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:
\[ Y = MX + N \]
\[ (P \times 1) \quad (P \times R)(R \times 1) \quad (P \times 1) \]

Assumptions:
- columns of \( M \) are linearly independent
- components of \( X \) are statistically independent

Goal:
Identification of \( M \) 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 \((Y = MX)\)

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

<table>
<thead>
<tr>
<th>Condition</th>
<th>Identification</th>
<th>Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_i) uncorr.</td>
<td>column space (M)</td>
<td>matrix EVD/SVD</td>
</tr>
<tr>
<td>(X_i) indep.</td>
<td>(M)</td>
<td>tensor EVD/SVD</td>
</tr>
</tbody>
</table>
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:** \( C = A \circ B \iff c_{ijklm} = a_{ij} b_{klm} \)
  
  E.g. \( C = U V^T = U \circ V \)

  \[
  \begin{array}{c}
  U^{(3)} \\
  \hline \\
  U^{(2)} \\
  \hline \\
  U^{(1)}
  \end{array}
  \]

  \[ A = U^{(1)} \circ U^{(2)} \circ U^{(3)} \]

- **Matrix multiplication:**
  \( A = B \times_2 U \iff a_{ijk} = \sum_p b_{ipk} u_{jp} \)

  E.g. \( A = U \cdot S \cdot V^T = S \times_1 U \times_2 V \)

  \[ [Tucker '64] \]

\[ A = B \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \]
**HOS definitions**

Moments and cumulants of a random variable:

<table>
<thead>
<tr>
<th>Moments</th>
<th>Cumulants</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1^X = E{X}$</td>
<td>$C_1^X = E{X}$</td>
</tr>
<tr>
<td>“mean” ($m_X$)</td>
<td>“mean”</td>
</tr>
<tr>
<td>$M_2^X = E{X^2}$</td>
<td>$C_2^X = E{(X - m_X)^2}$</td>
</tr>
<tr>
<td>“correlation” ($R_X$)</td>
<td>“variance” ($\sigma_X^2$)</td>
</tr>
<tr>
<td>$M_3^X = E{X^3}$</td>
<td>$C_3^X = E{(X - m_X)^3}$</td>
</tr>
<tr>
<td>$M_4^X = E{X^4}$</td>
<td>$C_4^X = E{(X - m_X)^4} - 3\sigma_X^4$</td>
</tr>
</tbody>
</table>
Moments and cumulants of a set of random variables:

Moments:

\[
(M_x^{(N)})_{i_1 i_2 \ldots i_N} = \text{Mom}(x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \overset{\text{def}}{=} \mathbb{E}\{x_{i_1} x_{i_2} \ldots x_{i_N}\}
\]

Cumulants:

\[
(c_x)_i = \text{Cum}(x_i) \quad \overset{\text{def}}{=} \quad \mathbb{E}\{x_i\}
\]

\[
(C_x)_{i_1 i_2} = \text{Cum}(x_{i_1}, x_{i_2}) \quad \overset{\text{def}}{=} \quad \mathbb{E}\{x_{i_1} x_{i_2}\}
\]

\[
(C_x^{(3)})_{i_1 i_2 i_3} = \text{Cum}(x_{i_1}, x_{i_2}, x_{i_3}) \quad \overset{\text{def}}{=} \quad \mathbb{E}\{x_{i_1} x_{i_2} x_{i_3}\}
\]

\[
(C_x^{(4)})_{i_1 i_2 i_3 i_4} = \text{Cum}(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}) \quad \overset{\text{def}}{=} \quad \mathbb{E}\{x_{i_1} x_{i_2} x_{i_3} x_{i_4}\} - \mathbb{E}\{x_{i_1} x_{i_2}\} \mathbb{E}\{x_{i_3} x_{i_4}\} - \mathbb{E}\{x_{i_1} x_{i_3}\} \mathbb{E}\{x_{i_2} x_{i_4}\} - \mathbb{E}\{x_{i_1} x_{i_4}\} \mathbb{E}\{x_{i_2} x_{i_3}\}
\]

