An Introduction to Multilinear Algebra Based Independent Component Analysis

Lieven De Lathauwer

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

Overview

- Problem definition
- Multilinear algebra: notation
- Basics of Higher-Order Statistics
- Prewhitening-based computation: general scheme
- Specific prewhitening-based multilinear algorithms
- A higher-order-only scheme
- A variant for coloured sources
- Dimensionality reduction
- Conclusions

Independent Component Analysis (ICA)

Model:

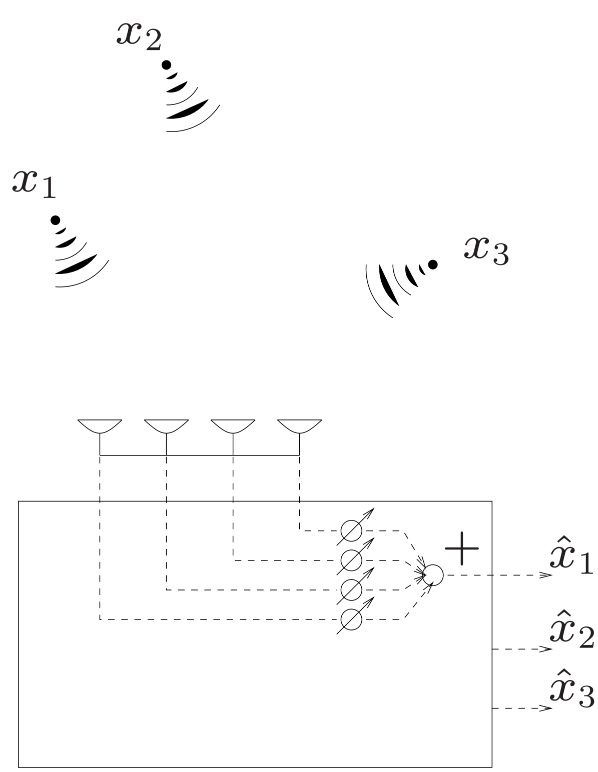
$$\begin{array}{ccc} Y & = & \mathbf{M}X + N \\ (P \times 1) & & (P \times R)(R \times 1)\ (P \times 1) \end{array}$$

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

Condition	Identification	Tool
X_i uncorr.	column space **M**	matrix EVD/SVD
X_i indep.	**M**	tensor EVD/SVD

Applications

- Speech and audio
- Image processing

  feature extraction, image reconstruction, video
- Telecommunications

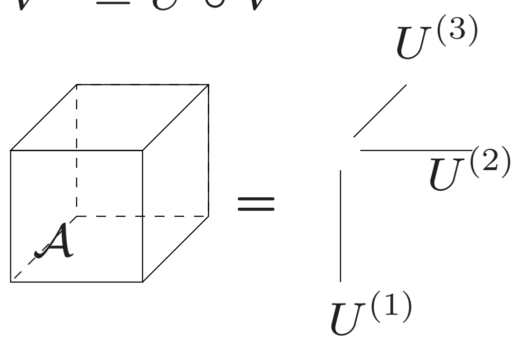
  OFDM, CDMA, . . .
- Biomedical applications

  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, . . .

Multilinear algebra: notation

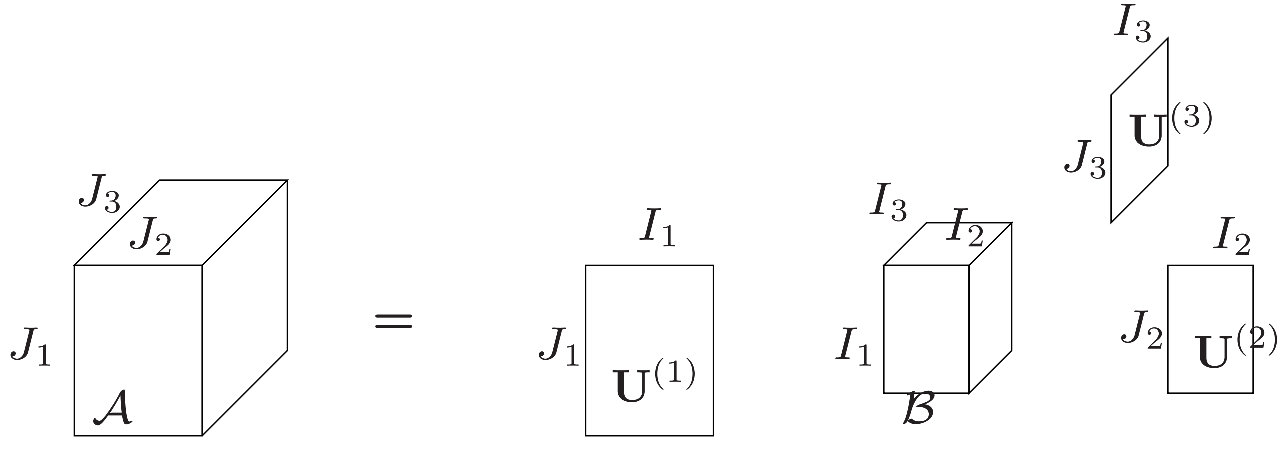
- Outer product: $\mathcal{C} = \mathcal{A} \circ \mathcal{B} \Leftrightarrow c_{ijklm} = a_{ij} b_{klm}$
  E.g. $\mathbf{C} = UV^T = U \circ V$



