MMDS Workshop, Stanford, June 24, 2006

Independent Component Analysis (ICA)
viewed as
a Tensor Decomposition

Pierre COMON

Lab. I3S
CNRS & University of Nice
Sophia-Antipolis, France

Modeling

Observation model

$$\boldsymbol{x} = \boldsymbol{H}\,\boldsymbol{s} + \boldsymbol{v} \tag{1}$$

- $\boldsymbol{x}$: observed, dim K
- P: source vector, dim P
- $\boldsymbol{H}$: $K \times P$ mixing matrix
- $\boldsymbol{v}$: additive noise

Taxonomy

One additional assumption is required on sources s_i:

- mutually independent sources
- discrete sources
- colored sources
- nonstationary sources

General Concepts

Principal component Analysis (PCA)

Goal

Given a K-dimensional r.v., $\boldsymbol{x}$, find $\boldsymbol{U}$ and $\boldsymbol{z}$ such that

- Observation

$$\boldsymbol{x} = \boldsymbol{U}\,\boldsymbol{z}$$

- $\boldsymbol{z}$ has uncorrelated components z_i

NB: Because of lack of uniqueness, $\boldsymbol{U}$ is often assumed to be unitary.

Independent Component Analysis (ICA)

Goal

Given a K-dimensional r.v., $\boldsymbol{x}$, find $\boldsymbol{H}$ and $\boldsymbol{s}$ such that

- Observation

$$\boldsymbol{x} = \boldsymbol{H}\,\boldsymbol{s} \tag{2}$$

- $\boldsymbol{s}$ has mutually statistically independent components s_i

➨ *"Blind"* Source Separation: only outputs x_i are observed.

General Concepts

Uniqueness

Inherent indeterminations

if $\boldsymbol{s}$ has independent components s_i, so has $\boldsymbol{\Lambda P}\,\boldsymbol{s}$

where $\boldsymbol{\Lambda}$ is invertible diagonal and $\boldsymbol{P}$ permutation

Solutions

If $(\boldsymbol{A},\ \boldsymbol{s})$ solution, then $(\boldsymbol{A\Lambda P},\ \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{s})$ also is.

- *"Essential uniqueness"*: unique up to a *trivial filter*, i.e. a scale-permutation
- Whole equivalence class of solutions ⇒ Look for one representative.

General Concepts

Decorrelation vs Independence

Example 1: Mixture of 2 identically distributed sources

Consider the mixture of two independent sources

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

where $\mathrm{E}\{s_i^2\} = 1$ and $\mathrm{E}\{s_i\} = 0$. Then x_i are *uncorrelated*:

$$\mathrm{E}\{x_1\, x_2\} = \mathrm{E}\{s_1^2\} - \mathrm{E}\{s_2^2\} = 0$$

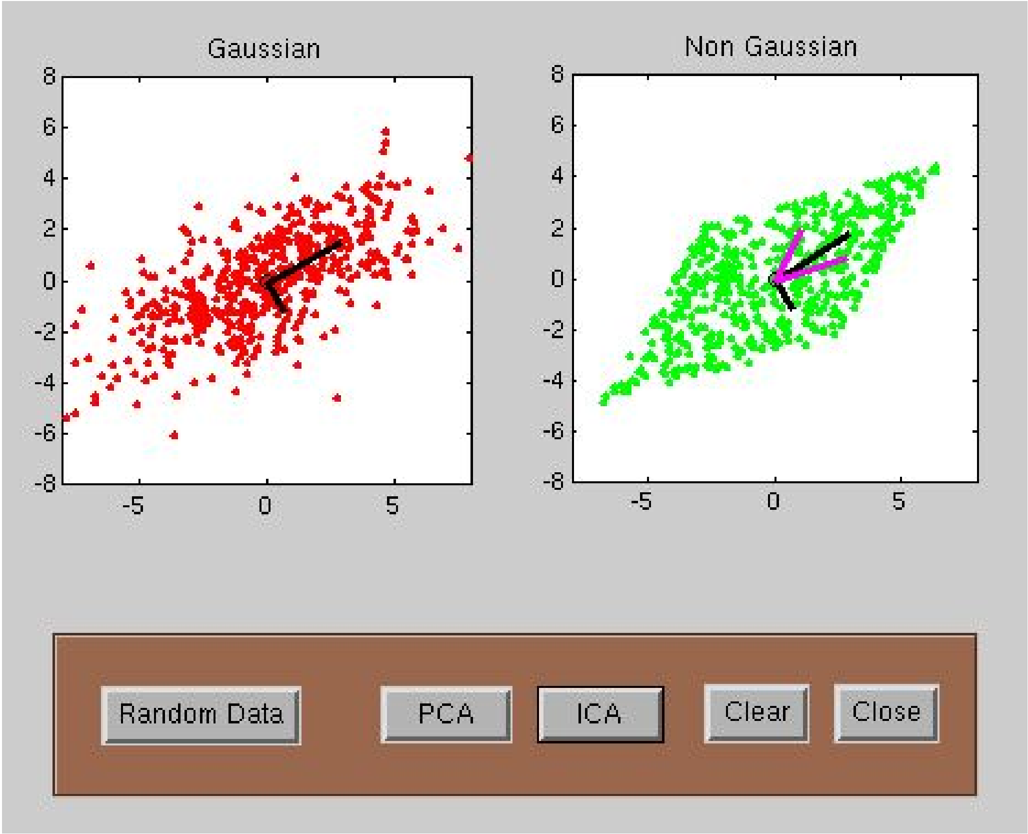
But x_i are *not independent* since, for instance:

$$\mathrm{E}\{x_1^2\, x_2^2\} - \mathrm{E}\{x_1^2\}\mathrm{E}\{x_2^2\} = \mathrm{E}\{s_1^4\} + \mathrm{E}\{s_2^4\} - 6 \neq 0$$

