Blind Techniques

Pierre Comon

I3S, CNRS, Univ. of Nice, Sophia-Antipolis, France

Contents

Part I

Principle & Tools

Introduction

- Modeling
- General concepts, a 2×2 example
- Historical survey, Origins

Observation model

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

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

Taxonomy (1)

Static/Dynamic and **Noisy/Noiseless**:

$$\mathbf{x}[n] = \mathbf{H} \star \mathbf{s}[n] + \mathbf{v}[n] \qquad (2)$$

Over/Under-Determined:

$$\textit{Number of sources}: \; P \lesseqgtr^{\text{Underdet}} K \; : \textit{Number of sensors}$$

Taxonomy (2)

Transmit/Receive diversity:

Sources	Sensors	
	1	$K > 1$
1	**SISO**	**SIMO**
$P > 1$	**MISO**	**MIMO**

Taxonomy (3)

One additional assumption required on sources:

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

Principal Component Analysis (PCA)

Goal

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

- Observation
$$\mathbf{x} = \mathbf{U}\,\mathbf{z}$$
- $\mathbf{z}$ has uncorrelated components z_i

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

Independent Component Analysis (ICA)

Goal

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

- Observation
$$\mathbf{x} = \mathbf{H}\,\mathbf{s} \tag{3}$$
- $\mathbf{s}$ has mutually statistically independent components s_i

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

Uniqueness

Inherent indeterminations
if $\mathbf{s}$ has independent components s_i, so has $\mathbf{\Lambda P\, s}$
where $\mathbf{\Lambda}$ is invertible diagonal and $\mathbf{P}$ permutation

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

- *"Essential uniqueness"*: unique up to a *trivial filter*, i.e. a scale-permutation (cf. slide 67)
- Whole equivalence class of solutions $\Rightarrow$ Look for one representative.

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*:

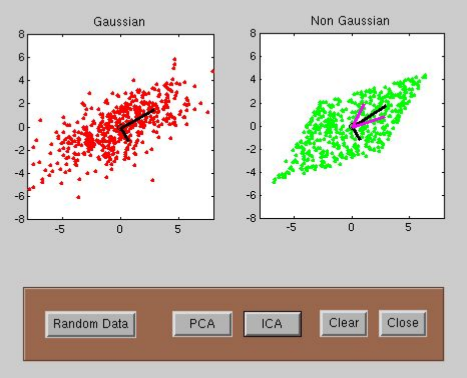
$$\mathrm{E}\{x_1\, x_2\} = \mathrm{E}\{s_1^2\} - \mathrm{E}\{s_2^2\} = 0$$

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

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

PCA vs ICA

Example 2: 2 sources and 2 sensors



Historical survey: Static MIMO

- **The ancestors:** Dugué'51, Darmois'53, Feller'66, Friedman'74, Donoho'80
- **The first shy steps in ICA:** Bar-Ness'82, Jutten'83, Fety'88
- **The first steps in Multi-Way:** Carroll-Chang'70, Harshman'70, Kruskal'77
- **First closed-form solutions:** Comon'89, Cardoso'92
- **First IT frameworks:** Comon'91, Cardoso'93, Comon'94, Bell'95, Delfosse-Loubaton'95
- **Specific applications:** Hyvarinen'97, Pajunen'97, Amari'98, Grellier'98, Parra'2000
- **Discrete/CM:** Talwar'96, VanderVeen'97Grellier'00

Historical survey: Static MIMO (cont'd)

- **Other:** Cao-Liu'96, VanDerVeen-Paulraj'96, Moreau-Pesquet'97, Taleb-Jutten'97, Comon'96, Ferreol-Chevalier'98, Belouchrani'98, Lee-Lewicki'99, deLathauwer'00, Pham-Cardoso'2000, Yeredor'2000, Sidiropoulos-Bro'00, Albera'04, Comon-Rajih'05, deLathauwer'05...

Historical survey: Convolutive SISO

- **Identification**
  - **Kurtosis** Benveniste-Ruget'80, Tugnait'89
  - **Non circularity/Alphabet:** Yellin-Porat'93, Grellier-Comon'99, Ciblat-Loubaton'02, Lebrun-Comon'03
- **Equalization**
  - **CMA:** Sato'75, Godard'80, Treichler'85
  - **Kurtosis:** Benveniste-Goursat'84, Donoho'81, Shalvi-Weinstein'90
  - **Bispectrum:** Marron'90, Matsuoka'84, LeRoux'93

NB: Earlier equalization algorithms, e.g. Decision-Directed, need the eye to be open.

Historical survey: Convolutive SIMO

- **Subspace:** Slock'94, Xu-Tong'95, Moulines-Duhamel'95, Xu-Liu-Tong'95, Gurelli-Nikias'95, Gesbert-Duhamel'97
- **Linear Prediction:** AbedMeraim-Moulines-Loubaton'97, Gesbert-Duhamel'00

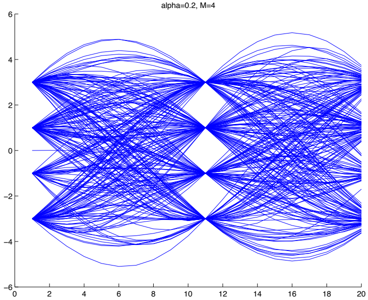
Historical survey: Convolutive MIMO

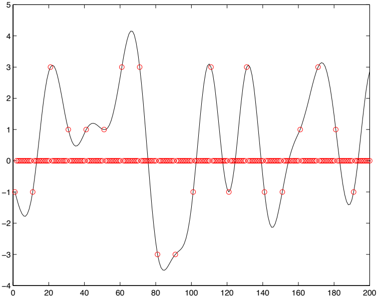
- **Subspace:** Gorokhov-Loubaton'97, Chevreuil-Loubaton'97, Loubaton-Moulines'01
- **Linear Prediction:** Comon'90, Ding'96, AbedMeraim-Loubaton'97, Gorokhov-Loubaton'99
- **Kurtosis:** Comon'96, Tugnait'97, Simon-Loubaton'98, Touzni'98
- **Discrete/CM:** Touzni-Fijalkow'98, VanDerVeen-Talwar'95, Ayadi-Slock'98

Origins of “Blind” Techniques

Pulse Amplitude Modulation (PAM) in baseband:

$$x(t) = a \sum_k h(t - k\,T)\, u_k$$





PAM4: symbols $u_k \in \{-3, -1, 1, 3\}$

General bibliography

- **Books on HOS, ICA, or Multi-Way:**
  Lacoume-Amblard-Comon'97 [LAC97] (freely downloadable, but in French)
  Hyvarinen-Karhunen-Oja'01 (but dedicated only to FastICA)
  Smilde-Bro-Geladi'04 [SBG04] (but dedicated only to Factor Analysis)
  Cichocki-Amari'02 [CA02] (but Neural Networks oriented)
  Comon-Jutten'06 [CJ07] [JC07] (but in French)
  Comon-Jutten'08 (will cover more topics, but you have to wait!)
- **Other related books:**
  Kagan-Linnik-Rao'73 [KLR73]
  McCullagh'87 [McC87]
  Nikias-Petropulu'93 [NP93]
  Haykin'2000 [HAY00a] [HAY00b]

Algebraic tools

- Singular Value decomposition (SVD)
- Spatial whitening (Standardization)
- PCA by pair sweeping
- Filter decomposition
- Time Whitening
- Space-time Whitening
- Matched filter

Singular Value Decomposition (SVD)

Every matrix $\mathbf{M}$ may be decomposed into:

$$\mathbf{M} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathsf{H}}$$

where

- $\mathbf{U}$ and $\mathbf{V}$ are unitary
- $\mathbf{\Sigma}$ is positive real diagonal
- $\mathbf{u}_i$ and $\mathbf{v}_i$ of $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors:

$$\mathbf{M}\,\mathbf{v}_i = \mathbf{u}_i\,\sigma_i \quad \mathbf{M}^{\mathsf{H}}\mathbf{u}_i = \mathbf{v}_i\,\sigma_i$$

- $\mathbf{u}_i$ are eigenvectors of $\mathbf{M}\mathbf{M}^{\mathsf{H}}$, and $\mathbf{v}_i$ those of $\mathbf{M}^{\mathsf{H}}\mathbf{M}$, associated with σ_i^2.

Spatial whitening (1)

Standardization via Cholesky or QR Let $\mathbf{x}$ be a zero-mean r.v. with covariance matrix:

$$\mathbf{\Gamma}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{x}\,\mathbf{x}^{\mathsf{H}}\}$$

Then Cholesky yields:

$$\exists \mathbf{L} \,/\quad \mathbf{L}\,\mathbf{L}^{\mathsf{H}} = \mathbf{\Gamma}_x$$

Consequence: $\mathbf{L}^{-1}\mathbf{x}$ has a unit variance.
Variable $\tilde{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{L}^{-1}\mathbf{x}$ is a *standardized random variable*.

- QR factorization of data matrix as $\mathbf{X} = \mathbf{L}\,\tilde{\mathbf{X}}$ yields same $\mathbf{L}$ as Cholesky factorization of sample covariance, but more accurate.
- Limitation: $\mathbf{L}$ may not be invertible if the covariance $\mathbf{\Gamma}_x$ is not full rank.

Spatial whitening (2)

Standardization via PCA

Definition

PCA is based on second order statistics

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

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

  where *Principal Components* z_i are uncorrelated
  ith column $\mathbf{u}_i$ of $\mathbf{U}$ is called *ith PC Loading vector*
- Two possible calculations:
  - EVD of Covariance $\mathbf{R}_x$: $\mathbf{R}_x = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^{\mathsf{H}}$
  - Sample estimate by SVD: $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}}$

Spatial whitening (3)

Summary

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

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

Remarks

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

Plane rotations

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

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

Same result obtained:

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

Jacobi sweeping for PCA

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

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

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

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

Scalar Filter Decomposition

- Any rational scalar filter $g[z]$ can be decomposed into:

$$\gamma[z] = u[z]\,\ell[z], \quad u[1/z^*]\,u[z] = 1,\ \forall z \tag{4}$$

- $\ell[z]$ is *minimum phase*: all its roots inside the unit circle
- $u[z]$ is *all-pass*, and hence *lossless*: flat frequency response (only phase varies).