$$\mathcal{A} = U^{(1)} \circ U^{(2)} \circ U^{(3)}$$

- Matrix multiplication:
  $\mathcal{A} = \mathcal{B} \times_2 \mathbf{U} \Leftrightarrow a_{ijk} = \sum_p b_{ipk} u_{jp}$
  E.g. $\mathbf{A} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T = \mathbf{S} \times_1 \mathbf{U} \times_2 \mathbf{V}$

[*Tucker '64*]



$$\mathcal{A} = \mathcal{B} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}$$

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"
$M_2^X = E\{X^2\}$ "correlation" (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:

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

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

Multivariate case: e.g. moments:

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

HOS properties

Multilinearity: for $\tilde{X} = \mathbf{A} \cdot X$

$$\begin{aligned} \mathcal{M}_N^{\tilde{X}} &= \mathcal{M}_N^{X} \times_1 \mathbf{A} \times_2 \mathbf{A} \ldots \times_N \mathbf{A} \\ \mathcal{C}_N^{\tilde{X}} &= \mathcal{C}_N^{X} \times_1 \mathbf{A} \times_2 \mathbf{A} \ldots \times_N \mathbf{A} \end{aligned}$$

E.g. $E\{\tilde{X}\tilde{X}^T\} = \mathbf{A} \cdot E\{XX^T\} \cdot \mathbf{A}^T$

Symmetry:

$$\begin{aligned} \mathtt{Mom}(X_1, X_2, \ldots, X_I) &= \mathtt{Mom}(X_{\mathrm{P}(1)}, X_{\mathrm{P}(2)}, \ldots, X_{\mathrm{P}(I)}) \\ \mathtt{Cum}(X_1, X_2, \ldots, X_I) &= \mathtt{Cum}(X_{\mathrm{P}(1)}, X_{\mathrm{P}(2)}, \ldots, X_{\mathrm{P}(I)}) \end{aligned}$$

$\longrightarrow$ supersymmetric higher-order tensors

Even distribution: odd moments and cumulants = 0

Partitioning of independent variables:

$$\mathtt{Cum}(X_1, X_2, \ldots, X_I) = 0$$

Stochastic vector of which components are mutually independent: cumulant tensor = diagonal

Moments: e.g. $\mathtt{Mom}(x, x, y, y) = \mathtt{E}\{x^2\} \cdot \mathtt{E}\{y^2\} \neq 0$

Non-Gaussianity:

Higher-order cumulants of a Gaussian variable are 0

Higher-order cumulants are blind to additive Gaussian noise:
for $\hat{x} = x + n$: $\mathcal{C}_{\hat{x}}^{(N)} = \mathcal{C}_{x}^{(N)} + \mathcal{C}_{n}^{(N)} = \mathcal{C}_{x}^{(N)}$

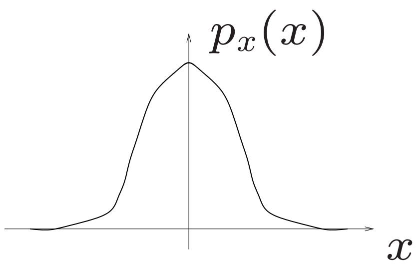
Estimation:

Higher order: - harder to estimate
- more entries

$\Rightarrow$ (third- and) fourth-order cumulants

HOS example

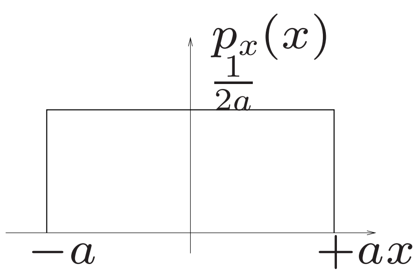
Gaussian distribution



$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

n	$m_x^{(n)}$	$c_x^{(n)}$
1	0	0
2	σ^2	σ^2
3	0	0
4	$3\ \sigma^4$	0

Uniform distribution



$$p_x(x) = \frac{1}{2a} \qquad (x \in [-a, +a])$$

n	$m_x^{(n)}$	$c_x^{(n)}$
1	0	0
2	$a^2/3$	$a^2/3$
3	0	0
4	$3a^4/5$	$-2a^4/15$

ICA: basic equations

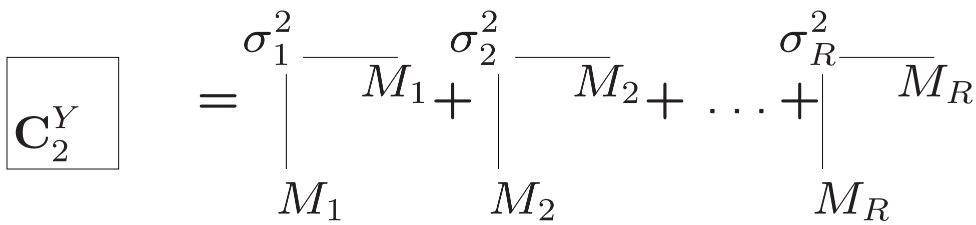
Model:

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

Second order:

$$\begin{aligned} \mathbf{C}_2^Y &= E\{YY^T\} \\ &= \mathbf{M} \cdot \mathbf{C}_2^X \cdot \mathbf{M}^\mathrm{T} \\ &= \mathbf{C}_2^X \times_1 \mathbf{M} \times_2 \mathbf{M} \end{aligned}$$

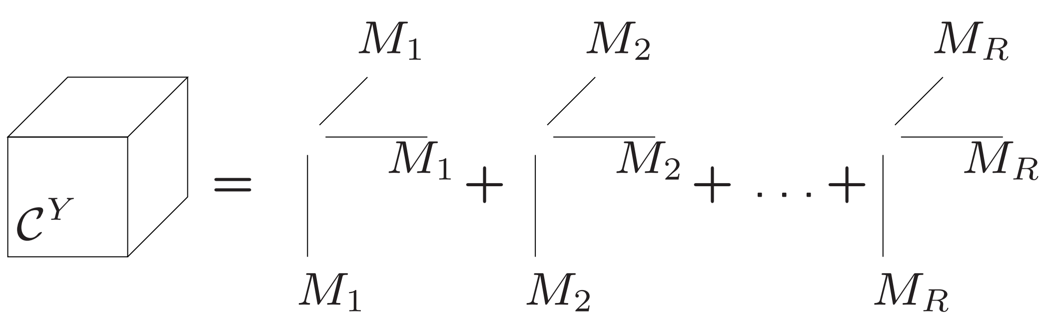
uncorrelated sources: $\mathbf{C}_2^X$ is diagonal
"diagonalization by congruence"



Higher order:

$$\mathcal{C}_4^Y = \mathcal{C}_4^X \times_1 \mathbf{M} \times_2 \mathbf{M} \times_3 \mathbf{M} \times_4 \mathbf{M}$$

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



Prewhitening-based computation

Model:

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

First order:

$$0 = 0$$

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

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

Prewhitening-based computation (2)

Matrix factorization:

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

Second order:

$$\mathbf{C}_2^Y = \mathbf{C}_2^X \times_1 \mathbf{M} \times_2 \mathbf{M} = \mathbf{T} \cdot \mathbf{T}^T$$