General Concepts

PCA vs ICA

Example 2: 2 sources and 2 sensors



Application Areas (1)

- Sensor Array Processing
  - Speech
  - Localization with ill calibrated antennas
  - Detection and/or extraction with unknown antennas
    (eg. sonar buoys, biomedical, audio, nuclear plants...)
  - Blind extraction (eg. ComInt: interception, surveillance)
  - Localization with reduced diversity (eg. Air traffic control)

Application Areas (2)

- Factor Analysis
  - Chemometrics
  - Econometrics
  - Psychology
- Compression
- Arithmetic Complexity
- Machine Learning
- Exploratory Analysis

General bibliography

- **Books on HOS, ICA, or Multi-Way:**

  Lacoume-Amblard-Comon'97 (but in French)

  Hyvarinen-Karhunen-Oja'01 (but dedicated only to FastICA)

  Smilde-Bro-Geladi'04 (but dedicated only to Factor Analysis)

  Comon-DeLathauwer (will cover more topics, but you have to wait!)

- **Other related books:**

  Kagan-Linnik-Rao'73

  McCullagh'87

  Nikias-Petropulu'93

  Haykin'2000

Spatial whitening

Standardization via PCA

Definition

PCA is based on second order statistics

- Observed random variable $\boldsymbol{x}$ of dimension K. Then $\exists(\boldsymbol{U}, \boldsymbol{z})$:

$$\boldsymbol{x} = \boldsymbol{U}\boldsymbol{z},\ \boldsymbol{U} \text{ unitary}$$

  where *Principal Components* z_i are uncorrelated

  ith column $\boldsymbol{u}_i$ of $\boldsymbol{U}$ is called *ith PC Loading vector*

- Two possible calculations:
  - EVD of Covariance $\boldsymbol{R}_x$: $\boldsymbol{R}_x = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\mathsf{H}}$
  - Sample estimate by SVD: $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$

Spatial whitening

Summary

Find a linear transform $\boldsymbol{L}$ such that vector $\tilde{\boldsymbol{x}} \stackrel{\text{def}}{=} \boldsymbol{L}\boldsymbol{x}$ has unit covariance. Many possibilities, including:

- PCA yields $\tilde{\boldsymbol{x}} = \boldsymbol{\Sigma}^{-1}\,\boldsymbol{U}^{\mathsf{H}}\,\boldsymbol{x}$
- Cholesky $\boldsymbol{R}_x = \boldsymbol{L}\,\boldsymbol{L}^{\mathsf{H}}$ yields $\tilde{\boldsymbol{x}} = \boldsymbol{L}^{-1}\,\boldsymbol{x}$

Remarks

- Infinitely many possibilities: $\boldsymbol{L}$ is as good as $\boldsymbol{L}\,\boldsymbol{Q}$, for any unitary $\boldsymbol{Q}$.
- If $\boldsymbol{R}_x$ not invertible, then $\boldsymbol{L}$ not invertible (ill-posed). One may use pseudo-inverse of $\boldsymbol{\Sigma}$ in PCA to compute $\boldsymbol{L}$, or regularize $\boldsymbol{R}_x$.

PCA by pair sweeping

Plane rotations

Application of a Givens rotation on both sides of a matrix allows to set a pair of zeros in a symmetric matrix:

$$\begin{pmatrix} c & . & s & . \\ . & 1 & . & . \\ -s & . & c & . \\ . & . & . & 1 \end{pmatrix} \boldsymbol{A} \begin{pmatrix} c & . & -s & . \\ . & 1 & . & . \\ s & . & c & . \\ . & . & . & 1 \end{pmatrix} = \begin{pmatrix} X & x & 0 & x \\ x & . & x & . \\ 0 & x & X & x \\ x & . & x & . \end{pmatrix}$$

Same result obtained:

- either by setting 0
- or by maximizing X's

PCA by pair sweeping

Jacobi sweeping for PCA

Cyclic by rows/columns algorithm for a 4×4 real symmetric matrix

$$\begin{pmatrix} . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{pmatrix} \rightarrow \begin{pmatrix} X & 0 & x & x \\ 0 & X & x & x \\ x & x & . & . \\ x & x & . & . \end{pmatrix} \rightarrow \begin{pmatrix} X & x & 0 & x \\ x & . & x & . \\ 0 & x & X & x \\ x & . & x & . \end{pmatrix} \rightarrow \begin{pmatrix} X & x & x & 0 \\ x & . & . & x \\ x & . & . & x \\ 0 & x & x & X \end{pmatrix} \rightarrow$$

$$\begin{pmatrix} . & x & x & 0 \\ x & X & 0 & x \\ x & 0 & X & x \\ 0 & x & x & . \end{pmatrix} \rightarrow \begin{pmatrix} . & x & . & x \\ x & X & x & 0 \\ . & x & . & x \\ x & 0 & x & X \end{pmatrix} \rightarrow \begin{pmatrix} . & . & x & x \\ . & . & x & x \\ x & x & X & 0 \\ x & x & 0 & X \end{pmatrix}$$

X: maximized, x: minimized, 0: canceled, $.$: unchanged

Statistical Independence

Definition

Components s_k of a K-dimensional r.v. $\boldsymbol{s}$ are *mutually independent*

$$\Updownarrow$$

The *joint* pdf equals the *product of marginal* pdf's:

$$p_{\boldsymbol{s}}(\boldsymbol{u}) = \prod_k p_{s_k}(u_k) \tag{3}$$

Definition

Components s_k of $\boldsymbol{s}$ are *pairwise independent* $\Leftrightarrow$ Any pair of components (s_k, s_ℓ) are mutually independent.

Mutual vs Pairwise independence (1)

Example 3: Pairwise but not Mutual independence

- 3 mutually independent BPSK sources, $x_i \in \{-1, 1\}$, $1 \leq i \leq 3$
- Define $x_4 = x_1 x_2 x_3$. Then x_4 is also BPSK, *dependent on x_i*
- x_k are *pairwise independent*:

  $p(x_1 = a, x_4 = b) = p(x_4 = b \,|\, x_1 = a).p(x_1 = a) =$

  $p(x_2\, x_3 = b/a).p(x_1 = a)$

  But x_1 and $x_2\, x_3$ are BPSK $\Rightarrow$

  $p(x_2\, x_3 = b/a).p(x_1 = a) = \frac{1}{2} \cdot \frac{1}{2}$
- But x_k obviously not mutually independent, $1 \leq k \leq 4$

  In particular, $\text{Cum}\{x_1, x_2, x_3, x_4\} = 1 \neq 0$

Mutual vs Pairwise independence (2)

Darmois's Theorem (1953)

Let two random variables be defined as linear combinations of independent random variables x_i:

$$X_1 = \sum_{i=1}^{N} a_i\, x_i, \quad X_2 = \sum_{i=1}^{N} b_i\, x_i$$

Then, if X_1 and X_2 are independent, those x_j for which $a_j b_j \neq 0$ are Gaussian.

Mutual vs Pairwise independence (3)

Corollary

If $\boldsymbol{z} = \boldsymbol{C}\,\boldsymbol{s}$, where s_i are independent r.v., with at most one of them being Gaussian, then the following properties are equivalent:

1. Components z_i are pairwise independent

2. Components z_i are mutually independent

3. $\boldsymbol{C} = \boldsymbol{\Lambda}\,\boldsymbol{P}$, with $\boldsymbol{\Lambda}$ diagonal and $\boldsymbol{P}$ permutation

Properties of Cumulants

- **Multi-linearity** (also enjoyed by moments):

$$\mathrm{Cum}\{\alpha X,\, Y, .., Z\} = \alpha\, \mathrm{Cum}\{X,\, Y, .., Z\} \quad (4)$$
$$\mathrm{Cum}\{X_1 + X_2,\, Y, .., Z\} = \mathrm{Cum}\{X_1,\, Y, .., Z\} + \mathrm{Cum}\{X_2,\, Y, .., Z\}$$

- **Cancellation**: If $\{X_i\}$ can be partitioned into 2 groups of independent r.v., then

$$\mathrm{Cum}\{X_1,\, X_2, .., X_r\} = 0 \quad (5)$$

- **Additivity**: If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are *independent*, then

$$\begin{aligned}\mathrm{Cum}\{X_1 + Y_1,\, X_2 + Y_2, .., X_r + Y_r\} &= \mathrm{Cum}\{X_1,\, X_2, .., X_r\} \\ &+ \mathrm{Cum}\{Y_1,\, Y_2, .., Y_r\}\end{aligned}$$

- **Inequalities**, e.g.:

$$\mathcal{K}_{(3)}^2 \leq \mathcal{K}_{(4)} + 2$$

Optimization Criteria

Contrast criteria: definition

Axiomatic definition

A *Contrast* optimization criterion Υ should enjoy 3 properties:

- *Invariance*: Υ should not change under the action of trivial filters (Permutation-Scale)
- *Domination*: If sources are already separated, any filter should decrease (or leave unchanged) Υ
- *Discrimination*: The maximum achievable value should be reached only when sources are separated (i.e. all absolute maxima are related to each other by trivial filters)

NB: idea first developed by Donoho for blind (scalar) equalization [DON81]

Optimization Criteria

Mutual Information

$\Upsilon \stackrel{\text{def}}{=} -I(p_{\boldsymbol{z}})$ is a contrast

- Invariant by scale change and permutation
- Always negative
- Null if and only if components are independent