Multivariate Filter Decomposition

- Any rational filter with Impulse Response matrix $\mathbf{F}[k]$ and complex gain $\check{\mathbf{F}}[z]$, can be decomposed into:

$$\check{\mathbf{F}}[z] = \check{\mathbf{U}}[z]\,\check{\mathbf{L}}[z], \quad \check{\mathbf{U}}[1/z^*]^{\mathsf{H}}\check{\mathbf{U}}[z] = \mathbf{I},\ \forall z \tag{5}$$

- $\mathbf{L}[k]$ is *triangular minimum phase* filter: roots of $\det(\check{\mathbf{L}}[z])$ inside unit circle
- $\mathbf{U}[k]$ *para-unitary* filter
- In static MIMO case, one gets QR:

$$\mathbf{F} = \mathbf{U}\,\mathbf{L}, \quad \mathbf{U}^{\mathsf{H}}\mathbf{U} = \mathbf{I} \tag{6}$$

where $\mathbf{L}$ is triangular and $\mathbf{U}$ unitary.

- Decomposition not unique.

Time Whitening

Let $x[k]$ be a scalar second order stationary process, $\check{x}[z]$ its $z-$ transform, and its power spectrum given by:

$$\gamma_x[z] \stackrel{\text{def}}{=} \mathrm{E}\{\check{x}[z]\,\check{x}[1/z^*]^*\}$$

From (4), the power spectrum can be decomposed as:

$$\exists \ell[z] \,/ \quad \ell[z]\,\ell[1/z^*]^* = \gamma_x[z]$$

where filter $\ell[z]$ is not unique, and defined up to an all-pass filter. $1/\ell[z]$ is a *whitening filter*, if it exists.

Space-time Whitening

- Let $x[k]$ be a multivariate second order stationary random process, $x[z]$ its $z-$ transform, and power spectral matrix:

$$\mathbf{\Gamma}_x[z] \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{x}[z]\,\mathbf{x}[1/z^*]^{\mathsf{H}}\}$$

Then, from (5)

$$\exists \check{\mathbf{L}}[z] \,/\quad \check{\mathbf{L}}[z]\,\check{\mathbf{L}}[1/z^*]^{\mathsf{H}} = \mathbf{\Gamma}_x[z]$$

- If $\check{\mathbf{L}}[z]$ admits an inverse, then we may take $\check{\mathbf{G}}[z] = \check{\mathbf{L}}[z]^{-1}$ as *whitening filter*, i.e. $\tilde{\mathbf{x}}[k] = \mathbf{G}[k] \star \mathbf{x}[k]$.

Spatial Matched Filter

If $\mathbf{x} = \mathbf{H}\,\mathbf{s} + \mathbf{v}$, where $\mathbf{H}$ is known, one can estimate $\mathbf{s}$ by spatial filtering as

$$\hat{\mathbf{s}} = \mathbf{W}\,\mathbf{x}$$

- Spatial Matched Filter: $\mathbf{W} = \mathbf{H}^{\mathsf{H}}\mathbf{R}_x^{-1}$
- Least Squares: $\mathbf{W} = (\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}\mathbf{H}^{\mathsf{H}}$
- Weighted Least Squares: $\mathbf{W} = (\mathbf{H}^{\mathsf{H}}\mathbf{B}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathsf{H}}\mathbf{B}^{-1}$ where $\mathbf{B}$ is the noise spatial coherence.

Statistical Tools

- Statistical Independence
- Mutual vs Pairwise Independence
- Cumulants of a scalar r.v.
- Multivariate Cumulants
- Complex variables, circularity
- Central limit, Edgeworth expansion
- Mutual Information, approximation

Statistical Independence

Definition

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

$$\Updownarrow$$

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

$$p_{\mathbf{s}}(\mathbf{u}) = \prod_k p_{s_k}(u_k) \tag{7}$$

Definition

Components s_k of $\mathbf{s}$ 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

- Bag containing 4 Bowls denoted {RB, YB, GB, RYB}:
  1 Red, 1 Yellow, 1 Green, 1 with the 3 colors.
- Equal drawing probabilities:
  $P(RB) = P(YB) = P(GB) = P(RYG) = 1/4$
- Event "R" $\stackrel{\text{def}}{=}$ draw a bowl containing Red $\Rightarrow$
  $P(R) = P(RB) + P(RYG) = 1/2$
- Then $P(R \cap Y) = P(RYG) = 1/4$
  equal to $P(R) * P(Y) \Rightarrow$ *Pairwise independent* Events
- But $P(R \cap Y \cap G) = P(RYG) = 1/4$
  not equal to $P(R) * P(Y) * P(G) = 1/8 \Rightarrow$
  Events are *not Mutually independent*

Mutual vs Pairwise independence (2)

Example 4: 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, $\mathrm{Cum}\{x_1, x_2, x_3, x_4\} = 1 \neq 0$

Mutual vs Pairwise independence (3)

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 (4)

Corollary

If $\mathbf{z} = \mathbf{C}\,\mathbf{s}$, 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. $\mathbf{C} = \mathbf{\Lambda}\,\mathbf{P}$, with $\mathbf{\Lambda}$ diagonal and $\mathbf{P}$ permutation

Characteristic functions

First c.f.

- Real Scalar: $\Phi_x(t) \stackrel{\text{def}}{=} \mathrm{E}\{e^{\jmath\, tx}\} = \int_u e^{\jmath\, tu}\, dF_x(u)$
- Real Multivariate: $\Phi_{\mathbf{x}}(\mathbf{t}) \stackrel{\text{def}}{=} \mathrm{E}\{e^{\jmath\, \mathbf{t}^{\mathsf{T}}\mathbf{x}}\} = \int_{\mathbf{u}} e^{\jmath\, \mathbf{t}^{\mathsf{T}}\mathbf{x}}\, dF_{\mathbf{x}}(\mathbf{u})$

Second c.f.

- $\Psi(\mathbf{t}) \stackrel{\text{def}}{=} \log \Phi(\mathbf{t})$
- Properties:
  - Always exists in the neighborhood of 0
  - Uniquely defined as long as $\Phi(\mathbf{t}) \neq 0$

Definition of Cumulants

- Moments:

$$\mu'_r \stackrel{\text{def}}{=} \mathrm{E}\{x^r\} = (-\jmath)^r \left. \frac{\partial^r \Phi(t)}{\partial t^r} \right|_{t=0} \tag{8}$$

- Cumulants:

$$\mathcal{C}_{x\,(r)} \stackrel{\text{def}}{=} \mathrm{Cum}\{\underbrace{x, \ldots, x}_{r\ \ times}\} = (-\jmath)^r \left. \frac{\partial^r \Psi(t)}{\partial t^r} \right|_{t=0} \tag{9}$$

- Needs the existence of the expansion. Counter example: Cauchy

$$p_x(u) = \frac{1}{\pi\,(1+u^2)}$$

- Relationship between Moments and Cumulants obtained by expanding both sides in Taylor series:

$$\mathrm{Log}\,\Phi_x(t) = \Psi_x(t)$$

First Cumulants

- $\mathcal{C}_{(2)}$ is the variance:
- For zero-mean r.v.: $\mathcal{C}_{(3)} = \mu_{(3)}$, and $\mathcal{C}_{(4)} = \mu_{(4)} - 3\,\mu_{(2)}^2$
- Warning: it is not true that $\mathcal{C}_{(r)}$ is the moment of a variable $x - x_g$, x_g Gaussian
- Standardized cumulants:

$$\mathcal{K}_{(r)} = \mathrm{Cum}_{(r)}\left\{\frac{x - \mu'_{(1)}}{\sqrt{\mu_{(2)}}}\right\}$$

e.g. *Skewness* $\mathcal{K}_3$, and *Kurtosis* $\mathcal{K}_4$.

Examples of cumulants (1)

Example 5: Zero-mean Gaussian

- Moments

$$\mu_{(2r)} = \mu_{(2)}^r \frac{(2r)!}{r!\,2^r}$$

  In particular:

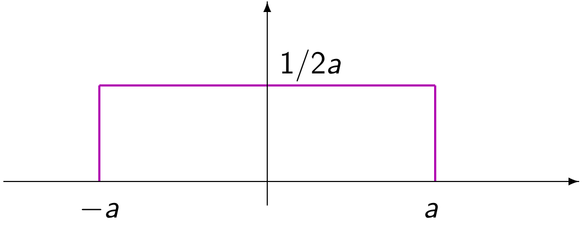
$$\mu_{(4)} = 3\mu_{(2)}^2, \;\; \mu_{(6)} = 15\mu_{(2)}^3$$

- $\mathcal{C}_{(4)} = 0, \quad \mathcal{K}_{(4)} = 0.$
- All Cumulants of order $r > 2$ are null

Examples of Cumulants (2)

Example 6: Uniform

- uniformly distributed in $[-a, +a]$ with probability $\frac{1}{2a}$
- Moments: $\mu_{(2k)} = \frac{a^{2k}}{2k+1}$
- 4th order Cumulant: $\mathcal{C}_{(4)} = \frac{a^4}{5} - 3\,\frac{a^4}{9} = -2\,\frac{a^4}{15}$
- Kurtosis: $\mathcal{K}_{(4)} = -\frac{6}{5}$.

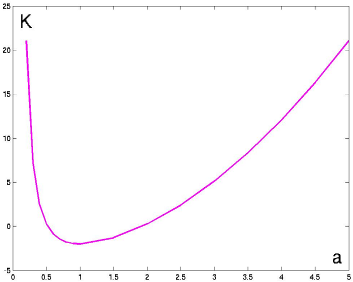


Proof...