Whitened r.v. $Z = \mathbf{Q} X = \mathbf{T}^{-1} Y$

Higher order: ICA:

$$\mathbf{Q} \prec \mathcal{C}_3^Y \text{ or } \mathcal{C}_4^Y$$

$$\begin{aligned} \mathcal{C}_4^Y &= \mathcal{C}_4^X \times_1 \mathbf{M} \times_2 \mathbf{M} \times_3 \mathbf{M} \times_4 \mathbf{M} \\ \mathcal{C}_4^Z &= \mathcal{C}_4^Y \times_1 \mathbf{T}^{-1} \times_2 \mathbf{T}^{-1} \times_3 \mathbf{T}^{-1} \times_4 \mathbf{T}^{-1} \\ \Rightarrow \mathcal{C}_4^Z &= \mathcal{C}_4^X \times_1 \mathbf{Q} \times_2 \mathbf{Q} \times_3 \mathbf{Q} \times_4 \mathbf{Q} \end{aligned}$$

"multilinear supersymmetric EVD"
"supersymmetric completely orthogonal rank decomposition"
"CANDECOMP/PARAFAC with orthogonality and symmetry constraints"

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

PCA versus ICA

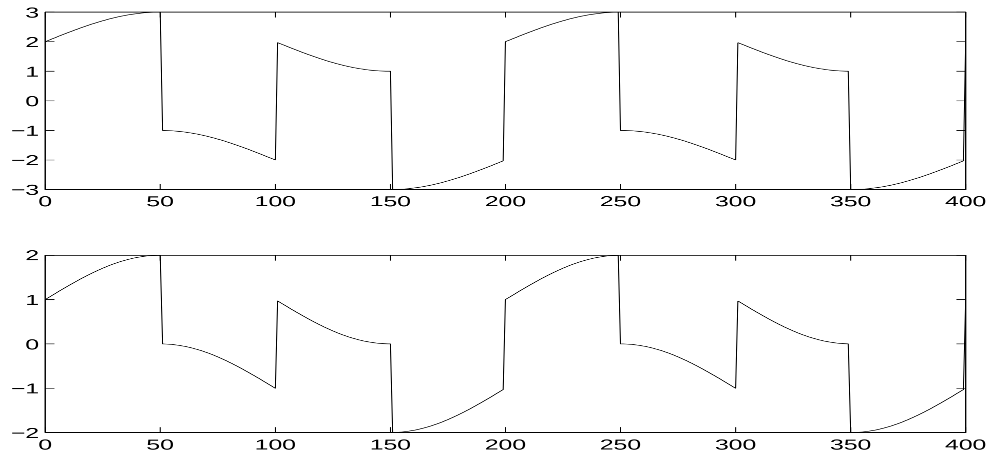
ICA = higher-order fine-tuning of PCA:

PCA	ICA
2nd-order	higher-order
matrix EVD	tensor EVD
uncorrelated sources	independent sources
column space **M**	**M** itself
always possible	depends on context

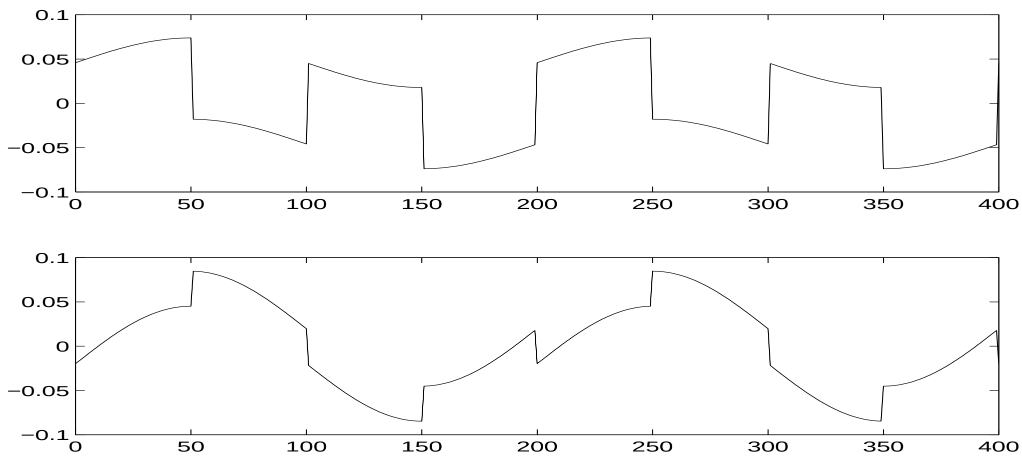
Computational cost:

cumulant estimation and diagonalization

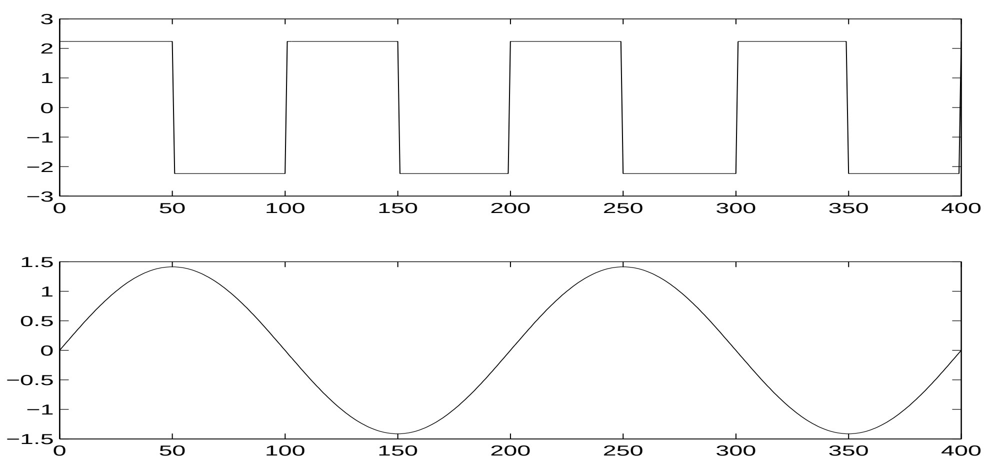
Illustration



Observations

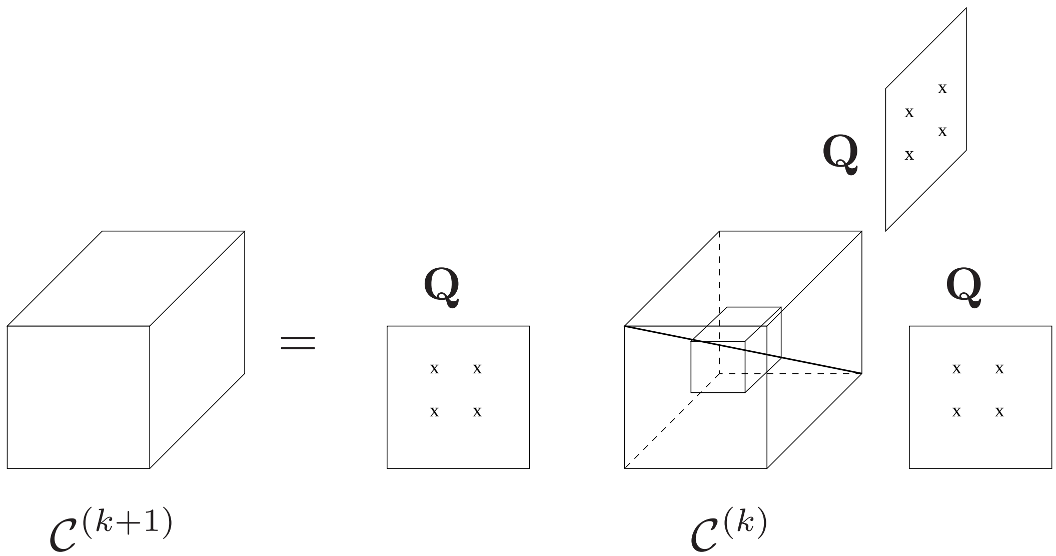


Sources estimated with PCA



Sources estimated with ICA

Algorithm 1: maximal diagonality

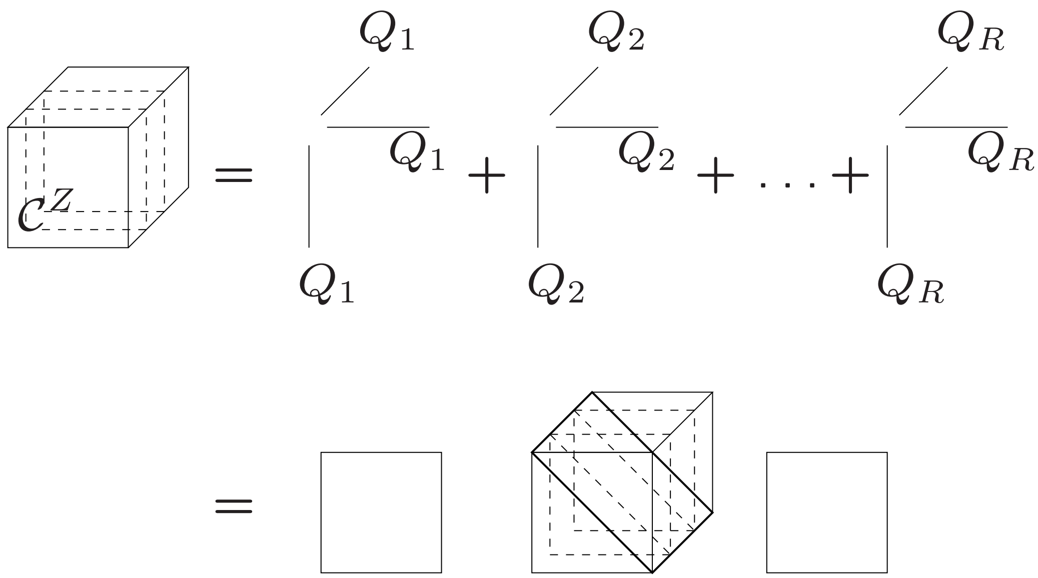


- Maximize energy on the diagonal by Jacobi-iteration
- Determination of optimal rotation angle:

order 3	real	roots polynomial degree 2