Optimization Criteria

CoM Family of contrasts

When observations are standardized, and when only *unitary transforms* are considered, then the following are contrast functions:

- If at most 1 source has a null skewness [COM94b]:

$$\Upsilon_{2,3} = \sum_{i=1}^{P} (\kappa_{iii})^2, \quad \kappa_{iii} \stackrel{\text{def}}{=} \mathcal{C}_{\boldsymbol{z}\, iii}$$

- If at most 1 source has a null kurtosis [COM94a]:

$$\Upsilon_{2,4} = \sum_{i=1}^{P} (\kappa_{ii}^{ii})^2, \quad \kappa_{ii}^{ii} \stackrel{\text{def}}{=} \mathcal{C}_{\boldsymbol{z}\, ii}^{\ \ ii}$$

- If at most 1 source has a null standardized Cumulant of order $r \stackrel{\text{def}}{=} p + q > 2$, and for any $\alpha \geq 1$:

$$\Upsilon_{\alpha,r} = \sum_{i=1}^{P} |\kappa_{i\,(p)}^{(q)}|^{\alpha}, \quad \kappa_{i\,(p)}^{(q)} \stackrel{\text{def}}{=} \text{Cum}\{\underbrace{z_i, \dots, z_i}_{p \text{ times}}, \underbrace{z_i^*, \dots, z_i^*}_{q \text{ times}}\}$$

Optimization Criteria

General Family of contrasts

- **Theorem** All CoM contrasts belong to the larger family :

$$\Upsilon_g(\boldsymbol{z}) = \sum_i g(|\kappa_{i\,(p)}^{(q)}|) \tag{6}$$

where $g(\cdot)$ is convex strictly increasing, and $p + q > 2$.

Numerical Algorithms

What problem are they supposed to solve?

- Find Absolute maximum of a rational function in several variables

What kind of algorithms?

- Gradient ascent: the simplest
- Gradient-based ascents (Newton, quasi-Newton, conjugate gradient..)
- Quasi-algebraic algorithms: try to avoid *local maxima*
- Algebraic algorithms: find all absolute maxima in *closed-form*

Algebraic algorithms

The 2-dimensional problem

- Assume data x have been standardized into $\tilde{\boldsymbol{x}}$.

- Then one looks for an estimate $\boldsymbol{z}$ of the source vector $\boldsymbol{s}$ as:

$$\boldsymbol{z} = \boldsymbol{Q}\,\tilde{\boldsymbol{x}}$$

where $\boldsymbol{Q}$ is unitary, and may be assumed of the form:

$$\boldsymbol{Q} = \begin{pmatrix} \cos\beta & \sin\beta\, e^{\jmath\varphi} \\ -\sin\beta\, e^{-\jmath\varphi} & \cos\beta \end{pmatrix} = \frac{1}{\sqrt{1+\theta\theta^*}} \begin{pmatrix} 1 & \theta \\ -\theta^* & 1 \end{pmatrix} \qquad (7)$$

where $\theta \stackrel{\text{def}}{=} \tan\beta\, e^{\jmath\varphi}$ denotes the complex tangent, and $\beta \in]-\pi/2,\ \pi/2]$.

Solution of the 2-dimensional problem (1)

Closed-form solution for absolute maximum of:

- $\Upsilon_{1,4}$ in $\mathbb{R}$
- $\Upsilon_{2,3}$ in $\mathbb{R}$ [COM94b]
- $\Upsilon_{2,4}$ in $\mathbb{R}$ [COM94a]
- $\Upsilon_{2,3}$ in $\mathbb{C}$ [dLdMV01]
- $\Upsilon_{1,4}$ in $\mathbb{C}$ [COM01]

Algebraic algorithms

Invariance & Indeterminacy (1)

- There is a whole class of equivalent absolute maxima, which can be deduced from each other by trivial filtering
- In the 2×2 real case, there are 8 equivalent absolute maxima, generated by two $\boldsymbol{P}\,\boldsymbol{\Lambda}$ transformations:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

- In the complex case, there are infinitely many, when $\varphi \in \mathbb{R}$.
- Expression (7) fixes this indeterminacy, so that only 2 solutions remain

Algebraic algorithms

What is the problem in dimension 2 ?

- $\Upsilon_{\alpha,r}$ is a homogeneous trigonometric polynomial in $(\cos\beta, \sin\beta)$ of *degree $\alpha\, r$*.
- And we want a closed-form (algebraic) solution
- But only polynomials of a single variable of *degree at most 4* can generally be rooted algebraically
- **Our problem:** check out whether $\Upsilon_{\alpha,r}$ could be transformed into a particular function that can be algebraically maximized