Examples of Cumulants (3)

Example 7: Zero-mean standardized binary

- x takes two values $x_1 = -a$ and $x_2 = 1/a$ with probabilities $P_1 = \frac{1}{1+a^2}$, $P_2 = \frac{a^2}{1+a^2}$
- Skewness is $\mathcal{K}_{(3)} = \dfrac{1}{a} - a$
- Kurtosis is $\mathcal{K}_{(4)} = \dfrac{1}{a^2} + a^2$
- Extreme values



Minimum Kurtosis
for $a = 1$ (symmetric):
$\mathcal{K}_{(4)} = -2$

Sub- and Super-Gaussian r.v.

Warning:
The concept of Sub/Super Gaussian is not uniquely defined in the literature. For instance, definitions below are *not equivalent*:

- Monotonicity of [BGR80]: $f(u) = -\frac{1}{u}\,\frac{d\,\log p_x(u)}{du}$.
- Tails of the standardized pdf are below/above those of Gaussian [ZIV95]
- Bsed on kurtosis [KS77]:
  - *Leptokurtic:* positive kurtosis
  - *mesokurtic:* zero kurtosis
  - *platykurtic:* negative kurtosis

Definition of Multivariate cumulants

- Notation: $\mathcal{C}_{ij..\ell} \stackrel{\text{def}}{=} \mathrm{Cum}\{X_i, X_j, ...X_\ell\}$
- First cumulants:

$$\begin{aligned} \mu'_i &= \mathcal{C}_i \\ \mu'_{ij} &= \mathcal{C}_{ij} + \mathcal{C}_i\mathcal{C}_j \\ \mu'_{ijk} &= \mathcal{C}_{ijk} + [3]\,\mathcal{C}_i\mathcal{C}_{jk} + \mathcal{C}_i\mathcal{C}_j\mathcal{C}_k \end{aligned}$$

with $[n]$: Mccullagh's *bracket notation*.

- Next, for zero-mean variables:

$$\begin{aligned} \mu_{ijk\ell} &= \mathcal{C}_{ijk\ell} + [3]\,\mathcal{C}_{ij}\mathcal{C}_{k\ell} \\ \mu_{ijk\ell m} &= \mathcal{C}_{ijk\ell m} + [10]\,\mathcal{C}_{ij}\mathcal{C}_{k\ell m} \end{aligned}$$

- General formula of Leonov Shiryayev obtained by Taylor expansion of both sides of $\Psi(\mathbf{t}) = \log \Phi(\mathbf{t})$...

Arrays and Tensors

Definitions Table $\mathbf{T} = \{T_{ij..k}\}$

- *Order* of $\mathbf{T} \stackrel{\text{def}}{=} \#$ of its ways $= \#$ of its indices
- *Dimension* $K_\ell \stackrel{\text{def}}{=}$ range of the ℓth index
- $\mathbf{T}$ is *Cubic* when all dimensions $K_\ell = K$ are equal
- $\mathbf{T}$ is *Symmetric* when it is square and when its entries do not change by *any* permutation of indices

NB: cf. course III for definitions and properties

Definition of Complex Cumulants

Definition

Let $\mathbf{z} = \mathbf{x} + \jmath\,\mathbf{y}$. Then pdf $p_{\mathbf{z}}$ = joint pdf $p_{\mathbf{x},\mathbf{y}}$

Notation

- Characteristic function:

$$\Phi_{\mathbf{z}}(\mathbf{w}) = \mathrm{E}\{\exp[\jmath(\mathbf{x}^{\mathsf{T}}\mathbf{u} + \mathbf{y}^{\mathsf{T}}\mathbf{v})]\} = \mathrm{E}\{\exp[\jmath\Re(\mathbf{z}^{\mathsf{H}}\mathbf{w})]\}$$

  where $\mathbf{w} \stackrel{\mathrm{def}}{=} \mathbf{u} + \jmath\mathbf{v}$.

- Generates Moments & Cumulants, e.g.:
  Variance: $Var\{\mathbf{z}\}_{ij} = \mathcal{C}_{\mathbf{z}\,i}^{\ \,j}$
  Higher orders: $\mathrm{Cum}\{z_i, \dots, z_j, z_k^*, \dots, z_\ell^*\} = \mathcal{C}_{\mathbf{z}\,ij}^{\ \,k\ell}$
  where *conjugated* r.v. are labeled *in superscript*.

Circularity (1)

- z is *circular in the strict sense* if its distribution does not depend on the phase of z. For a multivariate complex random variable $\mathbf{z}$, it means that:

$$\mathbf{z} \text{ and } \mathbf{z}\, e^{\jmath\theta},\ \forall\theta \in \mathbb{R}$$

have the same joint distribution.

- **Example 8: scalar circular complex Gaussian r.v.**

$$p_z(w) = \frac{1}{\pi\,\sigma^2}\,\exp -\frac{|w|^2}{\sigma^2}$$

defines a circular r.v.: only modulus appears.

Circularity (2)

- There exist up to 2^r distinct definitions of complex multivariate cumulants.
- At even order $2r$, cumulants having exacting r complex conjugations are termed *circular cumulants*.
- For instance, the cumulant below is circular

$$\mathcal{C}_{\mathbf{z}\, ij}^{\;\; k\ell} = \mathrm{Cum}\{z_i, z_j, z_k^*, z_\ell^*\}$$

whereas these ones are non circular

$$\mathcal{C}_{\mathbf{z}\, ijl}^{\;\; \ell} = \mathrm{Cum}\{z_i, z_j, z_k, z_\ell^*\}$$

$$\mathcal{C}_{\mathbf{z}\, ijk\ell} = \mathrm{Cum}\{z_i, z_j, z_k, z_\ell\}$$

- $\mathbf{z}$ is said to be *circular at order r* if its non circular cumulants of order r are all null:

$$p \neq r - p \Rightarrow \mathrm{Cum}\{z_1, .., z_p, z_{p+1}^*, .., z_r^*\} = 0 \qquad (10)$$

Example of complex r.v.

Example 9: PSK random variables For a PSK−4 random variable, $ZZ^* = 1$ and consequently:

$$\mathcal{C}_{(2)} = \mathrm{E}\{Z^2\} = 0, \mathcal{C}_{(2)}^{(2)} = -1, \mu_{(4)}^{(0)} = 1, \mathcal{C}_{(4)}^{(0)} = 1$$

It is thus circular up to order 3, but *non circular at order 4*.

Properties of Cumulants

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

$$\mathrm{Cum}\{\alpha X,\ Y, .., Z\} = \alpha\,\mathrm{Cum}\{X,\ Y, .., Z\} \quad (11)$$
$$\mathrm{Cum}\{X_1 + X_2,\ Y, .., Z\} = \mathrm{Cum}\{X_1,\ Y, .., Z\} + \mathrm{Cum}\{X_2,\ Y, .., Z\}$$

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

$$\mathrm{Cum}\{X_1,\ X_2, .., X_r\} = 0 \quad (12)$$

- **Additivity**: If **X** and **Y** are independent, then

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

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

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

Proof...

Central Limit Theorem

Let N independent scalar r.v., $x(n), 1 \leq n \leq N$ each with finite rth order Cumulant, $\kappa_{(r)}(n)$.
Define:

$$\bar{\kappa}_{(r)} = \frac{1}{N}\sum_{n=1}^{N}\kappa_{(r)}(n) \text{ and } y = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}(x(n) - \bar{\kappa}_{(1)}).$$

As $N \to \infty$, the pdf f_y tends to a Gaussian.

Proof.
Thanks to multi-linearity and additivity, $\mathcal{C}_{y\,(r)} = \frac{\bar{\kappa}_{(r)}}{N^{r/2-1}}, \forall r \geq 2$, tends to zero.

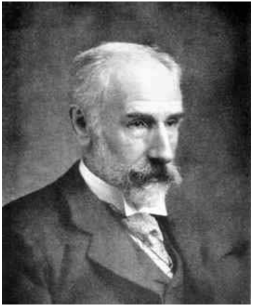
Edgeworth expansion of a pdf (1)

The pdf $p_{\mathbf{x}}(\mathbf{u})$ of a r.v. $\mathbf{x}$ can be expanded about the Gaussian density $g_{\mathbf{x}}(\mathbf{u})$ of same mean and variance, in terms of a combination of Hermite polynomials, ordered by decreasing significance in the sense of the Central Limit Theorem (CLT).

Order							
$m^{-1/2}$	κ_3						
m^{-1}	κ_4	κ_3^2					
$m^{-3/2}$	κ_5	$\kappa_3\kappa_4$	κ_3^3				
m^{-2}	κ_6	$\kappa_3\kappa_5$	$\kappa_3^2\kappa_4$	κ_4^2	κ_3^4		
$m^{-5/2}$	κ_7	$\kappa_3\kappa_6$	$\kappa_3^2\kappa_5$	$\kappa_4^2\kappa_3$	κ_3^5	$\kappa_4\kappa_5$	$\kappa_3^3\kappa_4$

From slide 53, rth order Cumulants $\sim O(m^{1-r/2})$.

Edgeworth expansion of a pdf (2)



Francis Edgeworth (1845-1926).

$$\frac{p_x(u)}{g_x(u)} = 1 + \frac{1}{3!}\kappa_3 h_3(v) + \frac{1}{4!}\kappa_4 h_4(v) + \frac{10}{6!}\kappa_3^2 h_6(v) + \frac{1}{5!}\kappa_5 h_5(v) + \frac{35}{7!}\kappa_3\kappa_4 h_7(v) + \frac{280}{9!}\kappa_3^3 h_9(v) + \dots$$

Mutual Information: definition

- According to the definition of page 34, one should measure a divergence:

$$\delta\left(p_{\mathbf{x}}, \prod_{i=1}^{N} p_{x_i}\right)$$

- If the *Kullback divergence* is used:

$$\mathrm{K}(p_{\mathbf{x}}, p_{\mathbf{y}}) \stackrel{\text{def}}{=} \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{p_{\mathbf{y}}(\mathbf{u})}\, d\mathbf{u},$$

then we get the *Mutual Information* as an independence measure:

$$I(p_{\mathbf{x}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{\prod_{i=1}^{N} p_{x_i}(u_i)}\, d\mathbf{u}. \qquad (13)$$

Mutual Information: properties

- MI always positive
- Cancels if r.v. are mutually independent
- MI is invariant by scale change

Proof...