order 3	complex	roots polynomial degree 3
order 4	real	roots polynomial degree 4
order 4	complex	-

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

Algorithm 2: simultaneous EVD



- Maximize energy on the diagonals by Jacobi-iteration
- Determination of optimal rotation angle:

real	roots polynomial degree 2
complex	roots polynomial degree 3

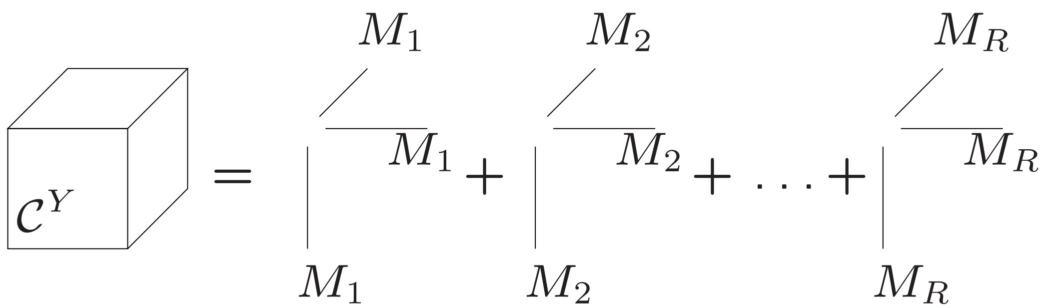
[*Cardoso '94 (JADE)*]