Algebraic algorithms

Example (1): maximization of : $\Upsilon_{2,3}$ in $\mathbb{R}$

$\Upsilon_{2,3} = \kappa_{111}^2 + \kappa_{222}^2$ can be proved to be a quadratic form $\boldsymbol{u}^\mathsf{T}\boldsymbol{B}\,\boldsymbol{u}$ where

$$\boldsymbol{u} \stackrel{\text{def}}{=} [\cos 2\beta, \;\; \sin 2\beta]^\mathsf{T} \tag{8}$$

and

$$\boldsymbol{B} \stackrel{\text{def}}{=} \begin{pmatrix} a_1 & 3\,a_4/2 \\ 3\,a_4/2 & 9\,a_2/4 + 3\,a_3/2 + a_1/4 \end{pmatrix}$$

with [dLdMV01]:

$$\begin{aligned} a_1 &= \gamma_{111}^2 + \gamma_{222}^2 \\ a_2 &= \gamma_{112}^2 + \gamma_{122}^2 \\ a_3 &= \gamma_{111}\,\gamma_{122} + \gamma_{112}\,\gamma_{222} \\ a_4 &= \gamma_{122}\,\gamma_{222} - \gamma_{111}\,\gamma_{112} \end{aligned}$$

Algebraic algorithms

Example (2): maximization of contrast $\Upsilon_{1,4}$ in $\mathbb{R}$

- **Input-Output relations**

$$
\begin{aligned}
\kappa_1 &= \gamma_1 \cos^4\beta + 4\gamma_{1112}\cos^3\beta\sin\beta + 6\gamma_{1122}\cos^2\beta\sin^2\beta \\
&+ 4\gamma_{1222}\cos\beta\sin^3\beta + \gamma_2\sin^4\beta \\
\kappa_2 &= \gamma_1 \sin^4\beta - 4\gamma_{1112}\cos\beta\sin^3\beta + 6\gamma_{1122}\cos^2\beta\sin^2\beta \\
&- 4\gamma_{1222}\cos^3\beta\sin\beta + \gamma_2\cos^4\beta
\end{aligned}
$$

- **Then** $\varepsilon\Upsilon_{1,4} = \kappa_1 + \kappa_2 =$

$$
[\cos 2\beta \;\; \sin 2\beta] \begin{pmatrix} \gamma_1+\gamma_2 & \gamma_{1112}-\gamma_{1222} \\ \gamma_{1112}-\gamma_{1222} & \frac{\gamma_1+\gamma_2}{2} + 3\gamma_{1122} \end{pmatrix} \begin{bmatrix} \cos 2\beta \\ \sin 2\beta \end{bmatrix}
$$

- **Conclusion:** again entirely *algebraic* since dominant eigenvector of a matrix of size < 4.

Algebraic algorithms

Example (3): maximization of of contrast $\Upsilon_{1,4}$ in $\mathbb{C}$

- Define $\kappa_i = \text{Cum}\{z_i, z_i, z_i^*, z_i^*\}$, $\gamma_{ij}^{k\ell} = \text{Cum}\{\tilde{x}_i, \tilde{x}_j, \tilde{x}_k^*, \tilde{x}_\ell^*\}$
- **Then...** again a quadratic form

$$\varepsilon\Upsilon_{1,4} \;=\; \kappa_1 + \kappa_2 = \boldsymbol{u}^{\mathsf{T}}\,\boldsymbol{B}\,\boldsymbol{u}$$

with

$$\boldsymbol{u}^{\mathsf{T}} = [\cos 2\beta \quad \sin 2\beta\cos\varphi \quad \sin 2\beta\sin\varphi]$$

and

$$\boldsymbol{B} \;=\; \begin{pmatrix} \gamma_{1111} + \gamma_{2222} & \Re\{\delta\} & -\Im\{\delta\} \\ \Re\{\delta\} & 2\gamma_{12}^{12} + \Re\{\gamma_{22}^{11}\} & \Im\{\gamma_{22}^{11}\} \\ -\Im\{\delta\} & \Im\{\gamma_{22}^{11}\} & 2\gamma_{12}^{12} - \Re\{\gamma_{22}^{11}\} \end{pmatrix};$$

$$\delta \;=\; \gamma_{12}^{11} - \gamma_{22}^{12}$$

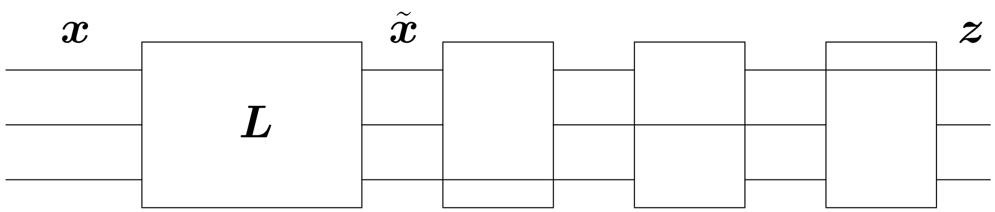
Conclusion: unexpectedly *entirely algebraic!* [COM01]

Quasi-algebraic algorithms

Jacobi Sweeping

Cyclic sweeping with fixed ordering

Example in dimension $P = 3$:



Carl Jacobi, 1804-1851

Quasi-algebraic algorithms

Jacobi Sweeping for tensors

Question: Why not select another ordering, e.g. process pairs having cross cumulants of largest magnitude?

Response: the computational complexity would be dominated by the computation of the tensor entries themselves!