- **Example 10: Gaussian case**

$$I(g_{\mathbf{x}}) = \frac{1}{2} \log \frac{\prod V_{ii}}{\det \mathbf{V}}$$

Mutual Information: decomposition

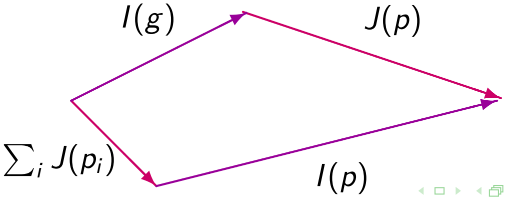
- Define the Negentropy as the divergence:

$$J(p_{\mathbf{x}}) = \mathrm{K}(p_{\mathbf{x}}, g_{\mathbf{x}}) = \int p_{\mathbf{x}}(\mathbf{u}) \log \frac{p_{\mathbf{x}}(\mathbf{u})}{g_{\mathbf{x}}(\mathbf{u})} \, d\mathbf{u}. \tag{14}$$

Negentropy is invariant by invertible transforms

- Then MI can be decomposed into:

$$I(p_{\mathbf{x}}) = I(g_{\mathbf{x}}) + J(p_{\mathbf{x}}) - \sum_i J(p_{x_i}). \tag{15}$$



Sample Measures of Statistical Independence

Independence at order r

- Definition:
  Components x_j of $\mathbf{x}$ are independent at order r if all *cross cumulants* of order r are null
- In other words: the *Cumulant tensor* $\mathcal{C}_{ij..\ell}$ is diagonal.

Example 11: Uncorrelated but not independent

$\mathbf{s}$ non Gaussian, s_i independent, then $\mathbf{x} = \mathbf{Q}\,\mathbf{s}$ has uncorrelated components *at order 2* if $\mathbf{Q}$ unitary $\rightarrow$ cf. example slide 12.

Edgeworth expansion of the MI

This yields for standardized random variables **x**, after lengthy calculations:

$$I(p_{\mathbf{x}}) = J(p_{\mathbf{x}}) - \frac{1}{48}\sum_i 4\,\mathcal{C}_{iii}^{\;2} + \mathcal{C}_{iiii}^{\;2} + 7\,\mathcal{C}_{iii}^{\;4} - 6\,\mathcal{C}_{iii}^{\;2}\,\mathcal{C}_{iiii} + o(m^{-2}). \tag{16}$$

- If 3rd order $\neq 0$, then $I(p_{\mathbf{x}}) \approx J(p_{\mathbf{x}}) - \frac{1}{12}\sum_i \mathcal{C}_{iii}^{\;2}$
- If 3rd order ≈ 0, then $I(p_{\mathbf{x}}) \approx J(p_{\mathbf{x}}) - \frac{1}{48}\sum_i \mathcal{C}_{iiii}^{\;2}$

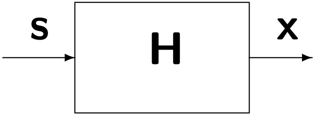
Optimization Criteria

- Cumulant matching
- Contrast criteria
- Mutual Information
- Maximum Likelihood vs MI
- CoM family
- Other criteria

Identification by Cumulant matching

Principle

- Estimate the mixture by solving the I/O Multi-linear equations
- Apply a separating filter based on the latter estimate



Noiseless mixture of 2 sources

Example 12: 2×2 **by Cumulant matching** (cf. demo p.13)

- After standardization, the mixture takes the form

$$\mathbf{x} = \begin{pmatrix} \cos\alpha & -\sin\alpha\, e^{\jmath\varphi} \\ \sin\alpha\, e^{-\jmath\varphi} & \cos\alpha \end{pmatrix} \mathbf{s} \qquad (17)$$

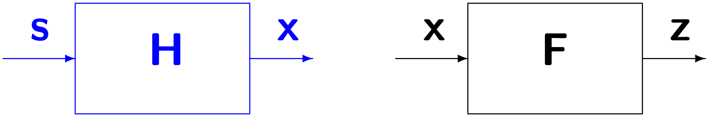
- Denote $\gamma_{ij}^{k\ell} = \mathrm{Cum}\{x_i, x_j, x_k^*, x_\ell^*\}$ and $\kappa_i = \mathrm{Cum}\{s_i, s_i, s_i^*, s_i^*\}$.
  Then by *Multi-linearity*:

$$\begin{aligned} \gamma_{12}^{12} &= \cos^2\alpha\, \sin^2\alpha\, (\kappa_1 + \kappa_2) \\ \gamma_{11}^{12} &= \cos^3\alpha\, \sin\alpha\, e^{\jmath\varphi}\, \kappa_1 - \cos\alpha\, \sin^3\alpha\, e^{\jmath\varphi}\, \kappa_2 \\ \gamma_{12}^{22} &= \cos\alpha\, \sin^3\alpha\, e^{\jmath\varphi}\, \kappa_1 - \cos^3\alpha\, \sin\alpha\, e^{\jmath\varphi}\, \kappa_2 \end{aligned}$$

- Compact solution: $\frac{\gamma_{12}^{22} - \gamma_{11}^{12}}{\gamma_{12}^{12}} = -2\, \cot 2\alpha\, e^{\jmath\varphi}$

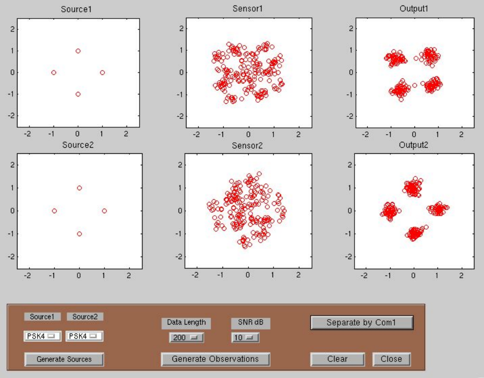
Now the inverse approach

- Cumulant matching (direct approach: identification)
- Contrast Criteria (inverse approach: equalization):



Noisy Mixtures of 2 sources

Example 13: Separation of 2 non Gaussian sources by contrast maximization



Source additional hypotheses

- **H1.** Each sources $s_j[k]$ is an i.i.d. sequence, for any fixed j

  H2. Sources s_j are mutually statistically independent

  H3. At most one source is Gaussian

  H4. At most one source has a null marginal cumulant

- **H5.** Sources are Discrete, and belong to some known alphabet (but may be stat. dependent)

  H6. Sources $s_j[k]$ are sufficiently exciting

- **H7.** Sources are colored, and the set of source spectra forms a family of linearly independent functions

- **H8.** Sources are non stationary, and have different time profiles

Trivial Filters

- They account for Inherent indeterminacies, remaining after assuming Source additional hypotheses
  For instance:
  - For dynamic (convolutive) mixtures, under **H1, H2, H3**, $\check{\mathbf{T}}[z] = \mathbf{P}\,\check{\mathbf{D}}[z]$, where $\mathbf{P}$ is a permutation, and $\check{\mathbf{D}}[z]$ a diagonal filter, with entries of the form $\check{D}_{pp}[z] = \lambda_p\, z^{\delta_p}$, where δ_p is an integer.
  - For static mixtures, under **H2, H3**, $\mathbf{T} = \mathbf{PD}$, where $\mathbf{P}$ permutation and $\mathbf{D}$ diagonal invertible.
- In other words, if $\mathbf{s}$ satisfies **Hi**, then so does $\mathbf{Ts}$

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 (as defined in slide 67)
- *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)

Mutual Information

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

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

Proof... cf slide 57

Likelihood

Given the source pdf's: $p_{\mathbf{s}}(\mathbf{u}) = \prod_i p_{s_i}(u_i)$, and a sample x_t, the ML approach consists of maximizing one of the criteria below w.r.t. $\mathbf{H}$:

- **Noiseless case**

$$p_{\mathbf{x}|\mathbf{H}}(\mathbf{x}_T|\mathbf{H}) = \frac{1}{|\det \mathbf{H}|}\, p_{\mathbf{s}}(\mathbf{H}^{-1}\mathbf{x})$$

- **Noisy case**

$$p_{\mathbf{x},\mathbf{s}|\mathbf{H}}(\mathbf{x}_T, \mathbf{s}|\mathbf{H}) = g(\mathbf{x}_T - \mathbf{H}\,\mathbf{s})\, p_{\mathbf{s}}(\mathbf{s})$$

- And the *Joint MAP-ML* criterion for a joint estimation of sources:

$$\begin{aligned}(\mathbf{s}_{MAP}, \mathbf{H}_{MV}) &= Arg\,\mathrm{Max}_{\mathbf{s},\mathbf{H}}\; p_{\mathbf{x},\mathbf{s}|\mathbf{H}}(\mathbf{x}_T, \mathbf{s}|\mathbf{H}) \\ &= Arg\,\mathrm{Max}_{\mathbf{s},\mathbf{H}}\; p(\mathbf{x}_T|\mathbf{s}, \mathbf{H})\, p_{\mathbf{s}}(\mathbf{s})\end{aligned}$$

Noiseless Maximum Likelihood (1)

- For an increasing number of independent observations, the average log-likelihood converges to

$$\mathcal{L}_T \stackrel{\text{def}}{=} \frac{1}{T} \log p(\mathbf{x}_1 \dots \mathbf{x}_T | \mathbf{H}) \longrightarrow \mathcal{L}_\infty = \int p_{\mathbf{x}}(\mathbf{u}) \log p_{\mathbf{x}|\mathbf{H}}(\mathbf{u}) \, d\mathbf{u}$$

which can be seen to be, by making the change $\mathbf{v} = \mathbf{H}^{-1}\mathbf{u}$, and up to a constant:

$$\Upsilon_{ML} \stackrel{\text{def}}{=} \mathcal{L}_\infty - S(p_{\mathbf{x}}) = -K(p_{\mathbf{z}}, p_{\mathbf{s}}) \tag{18}$$

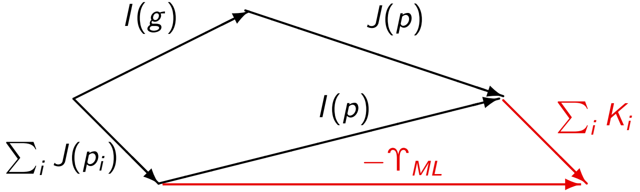
Proof...

