An Introduction to Tensor-Based Independent Component Analysis

Lieven De Lathauwer

K.U.Leuven
Belgium
Lieven.DeLathauwer@kuleuven-kortrijk.be

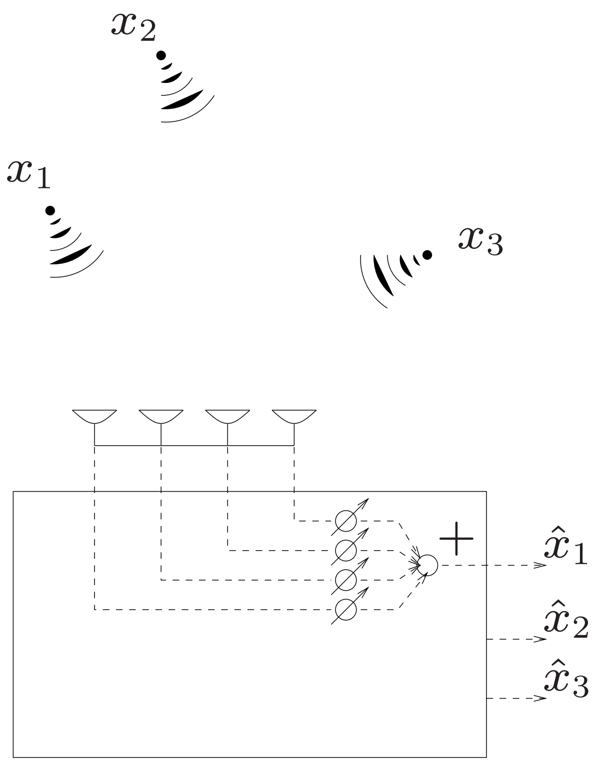
Overview

- Problem definition
- Higher-order statistics
- Basic ICA equations
- Specific prewhitening-based multilinear algorithms
- Application
- Higher-order-only schemes
- Variants for coloured sources
- Dimensionality reduction
- Conclusions

Independent Component Analysis (ICA)

Model:

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



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 **M**	matrix EVD/SVD
X_i indep.	**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)
$m_3^X = E\{X^3\}$	$c_3^X = E\{(X - m_X)^3\}$
$m_4^X = E\{X^4\}$	$c_4^X = E\{(X - m_X)^4\} - 3\sigma_X^4$

Characteristic Functions

First characteristic function:

$$\Phi_x(\omega) \stackrel{\text{def}}{=} \mathrm{E}\{e^{j\omega x}\} = \int_{-\infty}^{+\infty} p_x(x) e^{j\omega x} dx$$

Generates moments:

$$\Phi_x(\omega) = \sum_{k=0}^{\infty} m_k^X \frac{(j\omega)^k}{k!} \qquad (m_0 = 1)$$

Second characteristic function:

$$\Psi_x(\omega) \stackrel{\text{def}}{=} \ln \Phi_x(\omega)$$

Generates cumulants:

$$\Psi_x(\omega) = \sum_{k=1}^{\infty} c_k^X \frac{(j\omega)^k}{k!}$$

Moments and cumulants of a set of random variables:

Moments:

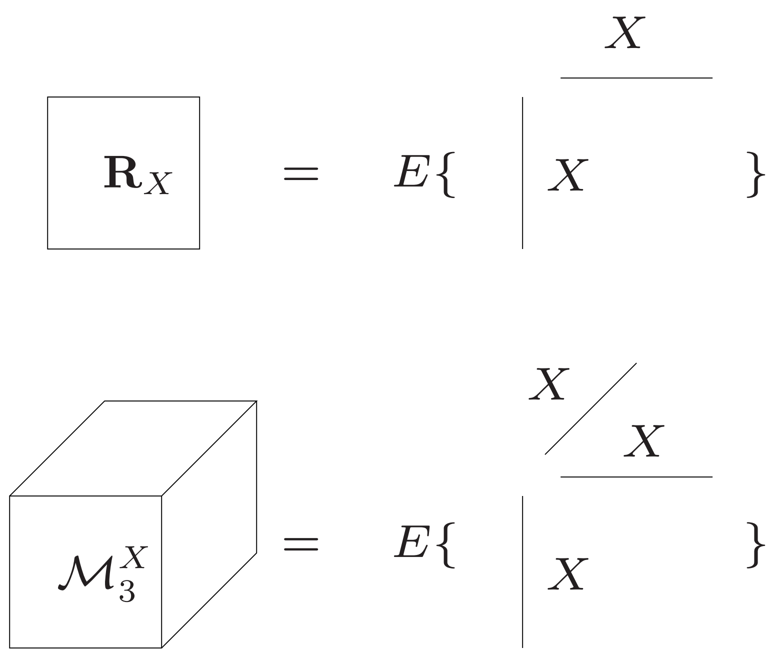
$$(\mathcal{M}_{\mathbf{x}}^{(N)})_{i_1 i_2 \ldots i_N} = \mathtt{Mom}(x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \stackrel{\text{def}}{=} \mathtt{E}\{x_{i_1} x_{i_2} \ldots x_{i_N}\}$$