Quasi-algebraic algorithms

Jacobi Sweeping for tensors

Joint Block Algorithm: Sweeping a $3 \times 3 \times 3$ tensor

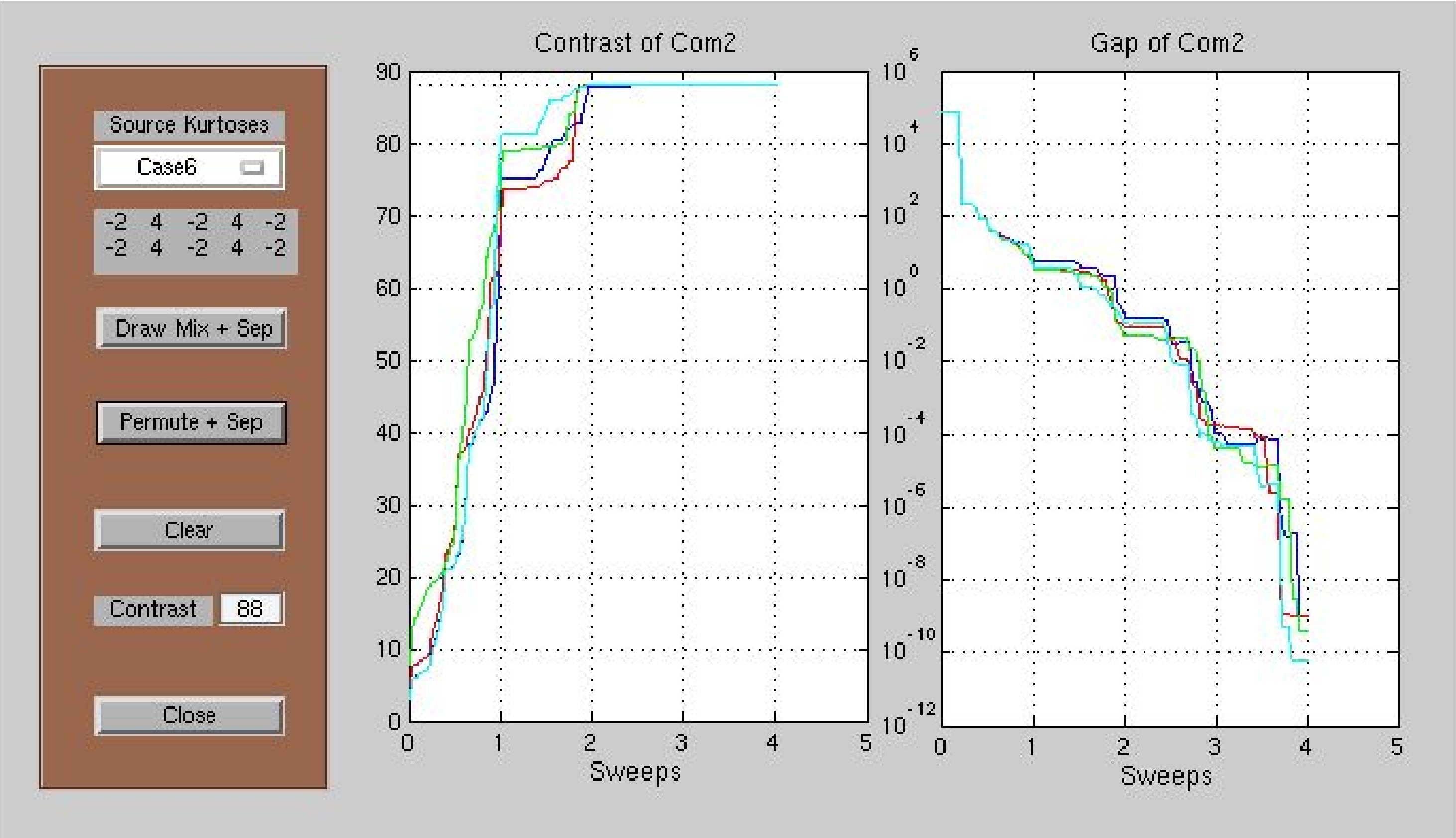
$$
\begin{array}{c}
\begin{pmatrix} X & x & x \\ x & x & x \\ x & x & . \end{pmatrix} \\
\begin{pmatrix} x & x & x \\ x & X & x \\ x & x & . \end{pmatrix} \\
\begin{pmatrix} x & x & x \\ x & x & x \\ x & x & . \end{pmatrix}
\end{array}
\longrightarrow
\begin{array}{c}
\begin{pmatrix} X & x & x \\ x & . & x \\ x & x & x \end{pmatrix} \\
\begin{pmatrix} x & x & x \\ x & . & x \\ x & x & x \end{pmatrix} \\
\begin{pmatrix} x & x & x \\ x & . & x \\ x & x & X \end{pmatrix}
\end{array}
\longrightarrow
\begin{array}{c}
\begin{pmatrix} . & x & x \\ x & x & x \\ x & x & x \end{pmatrix} \\
\begin{pmatrix} . & x & x \\ x & X & x \\ x & x & x \end{pmatrix} \\
\begin{pmatrix} . & x & x \\ x & x & x \\ x & x & X \end{pmatrix}
\end{array}
$$

$$
\left.\begin{array}{rl}
X: & \text{maximized} \\
x: & \text{minimized} \\
.: & \text{unchanged}
\end{array}\right\} \text{by the last Givens rotation} \quad \text{[COM89]}
$$

Quasi-algebraic algorithms

Influence of ordering

With update based on multilinearity.