Higher-order-only approach

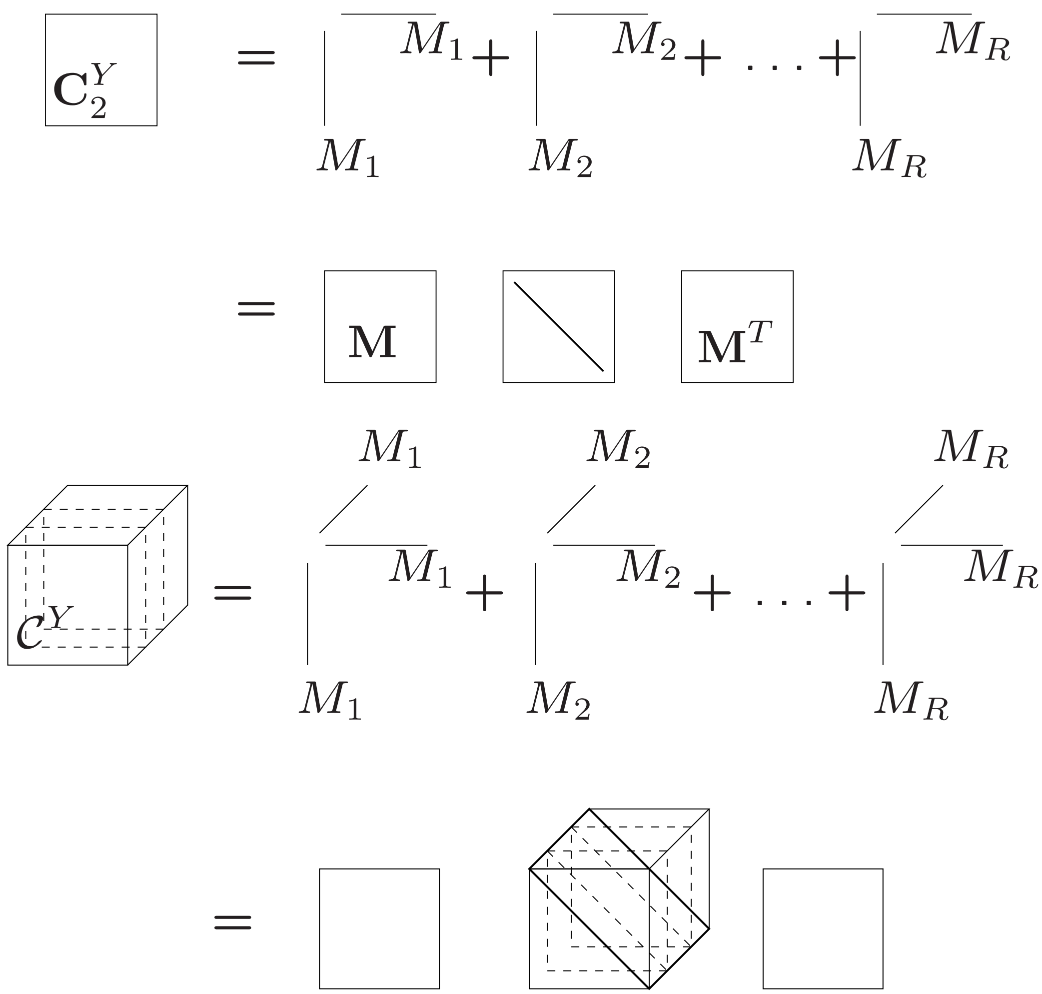
$$Y = \mathbf{M}X$$

$$\mathcal{C}_4^Y = \mathcal{C}_4^X \times_1 \mathbf{M} \times_2 \mathbf{M} \times_3 \mathbf{M} \times_4 \mathbf{M}$$

$\mathcal{C}_4^X$ is diagonal $\longrightarrow$ CANDECOMP / PARAFAC



Soft whitening



$\longrightarrow$ combine 2nd and HO information in a single tensor

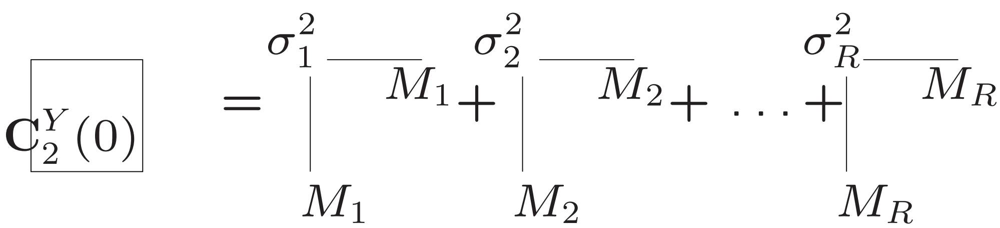
[*Yeredor '02*]

A variant for coloured sources

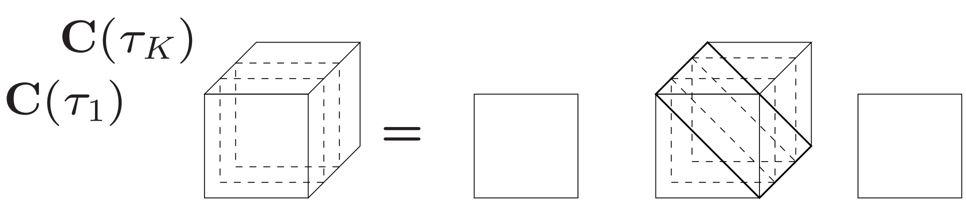
Condition: sources mutually uncorrelated, but individually correlated in time