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

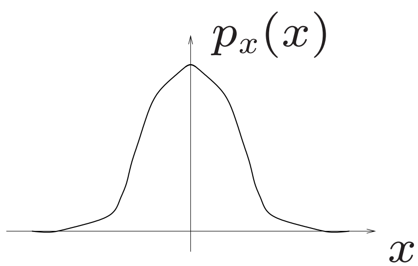
Multivariate case: e.g. moments:



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

HOS example

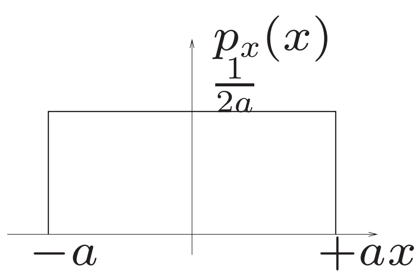
Gaussian distribution



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

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

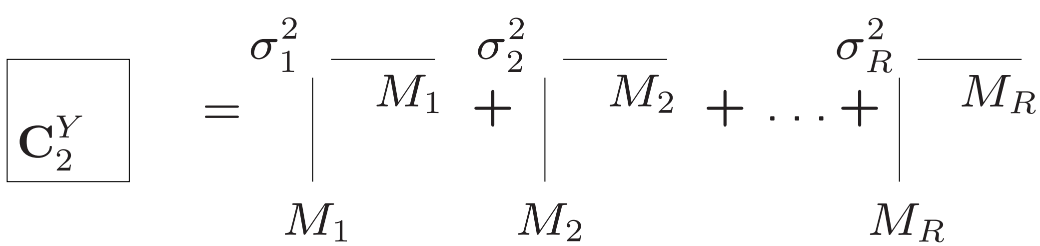
Model:

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

Second order:

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

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

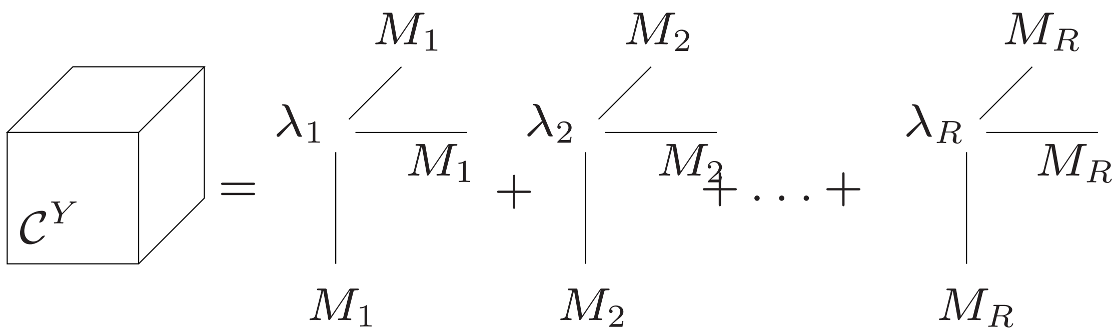


Higher order:

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

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

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

Prewhitening-based computation (2)

Matrix factorization:

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

Second order:

$$\mathbf{C}_2^Y = \mathbf{C}_2^X \bullet_1 \mathbf{M} \bullet_2 \mathbf{M} = \mathbf{T} \cdot \mathbf{T}^T$$

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

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
$\Rightarrow$ different solution strategies