Noiseless maximum Likelihood (2)

- Yet, since s_i are independent, it can be shown that

$$K(p_{\mathbf{z}}, p_{\mathbf{s}}) = \underbrace{K(p_{\mathbf{z}}, \prod_i p_{z_i})}_{MI} + \underbrace{\sum_i K(p_{z_i}, p_{s_i})}_{pdfdeviation}$$

This allows to take into account the source pdf's, if they are known



- **But** ML is not adequate if source pdf's are unknown
$\Rightarrow$ just use MI

CoM Family of contrast functionss

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

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

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

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

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

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

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

General Family of contrasts

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

$$\Upsilon_g(\mathbf{z}) = \sum_i g(|\kappa_{i\,(p)}^{(q)}|) \tag{19}$$

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

Proof...

Contrast CoM(1, 4)

Example 14: Kurtosis-based contrast without squaring

- In particular, if all source kurtosis have the same sign, ε, one can avoid the absolute value:

$$\Upsilon_{1,4} = \varepsilon \sum_{p=1}^{P} \kappa_{ii}^{ii}$$

Proof...

Other criteria

- Contrasts based on Matrix slices of Cumulant tensor
- Contrasts dedicated to Discrete source alphabets
- Contrasts for convolutive mixtures - basically the same!

Part II

Applications

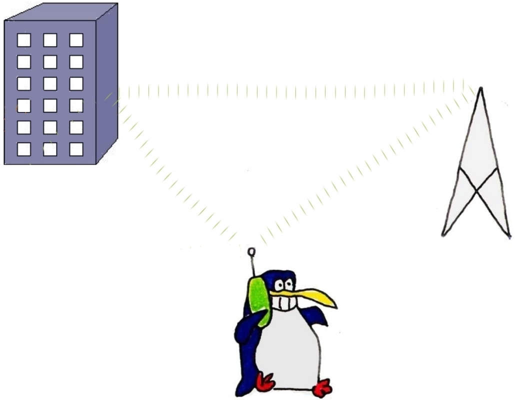
Applications

- Sensor Array Processing
- Telecommunications
- Speech
- Biomedical
- Machine Learning
- Exploratory Analysis...

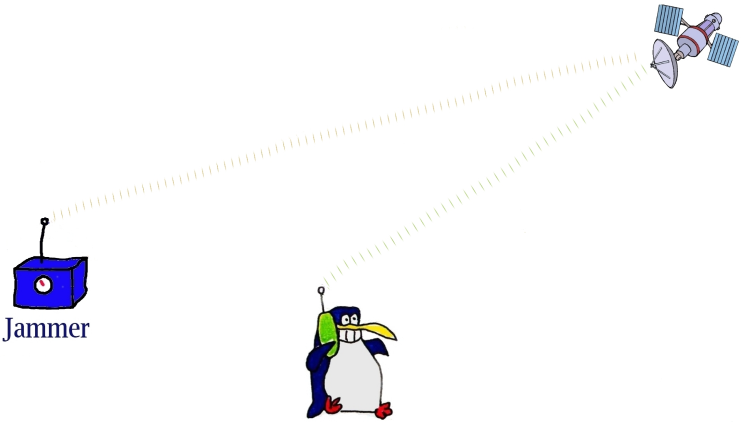
Application Areas (1)

- 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)

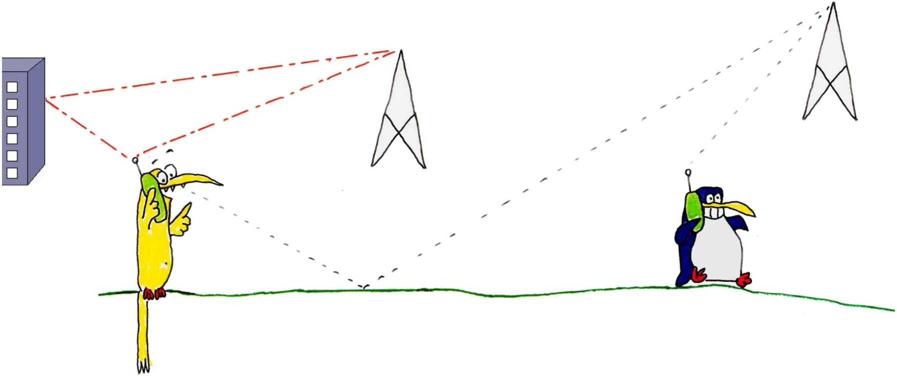
Telecommunications: SISO equalization



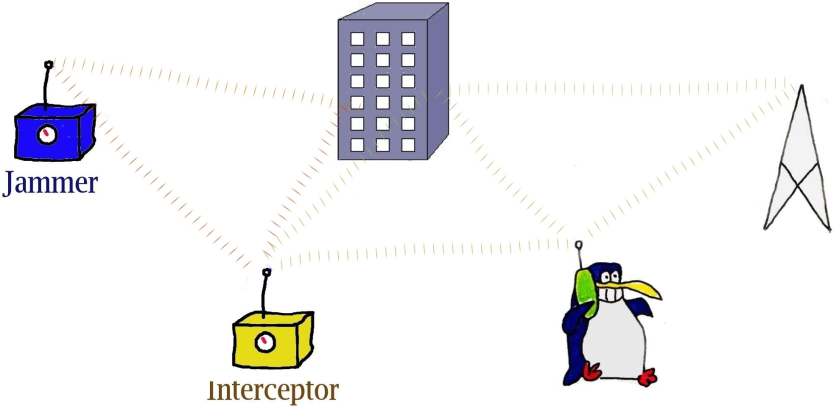
Telecommunications: MISO equalization



Telecommunications: MIMO equalization



ComInt: MIMO equalization



Speech

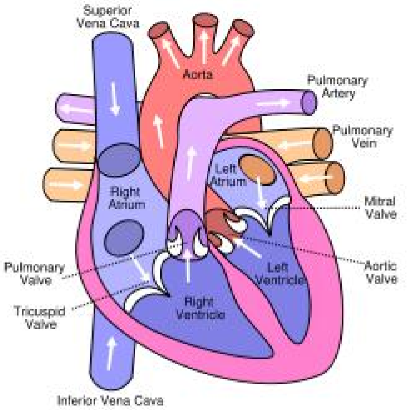
The Coktail Party problem

Deconvolution

Ill focussing is a 2-D convolution: mixture with neighboring pixels.

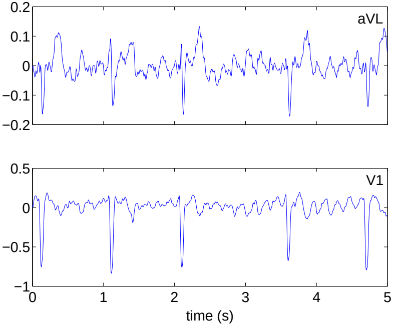
Electrocardiography (1)

Anatomy

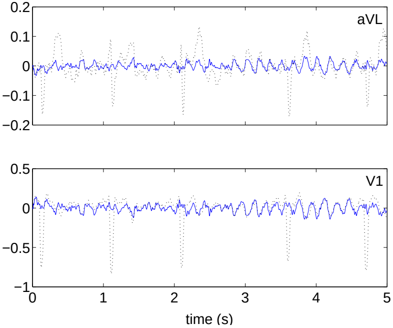


Electrocardiography (2)

Atrial fibrillation [RCS+04]



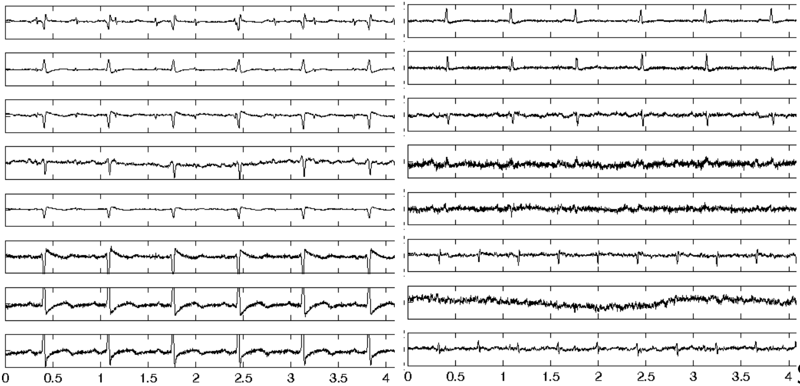
Atrial Fibrillation episode



Extracted Atrial activities

Electrocardiography (3)

Mother-Phoetus separation [dLdMV00a]



Data

Separation by ICA

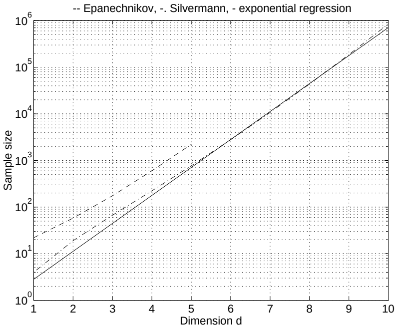
Application Areas (2)

