*Sanchez and Kowalski '91*], [*De Lathauwer '04*], [*De Lathauwer '06*]

SVD and Factor Analysis

- Decompose a data matrix in rank-1 terms
  E.g. independent component analysis, telecommunications, biomedical applications, chemometrics, data analysis, . . .

$$\mathbf{A} = \mathbf{F} \cdot \mathbf{G}^T$$



- Decomposition in rank-1 terms is not unique
- SVD made unique by adding orthogonality constraints

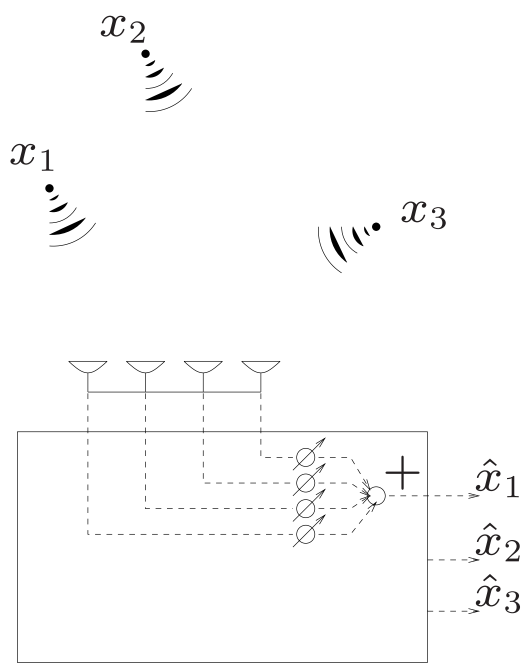
$$\mathbf{A} = \mathbf{U}^{(1)} \cdot \mathbf{\Sigma} \cdot \mathbf{U}^{(2)^T}$$

$\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ orthogonal, $\mathbf{\Sigma}$ diagonal

Independent Component Analysis (ICA)

Model:

$$Y = \mathbf{M}X + N$$



Model:

$$\begin{array}{ccc} Y & = & \mathbf{M}X + N \\ (P \times 1) & & (P \times R)(R \times 1)\ (P \times 1) \end{array}$$

Assumptions:

- columns of $\mathbf{M}$ are linearly independent
- components of X are statistically independent

Goal:

Identification of $\mathbf{M}$ and/or reconstruction of X while observing only Y

Independent Component Analysis (ICA)

Disciplines:

statistics, neural networks, information theory, *linear and multilinear algebra*, . . .

Indeterminacies:

ordering and scaling of the columns ($Y = \mathbf{M}X$)

Uncorrelated vs independent:

X, Y are uncorrelated iff $E\{XY\} = 0$
X, Y are independent iff $p_{XY}(x, y) = p_X(x)p_Y(y)$

statistical independence implies:

- the variables are uncorrelated
- additional conditions on the HOS

Algebraic tools:

Condition	Identification	Tool
X_i uncorr.	column space $\mathbf{M}$	matrix EVD/SVD
X_i indep.	$\mathbf{M}$	tensor EVD/SVD

Web site:

`http://www.tsi.enst.fr/icacentral/index.html`

mailing list, data sets, software

Applications

- Speech and audio
- Image processing

  feature extraction, image reconstruction, video
- Telecommunications

  OFDM, CDMA, . . .
- Biomedical applications

  functional Magnetic Resonance Imaging, electromyogram, electro-encephalogram, (fetal) electrocardiogram, mammography, pulse oximetry, (fetal) magnetocardiogram, . . .
- Other applications

  text classification, vibratory signals generated by termites (!), electron energy loss spectra, astrophysics, . . .

HOS definitions

Moments and cumulants of a random variable:

Moments	*Cumulants*
$m_1^X = E\{X\}$ "mean" (m_X)	$c_1^X = E\{X\}$ "mean"
$m_2^X = E\{X^2\}$ (R_X)	$c_2^X = E\{(X - m_X)^2\}$ "variance" (σ_X^2)
$m_3^X = E\{X^3\}$	$c_3^X = E\{(X - m_X)^3\}$
$m_4^X = E\{X^4\}$	$c_4^X = E\{(X - m_X)^4\} - 3\sigma_X^4$

Moments and cumulants of a set of random variables:

Moments:

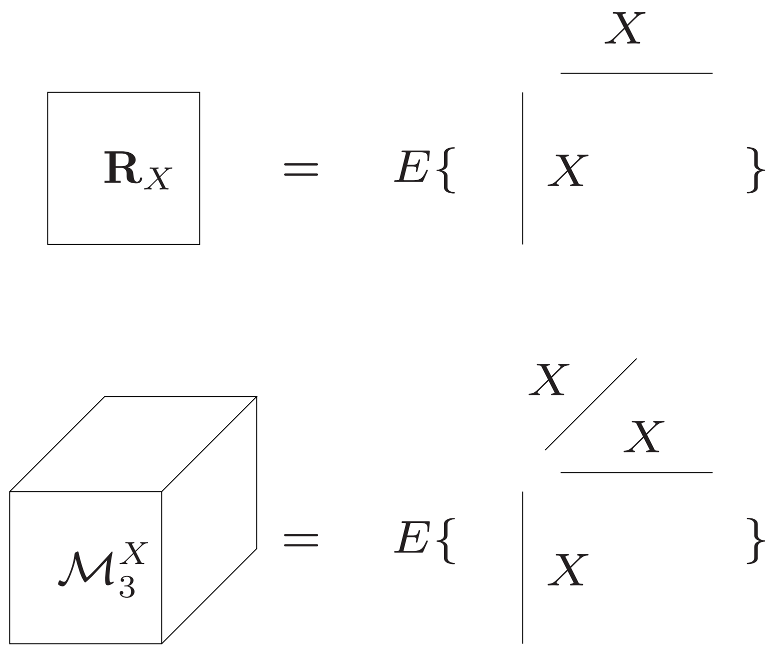
$$(\mathcal{M}_{\mathbf{x}}^{(N)})_{i_1 i_2 \ldots i_N} = \mathtt{Mom}(x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \stackrel{\text{def}}{=} \mathtt{E}\{x_{i_1} x_{i_2} \ldots x_{i_N}\}$$

Cumulants:

$$\begin{aligned}
(\mathbf{c}_{\mathbf{x}})_i = \mathtt{Cum}(x_i) &\stackrel{\text{def}}{=} \mathtt{E}\{x_i\} \\
(\mathbf{C}_{\mathbf{x}})_{i_1 i_2} = \mathtt{Cum}(x_{i_1}, x_{i_2}) &\stackrel{\text{def}}{=} \mathtt{E}\{x_{i_1} x_{i_2}\} \\
(\mathcal{C}_{\mathbf{x}}^{(3)})_{i_1 i_2 i_3} = \mathtt{Cum}(x_{i_1}, x_{i_2}, x_{i_3}) &\stackrel{\text{def}}{=} \mathtt{E}\{x_{i_1} x_{i_2} x_{i_3}\} \\
(\mathcal{C}_{\mathbf{x}}^{(4)})_{i_1 i_2 i_3 i_4} = \mathtt{Cum}(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}) &\stackrel{\text{def}}{=} \mathtt{E}\{x_{i_1} x_{i_2} x_{i_3} x_{i_4}\} - \mathtt{E}\{x_{i_1} x_{i_2}\}\mathtt{E}\{x_{i_3} x_{i_4}\} \\
&\quad - \mathtt{E}\{x_{i_1} x_{i_3}\}\mathtt{E}\{x_{i_2} x_{i_4}\} - \mathtt{E}\{x_{i_1} x_{i_4}\}\mathtt{E}\{x_{i_2} x_{i_3}\}
\end{aligned}$$

Order $\geqslant 2$: $x_i \leftarrow x_i - \mathtt{E}\{x_i\}$

Multivariate case: e.g. moments:



$$
\Longrightarrow \left\{ \begin{array}{llcc}
1: & m_X & \stackrel{\text{def}}{=} & E\{X\} \\
 & & \longrightarrow & \text{vector} \\
 & & & \\
2: & \mathbf{R}_X & \stackrel{\text{def}}{=} & E\{X X^T\} \\
 & & \longrightarrow & \text{matrix} \\
 & & & \\
3: & \mathcal{M}_3^X & \stackrel{\text{def}}{=} & E\{X \circ X \circ X\} \\
 & & \longrightarrow & \text{3rd order tensor} \\
 & & & \\
4: & \mathcal{M}_4^X & \stackrel{\text{def}}{=} & E\{X \circ X \circ X \circ X\} \\
 & & \longrightarrow & \text{4th order tensor}
\end{array} \right.
$$

ICA: basic equations

Model:

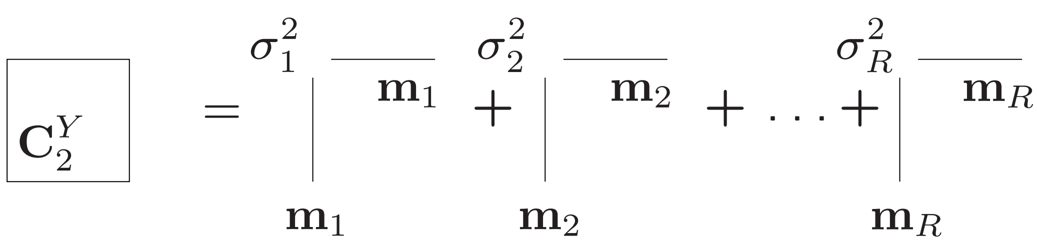
$$Y = \mathbf{M}X$$

Second order:

$$\begin{aligned} \mathbf{C}_2^Y &= E\{YY^T\} \\ &= \mathbf{M} \cdot \mathbf{C}_2^X \cdot \mathbf{M}^{\mathrm{T}} \\ &= \mathbf{C}_2^X \bullet_1 \mathbf{M} \bullet_2 \mathbf{M} \end{aligned}$$

uncorrelated sources: $\mathbf{C}_2^X$ is diagonal

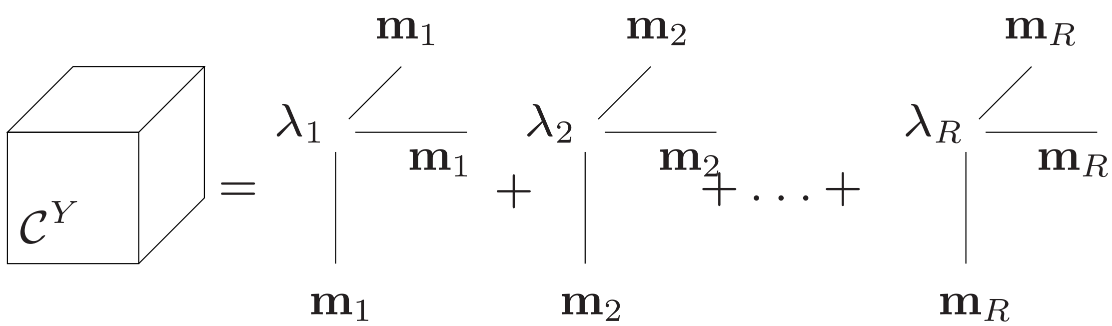
"diagonalization by congruence"



Higher order:

$$\mathcal{C}_4^Y = \mathcal{C}_4^X \bullet_1 \mathbf{M} \bullet_2 \mathbf{M} \bullet_3 \mathbf{M} \bullet_4 \mathbf{M}$$

independent sources: $\mathcal{C}_4^X$ is diagonal
"CANDECOMP / PARAFAC"



Prewhitening-based computation

Model:

$$Y = \mathbf{M}X$$

Second order:

$$\begin{aligned} \mathbf{C}_2^Y &= E\{YY^T\} \\ &= \mathbf{M} \cdot \mathbf{C}_2^X \cdot \mathbf{M}^T \\ &\Rightarrow \mathbf{M} \cdot \mathbf{I} \cdot \mathbf{M}^T \\ &= \mathbf{M} \cdot \mathbf{M}^T \\ &= (\mathbf{M} \cdot \mathbf{Q}) \cdot (\mathbf{M} \cdot \mathbf{Q})^T \end{aligned}$$

"square root": EVD, Cholesky, ...

Remark: PCA:

SVD of $\mathbf{M}$: $\mathbf{M} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T$

$$\Rightarrow \mathbf{C}_2^Y = (\mathbf{US}) \cdot (\mathbf{US})^T = \mathbf{U} \cdot \mathbf{S}^2 \cdot \mathbf{U}^T$$

Prewhitening-based computation (2)

Matrix factorization:

$$\mathbf{M} = \mathbf{T} \cdot \mathbf{Q}$$

Whitening:

$$\begin{aligned} Y &= \mathbf{M}X \\ Z &= \mathbf{T}^{-1}Y = \mathbf{Q}X \end{aligned}$$