Order $\geq 2$: \( x_i \leftarrow x_i - \mathbb{E}\{x_i\} \)
Multivariate case: e.g. moments:

\[
\begin{align*}
1: \quad m_X \quad &\overset{\text{def}}{=} \quad E\{X\} \\
&\rightarrow \quad \text{vector} \\
2: \quad R_X \quad &\overset{\text{def}}{=} \quad E\{XX^T\} \\
&\rightarrow \quad \text{matrix} \\
3: \quad M^X_3 \quad &\overset{\text{def}}{=} \quad E\{X \circ X \circ X\} \\
&\rightarrow \quad \text{3rd order tensor} \\
4: \quad M^X_4 \quad &\overset{\text{def}}{=} \quad E\{X \circ X \circ X \circ X\} \\
&\rightarrow \quad \text{4th order tensor}
\end{align*}
\]
**HOS properties**

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

$$
\mathcal{M}^X_N = \mathcal{M}^X_N \times_1 A \times_2 A \ldots \times_N A \\
\mathcal{C}^X_N = \mathcal{C}^X_N \times_1 A \times_2 A \ldots \times_N A
$$

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

**Symmetry:**

$$
\text{Mom}(X_1, X_2, \ldots, X_I) = \text{Mom}(X_{P(1)}, X_{P(2)}, \ldots, X_{P(I)}) \\
\text{Cum}(X_1, X_2, \ldots, X_I) = \text{Cum}(X_{P(1)}, X_{P(2)}, \ldots, X_{P(I)})
$$

→ supersymmetric higher-order tensors

**Even distribution:** odd moments and cumulants = 0

**Partitioning of independent variables:**

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

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

**Moments:** e.g. $\text{Mom}(x, x, y, y) = E\{x^2\} \cdot E\{y^2\} \neq 0$
Non-Gaussianity:

Higher-order cumulants of a Gaussian variable are 0

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

Estimation:

Higher order: - harder to estimate
  - more entries

\( \Rightarrow \) (third- and) fourth-order cumulants
HOS example

### Gaussian distribution

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

<table>
<thead>
<tr>
<th>( n )</th>
<th>( m_x^{(n)} )</th>
<th>( c_x^{(n)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>( \sigma^2 )</td>
<td>( \sigma^2 )</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3 ( \sigma^4 )</td>
<td>0</td>
</tr>
</tbody>
</table>

### Uniform distribution

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

<table>
<thead>
<tr>
<th>( n )</th>
<th>( m_x^{(n)} )</th>
<th>( c_x^{(n)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>( a^2/3 )</td>
<td>( a^2/3 )</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3( a^4/5 )</td>
<td>-2( a^4/15 )</td>
</tr>
</tbody>
</table>
ICA: basic equations

Model:

\[ Y = MX \]

Second order:

\[
C_Y^2 = E\{YY^T\} = M \cdot C_X^2 \cdot M^T = C_X^2 \times_1 M \times_2 M
\]

uncorrelated sources: \( C_X^2 \) is diagonal
“diagonalization by congruence”

\[
C_Y^2 = \begin{bmatrix}
\sigma_1^2 & \sigma_2^2 & \cdots & \sigma_R^2 \\
M_1 & M_2 & \cdots & M_R
\end{bmatrix}
\]

Higher order:

\[
C_Y^4 = C_X^4 \times_1 M \times_2 M \times_3 M \times_4 M
\]

independent sources: \( C_X^4 \) is diagonal
“CANDECOMP / PARAFAC”

\[
C_Y^4 = \begin{bmatrix}
M_1 \\
M_2 \\
M_3 \\
M_R
\end{bmatrix} + \begin{bmatrix}
M_1 \\
M_2 \\
M_3 \\
M_R
\end{bmatrix} + \cdots + \begin{bmatrix}
M_1 \\
M_2 \\
M_3 \\
M_R
\end{bmatrix}
\]
Prewhitening-based computation

Model:

\[ Y = MX \]

First order:

\[ 0 = 0 \]

Second order:

\[
C_Y^2 = E\{YY^T\} \\
= M \cdot C_X^2 \cdot M^T \\
guess: M \cdot I \cdot M^T \\
= M \cdot M^T \\
= (M \cdot Q) \cdot (M \cdot Q)^T
\]

“square root”
EVD, Cholesky, . . .