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

Machine Learning

Curse of dimensionality

- Number of samples required to reach a given relative error in pd.f. estimate, $O(\epsilon)$, is of order $O(\epsilon^{-1-d/4})$ [SIL86] $\Rightarrow$ exponential in d



- Split of space into two lower dimensional subspaces allows the approximation of the p.d.f. [COM95]:
$p_x(\mathbf{u}) \approx p_{x_1}(\mathbf{u}_1) \cdot p_{x_2}(\mathbf{u}_2)$

Factor Analysis

Food Sciences:
one of the numerous application areas

judges $\times$ *products* $\times$ *sensory properties*

Part III

Tensors

Contents

Introduction

- Striking facts
- Order, dimensions, outer & inner products
- Contraction
- Multi-linearity property
- Unfoldings & storage
- Symmetry

Striking facts

1. The rank of a matrix cannot be larger than its dimensions $\rightarrow$ possible for a tensor
2. Matrices with entries drawn randomly have maximal rank $\rightarrow$ not true for a tensor
3. The set of matrices of rank at most r is closed, $\forall r$, $\rightarrow$ not true for a tensor. Hence the approximation problem is generally ill-posed.
4. Worse: the maximal achievable rank of a tensor is generally still unknown.
5. There are several ways to extend the SVD to tensors
6. The computation of the rank of a given tensor still raises unsolved difficulties.
7. Rank and symmetric rank have not yet been proved to be the same
8. Subtraction of best rank-1 approximate does not necessarily decrease the rank

Tensor product

- Let $\mathcal{V}_\ell$ be vector spaces of dimension K_ℓ on a field $\mathbb{K}$, and let $\mathbf{v}_\ell \in \mathcal{V}_\ell$ be K_ℓ-dimensional vectors.
- A *tensor* $\mathbf{T}$ is an element of a tensor product $\mathcal{V}_1 \otimes \mathcal{V}_2 \otimes \ldots \otimes \mathcal{V}_P$. For instance

$$\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \ldots \otimes \mathbf{v}_P$$

  is a tensor of *dimensions* $K_1 \times K_2 \times \cdots \times K_P$.

Arrays

- If coordinates of $\mathbf{u} \in \mathcal{U}$, $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{W}$ are u_i, v_j, and w_k in canonical bases of $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$ respectively, then coordinates of tensor $\mathbf{T} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{w}$ are given by the array

$$T_{ijk} = u_i\, v_j\, w_k$$

- Given canonical bases, one often assimilates a *tensor* and its associated *array* of coordinates.

Order & Dimensions

Definitions Let the array $\mathbf{T} = \{T_{ij..k}\}$

- *Order* of $\mathbf{T} \stackrel{\mathrm{def}}{=} \#$ of its ways $= \#$ of its indices
- *Dimension* $K_\ell \stackrel{\mathrm{def}}{=}$ range of the ℓth index
- $\mathbf{T}$ is *Cubic* when all dimensions $K_\ell = K$ are equal
- $\mathbf{T}$ is *Symmetric* when it is cubic and when its entries do not change by *any* permutation of indices

Notation

- Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of dimensions $m_A \times n_A$ and $m_B \times n_B$, respectively
- Notation $\mathbf{A} \circ \mathbf{B}$ will be preferred to $\mathbf{A} \otimes \mathbf{B}$, to avoid possible confusion with the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ between matrices. In fact:
  - ☞ $\mathbf{A} \otimes \mathbf{B}$ is a *matrix* of size $m_A \, m_B \times n_A \, n_B$
  - ☞ $\mathbf{A} \otimes \mathbf{B}$ is a *tensor* of size $m_A \times n_A \times m_B \times n_B$

Matrix products

Again let $\mathbf{A}$ and $\mathbf{B}$ be matrices of dimensions $m_A \times n_A$ and $m_B \times n_B$, with entries $\{a_{ij}\}$ and $\{b_{ij}\}$, respectively

- **Kronecker product:** $\mathbf{A} \otimes \mathbf{B}$ is $m_A\, m_B \times n_A\, n_B$

$$\mathbf{A} \otimes \mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \end{pmatrix}$$

- **Khatri-Rao product** of matrices with same number of columns, n:

$$\mathbf{A} \odot \mathbf{B} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots \end{pmatrix}$$

  This is a *column-wise* Kronecker product. $\mathbf{A} \odot \mathbf{B}$ is $m_A\, m_B \times n$.

Outer product

- Outer or "tensor" product between two arrays, $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$:

$$C_{ij..\ell\, ab..d} = A_{ij..\ell}\, B_{ab..d}$$

  The *orders* add up

- **Example 15: Outer product between 2 vectors** The tensor

$$\mathbf{u} \circ \mathbf{v} = \mathbf{u}\, \mathbf{v}^{\mathsf{T}}$$

  has coordinates $u_i\, v_j$ and is of order 2, and is hence a matrix.

Arrays (cont'd)

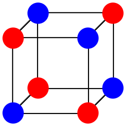
Example 16: Take

$$\mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Then

$$\mathbf{v}^{\circ 3} \stackrel{\text{def}}{=} \mathbf{v} \circ \mathbf{v} \circ \mathbf{v} = \left(\begin{array}{cc|cc} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{array} \right)$$

This is a "rank-1" symmetric tensor



blue bullets $= 1$, red bullets $= -1$.

Inner Product (1)

- **Mode-1 inner product:** $\mathbf{A} \bullet_1 \mathbf{B}$:

$$\{\mathbf{A} \underset{1}{\bullet} \mathbf{B}\}_{i_2..i_M, j_2..j_K} = \sum_k A_{ki_2..i_M} B_{kj_2..j_K}$$

This is a *contraction* on the 1st index

- **Mode$-p$ inner product:** similarly $\mathbf{A} \bullet_p \mathbf{B}$ is obtained by summing up (i.e. contracting) on the pth index
- **Example 17: Matrix-vector product** $\mathbf{A}\,\mathbf{u} = \mathbf{A}^\mathsf{T} \bullet_1 \mathbf{u}$
- **NB:**
there exists a (less convenient & less used) other notation: $\mathbf{A} \times_p \mathbf{B}$

Inner Product (2)

- **Example 18: Matrix products** are *contractions*

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{A} \underset{2}{\bullet} \mathbf{B}^{\mathsf{T}} = \mathbf{A}^{\mathsf{T}} \underset{1}{\bullet} \mathbf{B}$$

- **Example 19: Frobenius norm** of a Pth order tensor in $\mathbb{C}$:

$$||\mathbf{T}||^2 = \sum_{i_1\, i_2 \dots i_P} |T_{i_1\, i_2 \dots i_P}|^2 = \mathbf{T} \underset{1}{\bullet} \underset{2}{\bullet} \dots \underset{P}{\bullet} \mathbf{T}^*$$

One contracts on all indices

Inner Product (3)

- The *Contraction* is not associative

$$\mathbf{A} \underset{1}{\bullet} (\mathbf{B} \underset{1}{\bullet} \mathbf{C}) \neq (\mathbf{A} \underset{1}{\bullet} \mathbf{B}) \underset{1}{\bullet} \mathbf{C}$$

even for 2nd order tensors (matrices): $\mathbf{A}^{\mathsf{T}}\mathbf{B}^{\mathsf{T}}\mathbf{C} \neq \mathbf{B}^{\mathsf{T}}\mathbf{A}\,\mathbf{C}$

- **A convention** exists when a *single* tensor is contracted on several matrices, to avoid parentheses: the summation is always performed on the *second* matrix index.
Example 20: If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are matrices, and $\mathbf{T}$ a 3rd order tensor,

$$\mathbf{T}' = \mathbf{T} \underset{1}{\bullet} \mathbf{A} \underset{2}{\bullet} \mathbf{B} \underset{3}{\bullet} \mathbf{C} \Rightarrow T'_{pqr} = \sum_{ijk} A_{pi}\, B_{qj}\, C_{rk}\, T_{ijk} \qquad (20)$$

Change of basis

Assume a change of basis is performed in every linear space $\mathcal{V}_\ell$, e.g. defined by matrix **A** in $\mathcal{V}_1$, **B** in $\mathcal{V}_2$, ... and **C** in $\mathcal{V}_P$.

- Multilinearity. An order-P tensor **T** is transformed by the multi-linear map $\{\mathbf{A}, \mathbf{B}, ..\mathbf{C}\}$ into a tensor $\mathbf{T}'$:

$$T'_{ij..k} = \sum_{ab..c} A_{ia} B_{jb} .. C_{kc} T_{ab..c}$$

- Compact writing (with convention of slide 105):

$$\mathbf{T}' = \mathbf{T} \underset{1}{\bullet} \mathbf{A} \underset{2}{\bullet} \mathbf{B} \dots \underset{P}{\bullet} \mathbf{C}$$

Unfoldings (1)

- **Storage of a matrix in a vector** Let $\mathbf{A}$ be a $p \times q$ matrix, with columns $A_{:j}$. Then:

$$\mathbf{vec}\{\mathbf{A}\} \stackrel{\text{def}}{=} \begin{bmatrix} A_{:1} \\ A_{:2} \\ \vdots \\ A_{:q} \end{bmatrix} \quad (21)$$

- Conversely, $\mathbf{A} = \mathbf{Unvec}_q\,(\mathbf{vec}\{\mathbf{A}\})$, if q denotes the # of colums
- **Storage of a tensor in a vector**
  Similarly, the linear operator $\mathbf{vec}\{\cdot\}$ maps a $\alpha \times \beta \times \cdots \times \gamma$ tensor onto a vector ($\alpha\beta\ldots\gamma \times 1$ array)

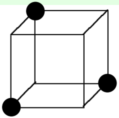
Unfoldings (2)

- **Storage of a tensor in a matrix**
  For a 3rd order tensor $\mathbf{T}$, one defines 3 *unfolding matrices*:

$$\mathbf{T}_{KI\times J} = \begin{bmatrix} \mathbf{T}_{::1} \\ \vdots \\ \mathbf{T}_{::k} \\ \vdots \\ \mathbf{T}_{::K} \end{bmatrix}, \quad \mathbf{T}_{IJ\times K} = \begin{bmatrix} \mathbf{T}_{1::} \\ \vdots \\ \mathbf{T}_{i::} \\ \vdots \\ \mathbf{T}_{I::} \end{bmatrix}, \quad \mathbf{T}_{JK\times I} = \begin{bmatrix} \mathbf{T}_{:1:}^{\mathsf{T}} \\ \vdots \\ \mathbf{T}_{:j:}^{\mathsf{T}} \\ \vdots \\ \mathbf{T}_{:J:}^{\mathsf{T}} \end{bmatrix},$$

- Conversely,
  $\textbf{Reshape}_{I,K,J}\,(\mathbf{T}_{KI\times J})$, $\textbf{Reshape}_{J,I,K}\,(\mathbf{T}_{IJ\times K})$ or
  $\textbf{Reshape}_{K,J,I}\,(\mathbf{T}_{JK\times I})$
  yield back $\mathbf{T}$ up to a permutation of the modes.
- Similar tools for higher orders...