Interpretation in terms of pairwise independence

- Pairs are processed in turns, so as to make outputs as independent as possible
- Ultimately: a set of *pairwise independent* outputs
- Legitimate because of corollary of Darmois's theorem (cf., slide 19)

Quasi-algebraic algorithms

Interpretation in terms of tensor diagonalization

Explanation for order 3 tensors

- Given a tensor g_{ijk}, find a matrix $\boldsymbol{Q}$ transforming g into $G_{pqr} = \sum_{ijk} Q_{pi} Q_{qj} Q_{rk}\, g_{ijk}$ such as to maximize:

$$\Psi_3(\boldsymbol{Q}) \stackrel{\text{def}}{=} \sum_i |G_{iii}|^2$$

- *Theorem:* if $\boldsymbol{Q}$ is unitary, then $\Omega \stackrel{\text{def}}{=} \sum_{ijk} |G_{ijk}|^2$ is constant independent of $\boldsymbol{Q}$

  Proof: uses $\sum_p Q_{ip} Q_{jp} = \delta_{ij}$

- *Corollary:* Maximize $\Upsilon_{3,2} \Leftrightarrow$ minimize all non diagonal entries

Hence: Approximate *"Tensor Diagonalization"*

Quasi-algebraic algorithms

Tensor diagonalization

Warning: Tensors cannot in general be diagonalized by congruent transforms, even non unitary!

Why?

because they have too many degrees of freedom ...

Stationary points

Example of diagonalization of real symmetric matrices

- Given a matrix g with components g_{ij}, it is sought for an orthogonal matrix Q such that ψ_2 is maximized:

$$\psi_2(G) = \sum_i G_{ii}^2; \quad G_{ij} = \sum_{p,q} Q_{ip} Q_{jq}\, g_{pq}.$$

- Stationary points of ψ_2 satisfy for any pair of indices $(q, r), q \neq r$:

$$G_{qq} G_{qr} = G_{rr} G_{qr}$$

- Next, $d^2\psi_2 < 0 \Leftrightarrow G_{qr}^2 < (G_{qq} - G_{rr})^2$, which proves that
  - $G_{qr} = 0, \forall q \neq r$ yields a maximum
  - $G_{qq} = G_{rr}, \forall q, r$ yields a minimum
  - Other stationary points are saddle points

Quasi-algebraic algorithms

Stationary points

Procedure applied to real 3rd or 4th order tensors

- Similarly, one can look at relations characterizing local maxima of criteria Ψ_3 and Ψ_4 [COM94b]:

$$\begin{aligned} G_{qqq}G_{qqr} - G_{rrr}G_{qrr} &= 0, \\ 4G_{qqr}^2 + 4G_{qrr}^2 - (G_{qqq} - G_{qrr})^2 - (G_{rrr} - G_{qqr})^2 &< 0; \\ G_{qqqq}G_{qqqr} - G_{rrrr}G_{qrrr} &= 0, \\ 4G_{qqqr}^2 + 4G_{qrrr}^2 - (G_{qqqq} - \frac{3}{2}G_{qqrr})^2 & \\ -(G_{rrrr} - \frac{3}{2}G_{qqrr})^2 &< 0. \end{aligned}$$

for any pair of indices $(p, q), p \neq q$. As a conclusion, contrary to Ψ_2 in the matrix case, Ψ_r might have theoretically spurious local maxima in the tensor case, $r > 2$

Quasi-algebraic algorithms

Tensors as Linear Operators

Overview

- Linear Operator Ω acting on square matrices:

$$\boldsymbol{M} \longrightarrow \Omega(\boldsymbol{M})_{ij} = \sum_{k\ell} \mathcal{C}_{ik}^{j\ell} M_{k\ell}$$

  admits eigen-matrices $\boldsymbol{N}(p)$, $1 \leq p \leq P^2$.

- In the absence of noise, P nonzero eigenvalues

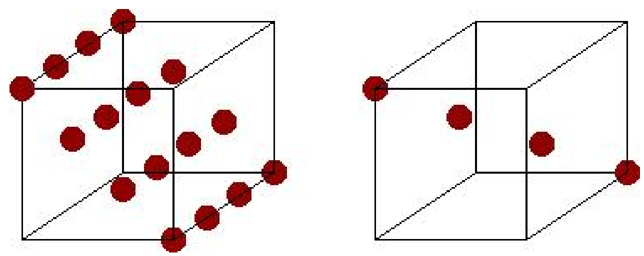
- In practice, retain P dominant eigen-matrices $\Rightarrow$ (i) reduced complexity P^2, and (ii) noise reduction

Quasi-algebraic algorithms

Joint Approximate Diagonalization (JAD)

Other idea: jointly diagonalize matrix slices

Example of $4 \times 4 \times 4$ tensors



Matrix slices diagonalization $\neq$ Tensor diagonalization

Performs less well, but computationnally attractive [CS93]