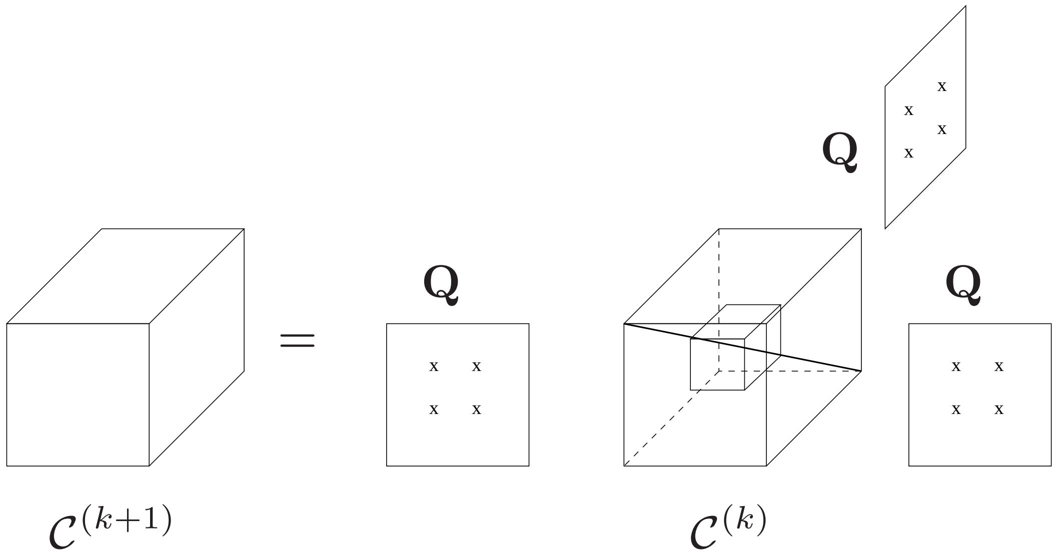
Higher order: ICA:

$$\begin{aligned} \mathcal{C}_4^Y &= \mathcal{C}_4^X \bullet_1 \mathbf{M} \bullet_2 \mathbf{M} \bullet_3 \mathbf{M} \bullet_4 \mathbf{M} \\ \Rightarrow \mathcal{C}_4^Z &= \mathcal{C}_4^X \bullet_1 \mathbf{Q} \bullet_2 \mathbf{Q} \bullet_3 \mathbf{Q} \bullet_4 \mathbf{Q} \end{aligned}$$

"multilinear symmetric EVD"
"CANDECOMP/PARAFAC with orthogonality and symmetry constraints"

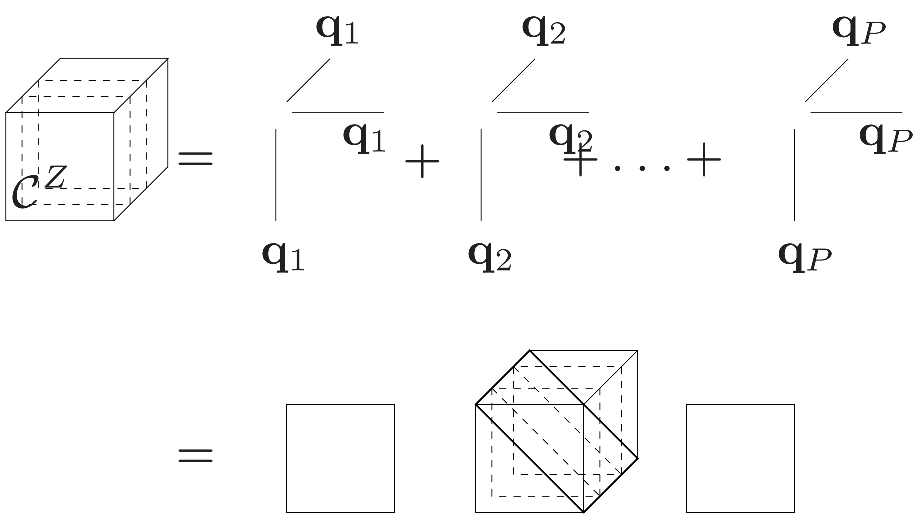
Source cumulant is theoretically diagonal
An arbitrary symmetric tensor cannot be diagonalized ⇒ different strategies

Algorithm 1: maximal diagonality



[*Comon '94*], [*De Lathauwer '01*]

Algorithm 2: simultaneous EVD



[*Cardoso '94 (JADE)*]

Conclusion

- Uniqueness and interpretability
- ICA uniqueness $<$ PARAFAC uniqueness
- Applications
- Numerical algorithms