ℓ-mode rank

- **Example 21:** $2 \times 2 \times 2$. Let $\mathbf{T} =$ 

  where bullets indicate nonzero entries, equal to 1 (see also slide 117). Then matrix unfoldings are

$$
\begin{aligned}
\mathbf{T}_{I \times JK} &= \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\
\mathbf{T}_{J \times KI} &= \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \\
\mathbf{T}_{K \times IJ} &= \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

- Note that ℓ *mode ranks can be different*:
  $\text{rank}_1 = \text{rank}_2 = 2 \neq \text{rank}_3 = 1$

Canonical Decomposition

- Tensor rank
- Properties of the CanD
- Normalized CanD
- Matrix writings of the CanD
- Rank can exceed dimensions
- Field can change rank

Tensor rank

- Any tensor or array $\mathbf{T}$, of dimensions $I \times J \times \cdots \times K$ can always be decomposed as

$$\mathbf{T} = \sum_q \mathbf{u}^{(q)} \circ \mathbf{v}^{(q)} \circ \ldots \circ \mathbf{w}^{(q)}$$

- The *tensor rank* of $\mathbf{T}$ is the minimal value of P such that equality holds
  This yields the *Canonical Decomposition* (CanD), sometimes referred to as *Parafac* decomposition:

$$\mathbf{T} = \sum_{q=1}^{\text{rank}\{\mathbf{T}\}} \mathbf{u}^{(q)} \circ \mathbf{v}^{(q)} \circ \ldots \circ \mathbf{w}^{(q)} \tag{22}$$

- Tensor rank is always larger than or equal to all ℓ-mode ranks:

$$\text{rank}_\ell\{\mathbf{T}\} \leq \text{rank}\{\mathbf{T}\}, \quad \forall \ell$$

Other writings (1)

- Vectors can be normalized to unit norm, yielding a normalized version:

$$\mathbf{T} = \sum_{q=1}^{\text{rank}\{\mathbf{T}\}} \lambda_q \, \mathbf{u}^{(q)} \circ \mathbf{v}^{(q)} \circ \ldots \circ \mathbf{w}^{(q)} \qquad (23)$$

☞ Will be useful for symmetric tensors in the real field

Other writings (2)

Let $\mathbf{T}$ be a 3rd order tensor, and denote $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ the matrices containing $\mathbf{u}^{(p)}$, $\mathbf{v}^{(p)}$, $\mathbf{w}^{(p)}$ as columns.

- Assuming $\boldsymbol{\Lambda}$ is a diagonal tensor of same order P as $\mathbf{T}$, with entries λ_q, the normalized CanD (23) admits a writing by contractions, with convention (20) of slide 105:

$$\mathbf{T} = \boldsymbol{\Lambda} \underset{1}{\bullet} \mathbf{U} \underset{2}{\bullet} \mathbf{V} \dots \underset{P}{\bullet} \mathbf{W}$$

  In other words, the CanD is a means to model a tensor as a transformation from a *diagonal* one.

☞ **Warning:** matrices $\mathbf{U}$, $\mathbf{V}$,... $\mathbf{W}$ may not be invertible nor even square!

Other writings (3)

- The CanD (22) can be written in matrix form:

$$\mathbf{T}_{I\times JK} = \mathbf{U}\,(\mathbf{W}\odot\mathbf{V})^{\mathsf{T}} \qquad (24)$$

- Alternatively, each matrix slice of **T** can be written as

$$\mathbf{T}_{::k} = \mathbf{U}\ \mathbf{Diag}\{\mathbf{W}(k,:)\}\ \mathbf{V}^{\mathsf{T}} \qquad (25)$$

NB: This extends to any order. In particular at order 4, with appropriate notations:

$$\mathbf{T}_{::k\ell} = \mathbf{A}\ \mathbf{Diag}\{\mathbf{C}(k,:)\}\ \mathbf{Diag}\{\mathbf{D}(\ell,:)\}\ \mathbf{B}^{\mathsf{T}}$$

Properties

- The CanD of a *multilinear transform* is the *transformed CanD*:
If $\mathbf{T} \stackrel{\text{def}}{=} \boldsymbol{\Lambda} \bullet_1 \mathbf{U} \bullet_2 \mathbf{V} \bullet_3 \mathbf{W}$ is transformed into
$\mathbf{T}' = \mathbf{T} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B} \bullet_3 \mathbf{C}$,
then $\mathbf{T}'$ admits the CanD:

$$\mathbf{T}' = \boldsymbol{\Lambda} \underset{1}{\bullet} (\mathbf{A}\,\mathbf{U}) \underset{2}{\bullet} (\mathbf{B}\,\mathbf{V}) \underset{3}{\bullet} (\mathbf{C}\,\mathbf{W})$$

- The CanD is valid in a ring (only multiplies)

Examples (1)

- **Example 22:** $2 \times 2 \times 2$ **tensor of rank 2**

$$\mathbf{T} = \left(\begin{array}{cc|cc} 1 & 2 & 2 & 4 \\ 3 & 4 & 6 & 8 \end{array}\right) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

here matrix slices are proportional

- **Example 23:** $2 \times 2 \times 2$ **of rank 2**

$$\mathbf{T} = \left(\begin{array}{cc|cc} 1 & 2 & 2 & 4 \\ 3 & 4 & 4 & 6 \end{array}\right) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \circ \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

even if matrix slices are not proportional

Examples (2)

- **Example 24:** $2 \times 2 \times 2$ **tensor of rank 3** [COM02b]

$$\mathbf{T} = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right)$$

and

$$2\,\mathbf{T} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\circ 3} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}^{\circ 3} + 2 \begin{pmatrix} 0 \\ -1 \end{pmatrix}^{\circ 3}$$

☞ This is the maximal rank in dimension $2 \times 2 \times 2$
☞ Here we have $\text{rank}_3 = 1 < \text{rank}_1 = \text{rank}_2 = 2 < \text{rank}\{\mathbf{T}\}$ (cf. slide 109).

NB: Other writing: $6x^2y = (x+y)^3 + (-x+y)^3 - 2y^3$

Field can change rank

- We have for any real tensor $\mathbf{T}$

$$\text{rank}\{\mathbf{T}\}_{\mathbb{C}} \leq \text{rank}\{\mathbf{T}\}_{\mathbb{R}}$$

Example 25: A $2 \times 2 \times 2$ tensor of rank 3 in $\mathbb{R}$, but 2 in $\mathbb{C}$ [CMLG06]

$$\mathbf{T} = \left(\begin{array}{cc|cc} -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array} \right)$$

In fact

$$\mathbf{T} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\circ 3} + \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}^{\circ 3} + 2 \begin{pmatrix} -1 \\ 0 \end{pmatrix}^{\circ 3} = \frac{\jmath}{2} \begin{pmatrix} -\jmath \\ 1 \end{pmatrix}^{\circ 3} - \frac{\jmath}{2} \begin{pmatrix} \jmath \\ 1 \end{pmatrix}^{\circ 3}$$

Symmetric tensors

- Usefulness
- Symmetric rank
- Link with quantics
- Why rank can exceed dimension
- Generic & typical ranks
- Clebsh's statement
- Topology
- Hirschowitz theorem

Usefulness of symmetric tensors

- They occur as derivatives of a multivariate function
  - Moments
  - Cumulants
  - Hessian

Space of symmetric tensors

- $\mathcal{S}_K$: symmetric tensors of dimensions K and order d
  ☞ space of dimension $D_s(K,d) = \binom{K+d-1}{d}$

K\d	quadric 2	cubic 3	quartic 4	quintic 5	sextic 6
2	3	4	5	6	7
3	6	10	15	21	28
4	10	20	35	56	84
5	15	35	70	126	210
6	21	56	126	252	462

Number of free parameters in a symmetric tensor of order d and dimension K

- $\mathcal{A}_K$: general tensors of dimensions $K_\ell = K$, $1 \leq \ell \leq d$
  ☞ space of dimension $D_A(K,d) = K^d$

Symmetric rank

- **Definition** For decomposing a *symmetric* tensor, one can impose symmetry of each rank-1 term. Hence the *symmetric rank*:

$$\mathbf{T} = \sum_{q=1}^{\text{rank}_s(\mathbf{T})} [\mathbf{u}^{(q)}]^{\circ P}$$

- **Property** We have that

$$\text{rank}\{\mathbf{T}\} \leq \text{rank}_s \mathbf{T}, \quad \forall\, \mathbf{T} \text{ symmetric}$$

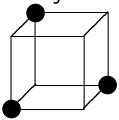
- It is not yet proved that both coincide for all values of order and dimensions:
  this is a conjecture [CGLM08].