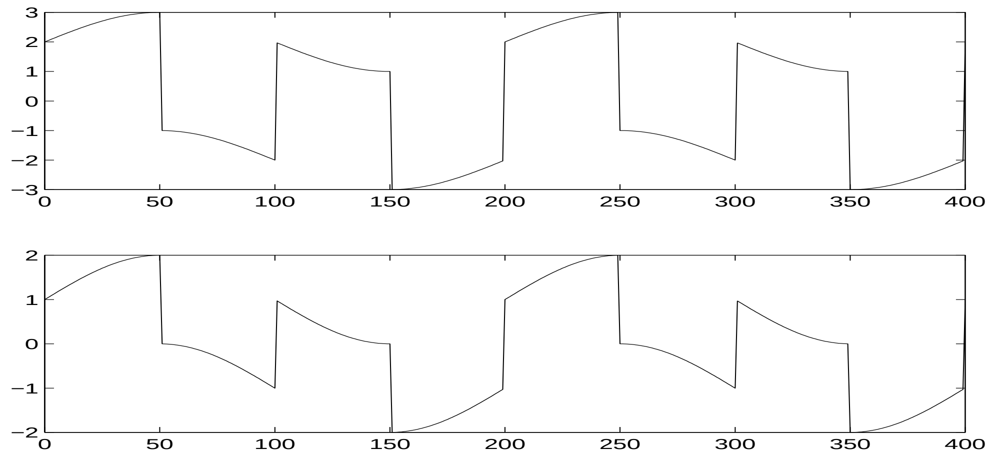
PCA versus ICA

ICA = higher-order fine-tuning of PCA:

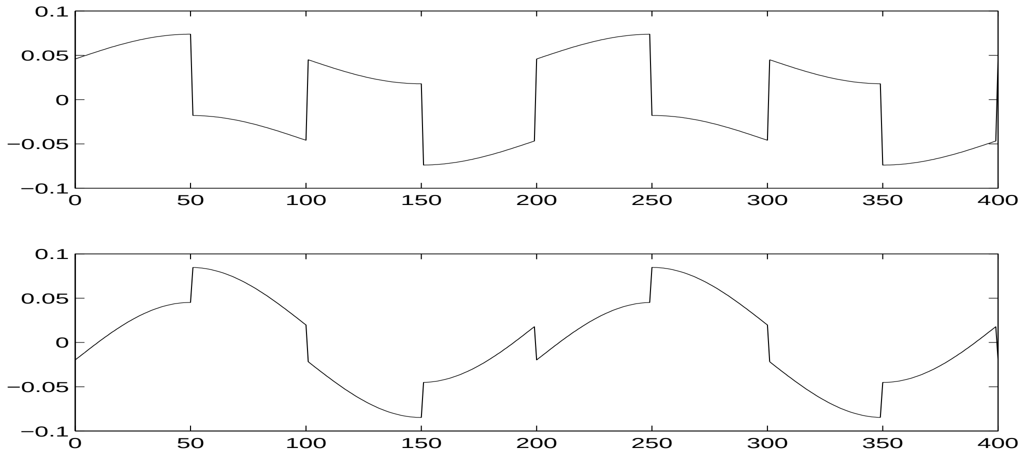
PCA	ICA
2nd-order	higher-order
matrix EVD	tensor EVD
uncorrelated sources	independent sources
column space **M**	**M** itself
always possible	depends on context