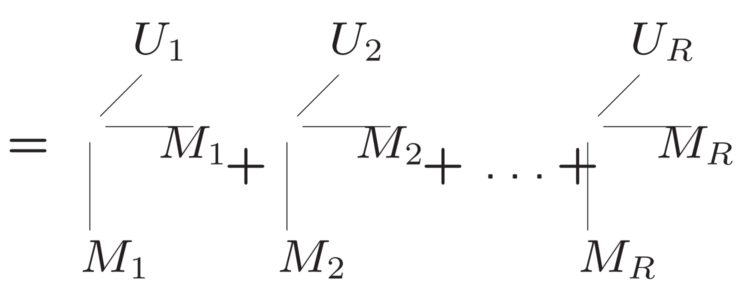
Basic equations:

$$\begin{aligned} \mathbf{C}_2^Y(0) &= E\{Y(t)Y(t)^T\} \\ &= \mathbf{M}\cdot\mathbf{C}_2^X(0)\cdot\mathbf{M}^T \end{aligned}$$



$$\begin{aligned} \mathbf{C}_2^Y(\tau) &= E\{Y(t)Y(t+\tau)^T\} \\ &= \mathbf{M}\cdot\mathbf{C}_2^X(\tau)\cdot\mathbf{M}^T \end{aligned}$$





[*Belouchrani et al. '97 (SOBI)*]

Large mixtures: more sensors than sources

Applications:

EEG, MEG, NMR, hyper-spectral image processing, data analysis, . . .

Prewhitening-based algorithms:

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

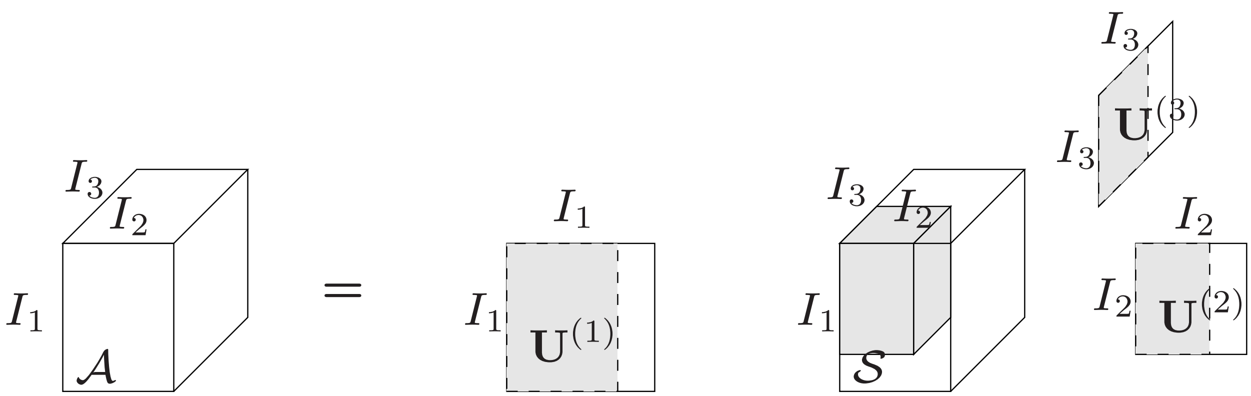
$$\begin{array}{ccc} \mathbf{M} & = & \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T \\ (P \times R) & & (P \times R)(R \times R)(R \times R) \end{array}$$

$$\begin{array}{ccc} Z & = & \mathbf{S}^{-1} \cdot \mathbf{U}^T Y \\ Z & = & \mathbf{V}^T X \\ (R \times 1) & & (R \times R)(R \times 1) \end{array}$$

Large mixtures: more sensors than sources (2)

Algorithms without prewhitening:

best rank-$(R_1, R_2, \ldots, R_N)$ reduction



orthogonal iteration:

[*Kroonenberg '83*, *De Lathauwer '00*]

Rayleigh quotient iteration:

[*Zhang and Golub '01*, *De Lathauwer '04*]

Large mixtures: many sensors and sources

Gauss-Newton method for simultaneous matrix diagonalization

[*van der Veen '01*]

Conclusion

- PCA: directions of extremal oriented energy
  ICA: directions of statistically independent contributions
- Independence is a stronger condition than uncorrelatedness $\rightarrow$ unique solution
- Solution by means of multilinear algebra:
  - maximal diagonality
  - simultaneous EVD
  - CANDECOMP/PARAFAC with symmetry constraint
- Broad application domain
- Generalizations for convolutive mixtures