Quasi-algebraic algorithms

STD (1)

One step forward: Slicing decreases the order

- Similarly, one can try to diagonalize a 4th order tensor $\boldsymbol{T} = [\gamma_{ijk\ell}]$ by jointly diagonalizing 3rd order slices $\boldsymbol{T}(\ell)$
- Algorithm: Each Givens rotation is obtained again by maximizing a quadratic form $\boldsymbol{u}^{\mathsf{T}}\boldsymbol{B}\,\boldsymbol{u}$
- Noise reduction possibility: replace slices by a family of 3rd order tensors forming a basis of the map $\mathbb{C}^K \longrightarrow \mathbb{C}^{K\times K\times K}$ (consider the 4th order tensor as a linear map; basis obtained by SVD)

Quasi-algebraic algorithms

STD (2)

In the real case, $\boldsymbol{B}$ is given as in slide 30 by:

$$\boldsymbol{B} = \begin{pmatrix} a_1 & 3\,a_4/2 \\ 3\,a_4/2 & 9\,a_2/4 + 3\,a_3/2 + a_1/4 \end{pmatrix}$$

with [dLdMV01]:

$$\begin{aligned} a_1 &= \sum_\ell \gamma_{111\ell}^2 + \gamma_{222\ell}^2 \\ a_2 &= \sum_\ell \gamma_{112\ell}^2 + \gamma_{122\ell}^2 \\ a_3 &= \sum_\ell \gamma_{111\ell}\,\gamma_{122\ell} + \gamma_{112\ell}\,\gamma_{222\ell} \\ a_4 &= \sum_\ell \gamma_{122\ell}\,\gamma_{222\ell} - \gamma_{111\ell}\,\gamma_{112\ell} \end{aligned}$$

Criteria

Comparison between CoM, JAD, and STD

$$\Upsilon_{CoM2}(\boldsymbol{Q}) = \sum_{i=1}^{P} |T_{iiii}|^2 = \Upsilon_{2,4}, \qquad (9)$$

$$\Upsilon_{STD}(\boldsymbol{Q}) = \sum_{i=1}^{P} \sum_{j=1}^{P} |T_{iiij}|^2, \qquad (10)$$

$$\Upsilon_{JAD}(\boldsymbol{Q}) = \sum_{i=1}^{P} \sum_{j=1}^{P} \sum_{k=1}^{P} |T_{iijk}|^2 \qquad (11)$$

Different Discrimination powers:

$$\Upsilon_{CoM2}(\boldsymbol{Q}) \leq \Upsilon_{STD}(\boldsymbol{Q}) \leq \Upsilon_{JAD}(\boldsymbol{Q})$$

i.e. CoM2 is the best (but may be computationnally heavy, e.g. in $\mathbb{C}$)

The End

Conclusion

- ICA is widely used, and related to approximate tensor diagonalization
- But still lack of efficient numerical algorithms

References

[COM89] P. COMON. Separation of stochastic processes. In *Proc. Workshop on Higher-Order Spectral Analysis*, pages 174–179, Vail, Colorado, June 28-30 1989. IEEE-ONR-NSF.

[COM94a] P. COMON. Independent Component Analysis, a new concept ? *Signal Processing, Elsevier*, 36(3):287–314, April 1994. Special issue on Higher-Order Statistics.

[COM94b] P. COMON. Tensor diagonalization, a useful tool in signal processing. In M. Blanke and T. Soderstrom, editors, *IFAC-SYSID, 10th IFAC Symposium on System Identification*, volume 1, pages 77–82, Copenhagen, Denmark, July 4-6 1994. invited session.

[COM01] P. COMON. From source separation to blind equalization, contrast-based approaches. In *Int. Conf. on Image and Signal Processing (ICISP'01)*, Agadir, Morocco, May 3-5, 2001. invited plenary.

[CS93] J. F. CARDOSO and A. SOULOUMIAC. Blind beamforming for non-Gaussian signals. *IEE Proceedings - Part F*, 140(6):362–370, December 1993. Special issue on Applications of High-Order Statistics.

[dL97] L. de LATHAUWER. *Signal Processing based on Multilinear Algebra*. Doctorate, Katholieke Universiteit Leuven, September 1997.

[dLdMV01] L. de LATHAUWER, B. de MOOR, and J. VANDEWALLE. Independent Component Analysis and (simultaneous) third-order tensor diagonalization. *IEEE Trans. Sig. Proc.*, pages 2262–2271, October 2001.

[DON81] D. DONOHO. On minimum entropy deconvolution. In *Applied time-series analysis II*, pages 565–609. Academic Press, 1981.

[LAC97] J. L. LACOUME, P. O. AMBLARD, and P. COMON. *Statistiques d'ordre supérieur pour le traitement du signal*. Collection Sciences de l'Ingénieur. Masson, 1997.

[MP97] E. MOREAU and J. C. PESQUET. Generalized contrasts for multichannel blind deconvolution of linear systems. *IEEE Signal Processing Letters*, 4(6):182–183, June 1997.

[SBG04] A. SMILDE, R. BRO, and P. GELADI. *Multi-Way Analysis*. Wiley, 2004.