Computational cost:
cumulant estimation and diagonalization

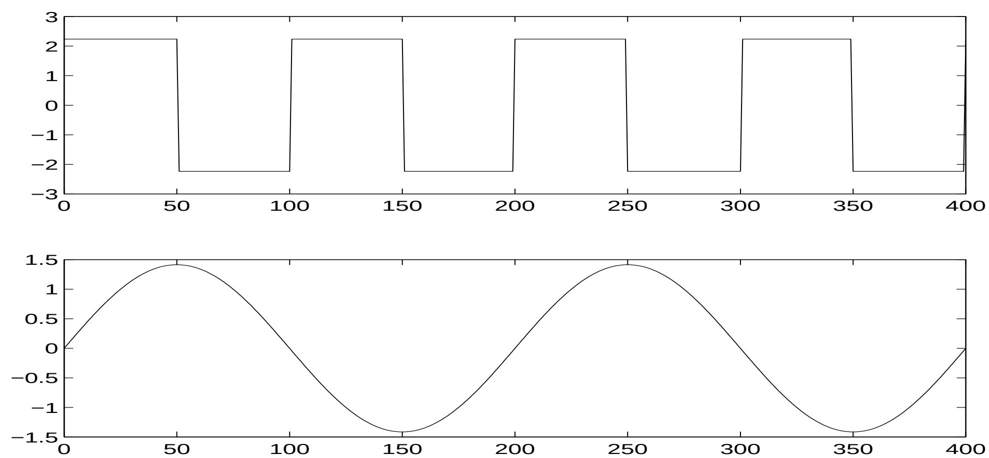
Illustration



Observations

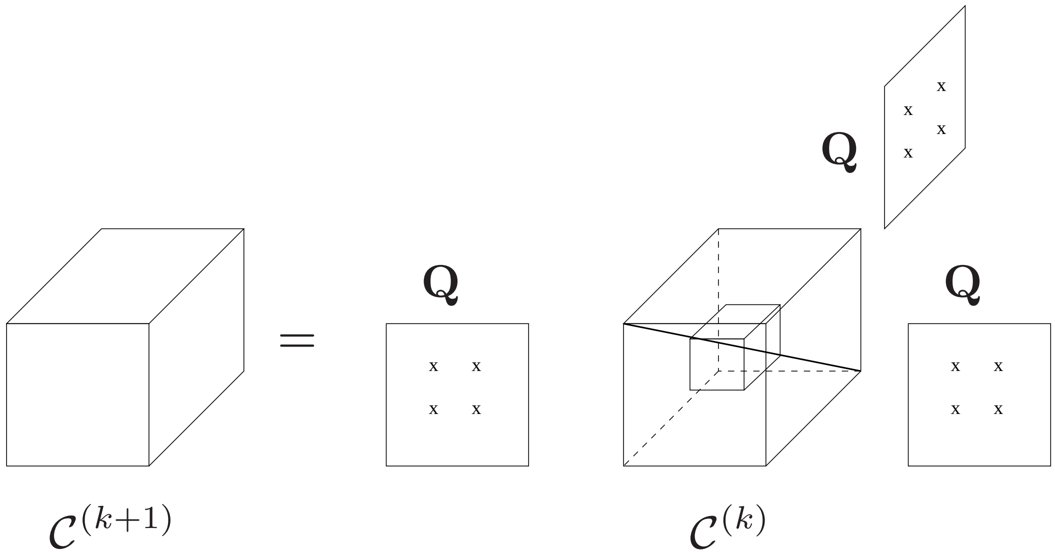


Sources estimated with PCA



Sources estimated with ICA

Algorithm 1: maximal diagonality

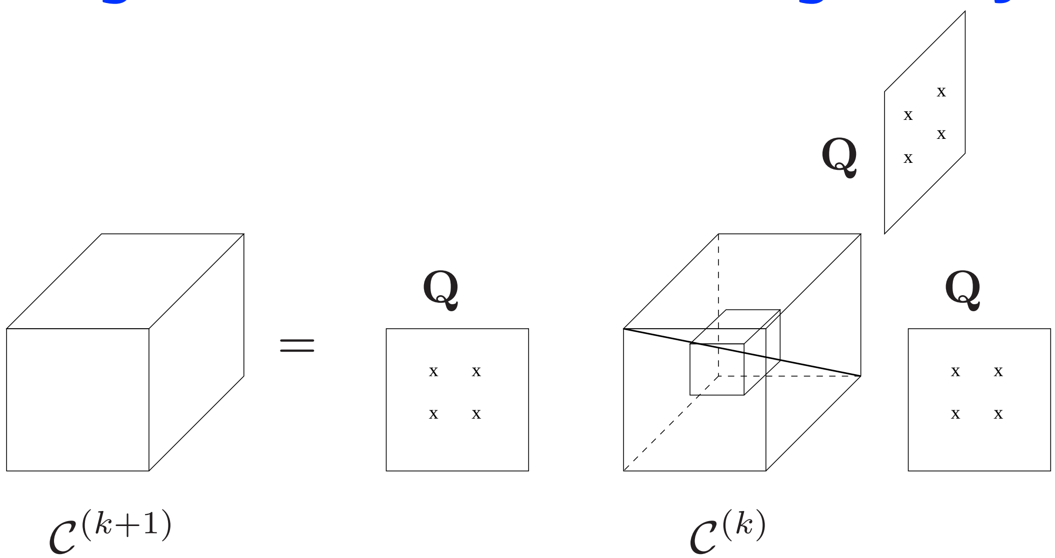


- Maximize energy on the diagonal by Jacobi-iteration
- Determination of optimal rotation angle:

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: maximal diagonality

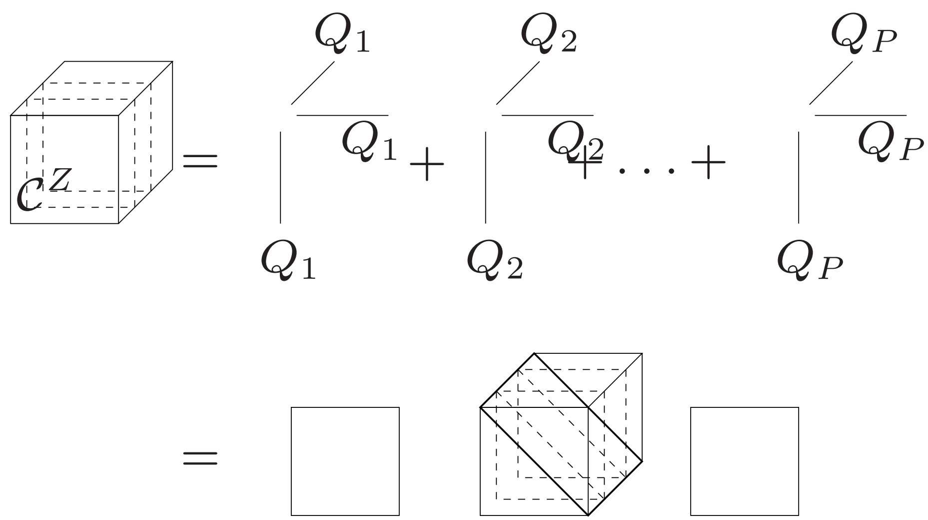


- Trace is not rotation invariant
- Maximize sum of diagonal entries by Jacobi-iteration
- Determination of optimal rotation angle:

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

[*Comon, Moreau, '97*]

Algorithm 3: simultaneous EVD



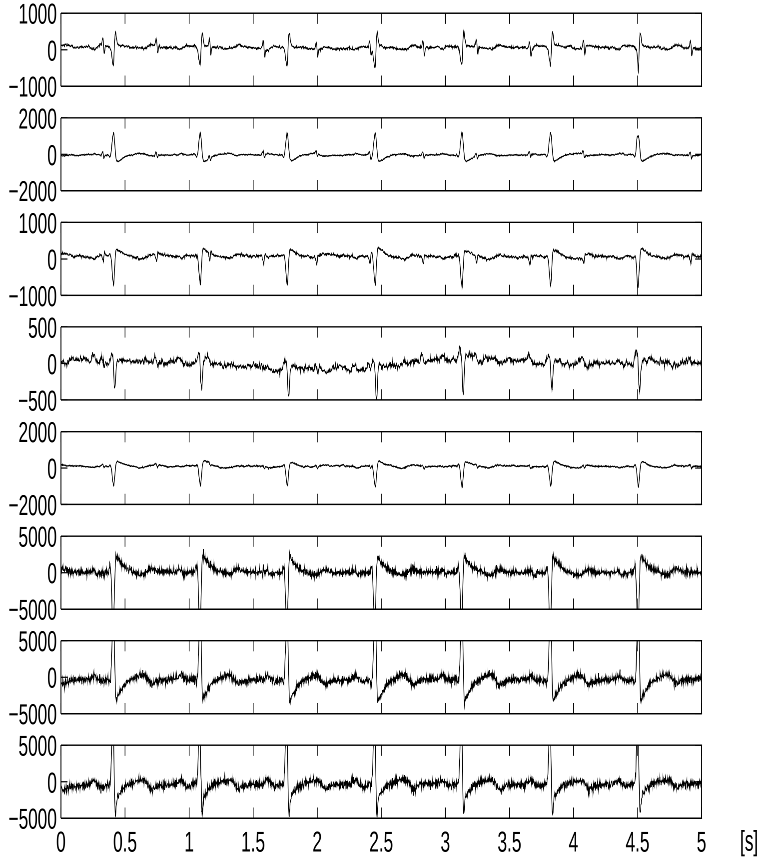
- Maximize energy on the diagonals by Jacobi-iteration
- Determination of optimal rotation angle:

real	roots polynomial degree 2
complex	roots polynomial degree 3

[*Cardoso '94 (JADE)*]