Remark: PCA:

SVD of \( M \):

\[ M = U \cdot S \cdot V^T \]

\[ \Rightarrow C_Y^2 = (US) \cdot (US)^T \]

\[ = U \cdot S^2 \cdot U^T \]
Prewhitening-based computation (2)

Matrix factorization:

\[ M = T \cdot Q \]

Second order:

\[ C^Y_2 = C^X_2 \times_1 M \times_2 M = T \cdot T^T \]

Whitened r.v. \( Z = QX = T^{-1}Y \)

Higher order: ICA:

\[ Q \preceq C^Y_3 \text{ or } C^Y_4 \]

\[
\begin{align*}
C^Y_4 &= C^X_4 \times_1 M \times_2 M \times_3 M \times_4 M \\
C^Z_4 &= C^Y_4 \times_1 T^{-1} \times_2 T^{-1} \times_3 T^{-1} \times_4 T^{-1}
\end{align*}
\]

\[ \Rightarrow C^Z_4 = C^X_4 \times_1 Q \times_2 Q \times_3 Q \times_4 Q \]

“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
\( \Rightarrow \) different solution strategies
PCA versus ICA

ICA = higher-order fine-tuning of PCA:

<table>
<thead>
<tr>
<th>PCA</th>
<th>ICA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2nd-order</td>
<td>higher-order</td>
</tr>
<tr>
<td>matrix EVD</td>
<td>tensor EVD</td>
</tr>
<tr>
<td>uncorrelated sources</td>
<td>independent sources</td>
</tr>
<tr>
<td>column space $M$</td>
<td>$M$ itself</td>
</tr>
<tr>
<td>always possible</td>
<td>depends on context</td>
</tr>
</tbody>
</table>

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

\[ \mathbf{Z} = \frac{Q_1}{Q_1} + \frac{Q_2}{Q_2} + \ldots + \frac{Q_R}{Q_R} \]

- 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

\[ Y = MX \]

\[ C_Y^4 = C_X^4 \times_1 M \times_2 M \times_3 M \times_4 M \]

\( C_X^4 \) is diagonal \( \rightarrow \) CANDECOMP / PARAFAC

\[ C_Y^4 = M_1 + M_2 + \ldots + M_R \]
Soft whitening

\[
C^Y_2 = \begin{bmatrix}
\overline{M_1} + \overline{M_2} + \ldots + \overline{M_R}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
M_1 \quad \big/ \\
M_2 \quad \big/ \\
M_R \quad \big/ \\
\end{bmatrix}
\]

→ 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:

\[
C_Y^2(0) = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_R^2 \\ \\ M_1 & M_2 & \cdots & M_R \end{bmatrix} = M \cdot C_X^2(0) \cdot M^T
\]

\[
C_Y^2(\tau) = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_R^2 \\ \\ M_1 & M_2 & \cdots & M_R \end{bmatrix} = M \cdot C_X^2(\tau) \cdot M^T
\]

\[
C(\tau_K) = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_R^2 \\ \\ M_1 & M_2 & \cdots & M_R \end{bmatrix} = U_1 \cdot M_1 + U_2 \cdot M_2 + \cdots + U_R \cdot M_R
\]

[Belouchrani et al. '97 (SOBI)]
Large mixtures: more sensors than sources

Applications:

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

Prewhitening-based algorithms:

\[
Y = MX
\]
\[
(P \times 1) \quad (P \times R)(R \times 1) \quad (P \gg R)
\]

\[
M = U \cdot S \cdot V^T
\]
\[
(P \times R) \quad (P \times R)(R \times R)(R \times R)
\]

\[
Z = S^{-1} \cdot U^TY
\]
\[
Z = V^TX
\]
\[
(R \times 1) \quad (R \times R)(R \times 1)
\]
Large mixtures: more sensors than sources (2)

Algorithms without prewhitening:

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

\[
\begin{align*}
A I_3 I_2 I_1 &= \begin{bmatrix} I_1 \end{bmatrix} U^{(1)} \quad & I_3 I_2 I_1 \quad & I_3 I_2 U^{(2)} \\
&
\end{align*}
\]

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 → unique solution
- Solution by means of multilinear algebra:
  - maximal diagonality
  - simultaneous EVD
  - CANDECOMP/PARAFAC with symmetry constraint
- Broad application domain
- Generalizations for convolutive mixtures