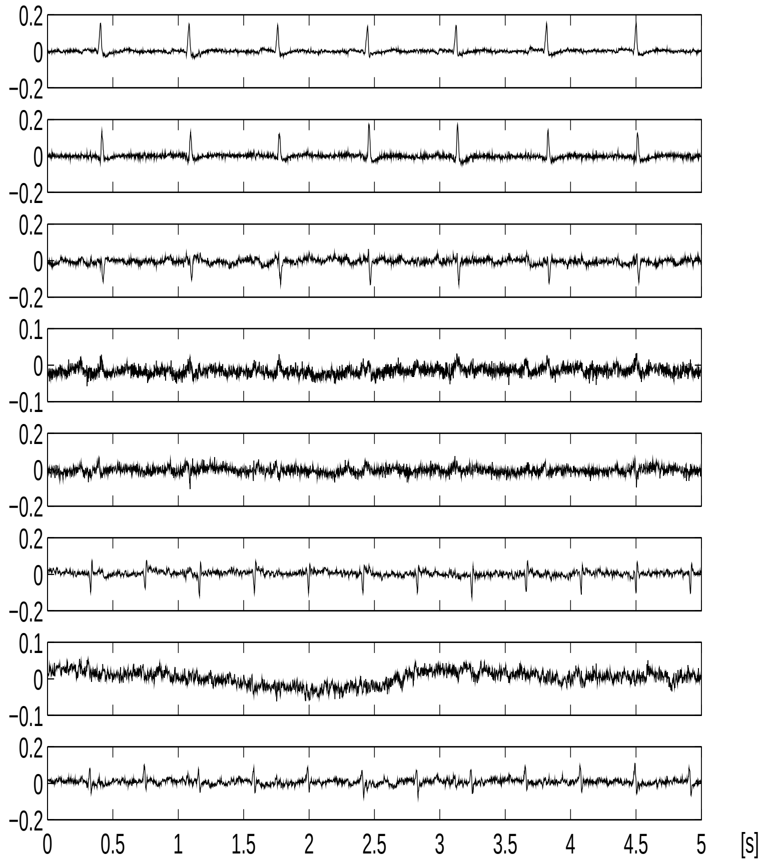
Application: fetal electrocardiogram extraction

Abdominal and thoracic recordings



ICA results for FECG extraction

Independent components:

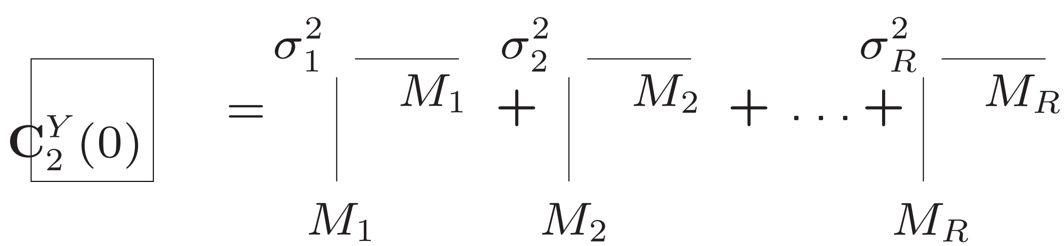


A variant for coloured sources

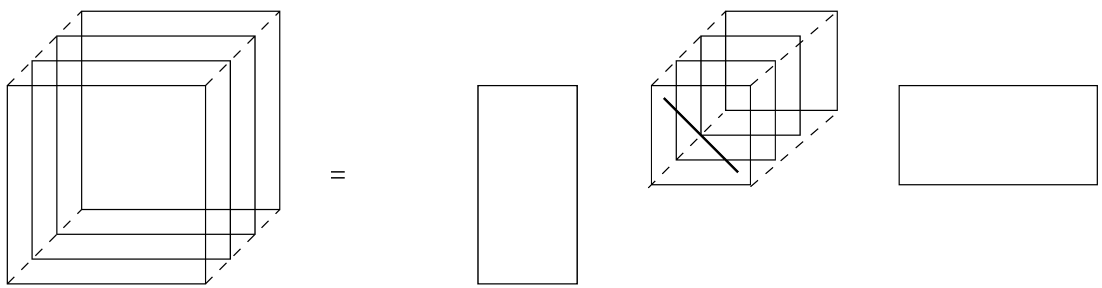
Condition: sources mutually uncorrelated, but individually correlated in time

Basic equations:

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



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



Variants: nonstationary sources, time-frequency representations, Hessian second characteristic function, . . .

[*Belouchrani et al. '97 (SOBI)*], [*De Lathauwer and Castaing '08*] (overcomplete)

Large mixtures: more sensors than sources

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

Prewhitening-based algorithms:

$$\underset{(P \times 1)}{Y} = \underset{(P \times R)(R \times 1)}{\mathbf{M} X} \qquad (P \gg R)$$

$$\underset{(P \times R)}{\mathbf{M}} = \underset{(P \times R)(R \times R)(R \times R)}{\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T}$$

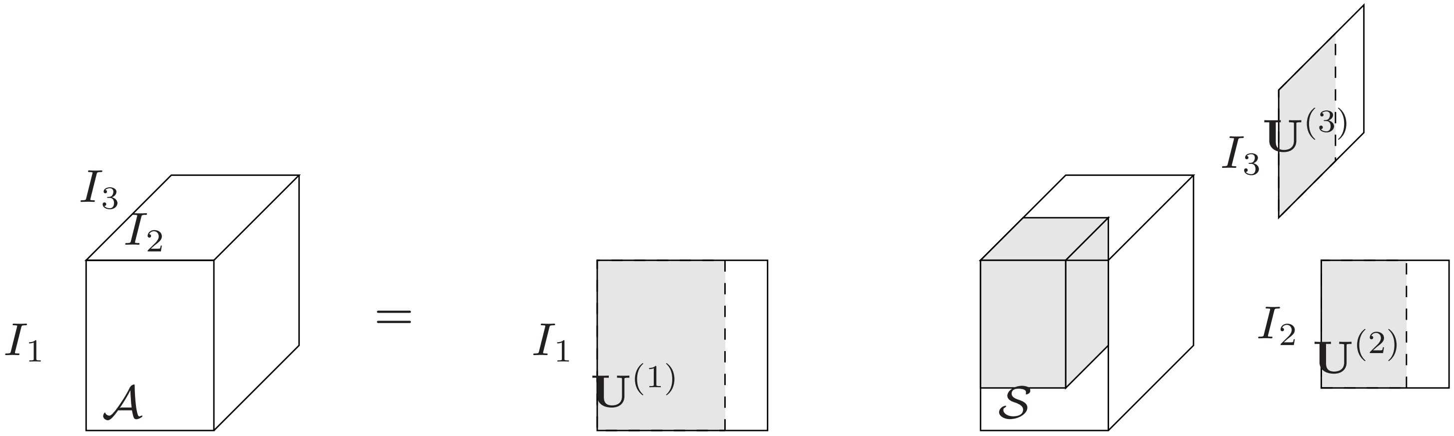
$$Z = \mathbf{S}^{-1} \cdot \mathbf{U}^T Y$$

$$\underset{(R \times 1)}{Z} = \underset{(R \times R)(R \times 1)}{\mathbf{V}^T X}$$

Large mixtures: more sensors than sources (2)

Algorithms without prewhitening:

best multilinear rank approximation



Tucker decomposition:

[*Tucker '64*], [*De Lathauwer '00*]

Large data sets:

[*Mahoney et al. '06*], [*Tyrtyshnikov et al. '06*], [*Oseledets et al. '08*]

Orthogonal iteration:

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

Optimization on manifolds:

- Newton [*Eldén and Savas '06*], [*Ishteva et al. '08*]
- Quasi-Newton [*Savas and Lim '08*]
- Trust region [*Ishteva et al. '09*]
- Conjugate gradient [*Ishteva et al. '09*]

Krylov method: [*Savas and Eldén '08*]

Conclusion

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