Link with quantics (1)

- A *quantic* is a homogeneous polynomial in several variables. For instance: quadric, cubic, quartic...
- **Example 26: Binary cubic** $(d, K) = (3, 2)$
  Take again example in slide 117:

$$p(x_1, x_2) = \sum_{i,j,k=1}^{2} T_{ijk}\, x_i\, x_j\, x_k$$

$$\mathbf{T} = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right) =$$



$$\Rightarrow \quad p(\mathbf{x}) = 3\, x_1^2 x_2 = 3\, \mathbf{x}^{[2,1]}$$

Link with quantics (2)

- **Bijection:** Symmetric tensor of order d and dimension K $\leftrightarrow$ quantic of degree d in K variables:

$$p(\mathbf{x}) = \sum_{\mathbf{j}} T_{\mathbf{j}}\, \mathbf{x}^{\mathbf{f}(\mathbf{j})} \tag{26}$$

  - integer vector $\mathbf{j}$ of dimension d $\leftrightarrow$ integer vector $\mathbf{f}(\mathbf{j})$ of dimension K
  - entry f_k of $\mathbf{f}(\mathbf{j})$ being $\stackrel{\text{def}}{=}$ #of times index k appears in $\mathbf{j}$
  - We have in particular $|\mathbf{f}(\mathbf{j}))| = d$.

- Standard conventions: $\mathbf{x}^{\mathbf{j}} \stackrel{\text{def}}{=} \prod_{k=1}^{K} x_k^{j_k}$ and $|\mathbf{f}| \stackrel{\text{def}}{=} \sum_{k=1}^{K} f_k$, where $\mathbf{j}$ and $\mathbf{f}$ are integer vectors.

Literature

Gauss'1825
Sylvester'1851
Cayley'1854
Clebsch'1861
Salmon'1874
Poincaré'1890
Hilbert'1900
Wakeford'1918
Grothendieck'1966
Dieudonné'1970
Shafarevich'1975

Ehrenborg, Kogan...

Why rank can exceed dimension

Theorem Let $\mathbf{v}_{(1)}$, $\mathbf{v}_{(2)}$, ... $\mathbf{v}_{(r)}$, be r *pairwise* linearly independent vectors, then for all $k \geq r-1$, the rank-1 symmetric tensors are *linearly independent*:

$$\mathbf{v}_{(1)}^{\circ k}, \ \mathbf{v}_{(2)}^{\circ k}, \dots \mathbf{v}_{(r)}^{\circ k}$$

Example 27: 3 vectors in dimension 2

$$\mathbf{v}_{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \mathbf{v}_{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \mathbf{v}_{(3)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

are pairwise linearly independent, but matrix of $\{\mathbf{v}_{(q)}^{\circ 2}\}$ is full rank:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Orbits (1)

- General Linear group $\mathcal{GL}$: group of invertible matrices
- Orbit of a polynomial p: all polynomials q that can be transformed into p by $\mathbf{A} \in \mathcal{GL}$: $q(\mathbf{x}) = p(\mathbf{A}\mathbf{x})$.
- Allows to classify polynomials

Orbits (2)

Example 28: Quadrics

- Binary quadrics are associated with 2×2 symmetric matrices (tensors of order 2)
  - Orbits in $\mathbb{R}$: $\{0,\ x^2,\ x^2+y^2,\ x^2-y^2\}$
    ☞ $2xy \in \mathcal{O}(x^2-y^2)$ in $\mathbb{R}[x,y]$
  - Orbits in $\mathbb{C}$: $\{0,\ x^2,\ x^2+y^2\}$
    ☞ $2xy \in \mathcal{O}(x^2+y^2)$ in $\mathbb{C}[x,y]$
- Set of singular matrices is closed
- Set $\mathcal{Y}_r$ of matrices of at most rank r is closed

3×3

Classification of ternary quadrics

Orbits in $\mathbb{C}$:

$\mathcal{GI}-$orbit	$\omega(p)$
0	0
x^2	1
x^2+y^2	2
$x^2+y^2+z^2$	3 (**generic**)

Question: what is the answer in $\mathbb{R}$?

CanD of polynomials

By using bijection (26), decomposing a dth order symmetric tensor into a sum of rank-1 tensors means

$$p(\mathbf{x}) = \sum_{q=1}^{r(p)} (\mathbf{v}_{(q)}^{\mathsf{T}} \mathbf{x})^d \qquad (27)$$

- This is a sum of powers of linear forms.
- $r(p)$ coincides with the rank of associated tensor
- $r(p)$ is sometimes called the *width* of p [REZ92].

Generic & Typical Ranks

- **Informal definition** A property is *typical* if it holds true on a non-zero-volume set
- **Informal definition** A property is *generic* if is true almost everywhere.
- There can be *several* typical ranks, but only *a single* generic rank.

Bounds on generic rank (1)

For quantics of degree d in K variables

- Lower bound

$$\left\lceil \frac{\binom{K+d-1}{d}}{K} \right\rceil \leq \overline{R}$$

- Upper bound [Reznick'92]

$$\overline{R} \leq \binom{K+d-2}{d-1}$$

Bounds on generic rank (2)

- Tensors of order d and dimensions $(K_1, ..K_d)$ without symmetry:
  - Upper bound
    $$\left\lceil \frac{\prod_{i=1}^{d} K_i}{1 + \sum_{i=1}^{d}(K_i - 1)} \right\rceil \leq \overline{R}$$
  - Square case $K_i = K$:
    $$K^d/(dK - d + 1) \leq \overline{R}$$
- Lower bound (Square case):
  $$K^d/(d\,K - d + 1) \leq \overline{R}$$

Topology of quantics

- Every elementary closed set $\stackrel{\text{def}}{=}$ varieties, defined by $p(\mathbf{x}) = 0$
- Closed sets = finite union of varieties
- Closure of a set $\mathcal{E}$: smallest closed set $\overline{\mathcal{E}}$ containing $\mathcal{E}$

➽ is called the *Zariski* topology in $\mathbb{C}$ [CLO92]

➽ this is not Euclidian topology, but results still apply [CGLM08]: Tensors with *entries randomly drawn* according to a continuous pdf are generic

Clebsch's statement

Alfred Clebsch (1833-1872)

The generic ternary quartic cannot be written as the sum of 5 fourth powers

- $D(3,4) = 15$
- $3\,r$ free parameters in the CAND
- But $r = 5$ is not enough $\rightarrow r = 6$ is generic !

Tensor subsets

- Set of tensors of rank *at most r* with values in $\mathbb{C}$:

$$\mathcal{Y}_r = \{\mathbf{T} \in \mathcal{T} :\ r(\mathbf{T}) \leq r\}$$

- Set of tensors of rank *exactly r*: $\mathcal{Z}_r = \{\mathbf{T} \in \mathcal{T} :\ r(\mathbf{T}) = r\}$

$$\mathcal{Z} = \mathcal{Y}_r - \mathcal{Y}_{r-1},\ r > 1$$

- Zariski closures: $\overline{\mathcal{Y}}_r$, $\overline{\mathcal{Z}}_r$?

Lack of closeness of $\mathcal{Z}_r$

- PROPOSITION
  $\mathcal{Z}_1$ is closed, *but not* $\mathcal{Z}_r$ for any $r > 1$

  [Burgisser'97] [Strassen'83]
- Proof
  If $\text{rank}\{\mathbf{T}\} > 1$, there exist $\mathbf{T}_0 \in \mathcal{Z}_{r-1}$ and $\mathbf{y} \neq 0$ such that

$$\mathbf{T} = \mathbf{T}_0 + \mathbf{y}^{\circ\, d}$$

  Then define $\mathbf{T}_\varepsilon = \mathbf{T}_0 + \varepsilon\, \mathbf{y}^{\circ\, d}$. This series converges to $\mathbf{T}_0 \notin \mathcal{Z}_r$ as $\varepsilon \to 0$

Lack of closeness of $\mathcal{Y}_r$ (1)

- PROPOSITION
  If $d > 2$, $\mathcal{Y}_r$ is not closed for $1 < r < R$.
- **Example 29: Sequence of rank-2 tensors converging towards a rank-4:**

$$\mathbf{T}_\varepsilon = \frac{1}{\varepsilon}\left[(\mathbf{u} + \varepsilon\,\mathbf{v})^{\circ 4} - \mathbf{u}^{\circ 4}\right]$$

In fact, as $\varepsilon \to 0$, it tends to:

$$\mathbf{T}_0 = \mathbf{u} \circ \mathbf{u} \circ \mathbf{u} \circ \mathbf{v} + \mathbf{u} \circ \mathbf{u} \circ \mathbf{v} \circ \mathbf{u} + \mathbf{u} \circ \mathbf{v} \circ \mathbf{u} \circ \mathbf{u} + \mathbf{v} \circ \mathbf{u} \circ \mathbf{u} \circ \mathbf{u}$$

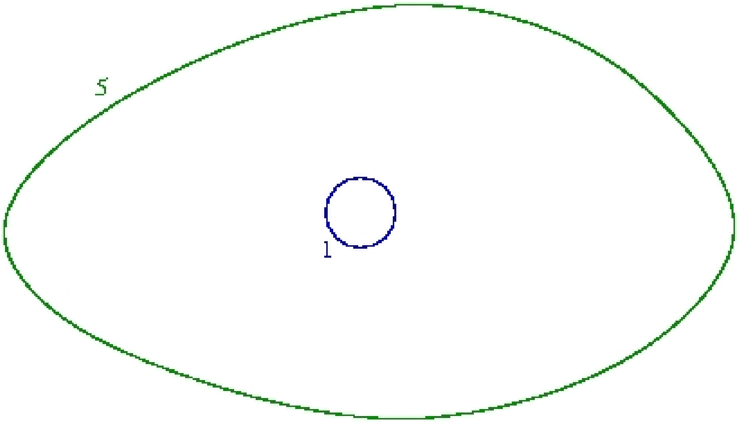
which can be shown to be proportional to the rank-4 tensor:

$$3\mathbf{T}_0 = 8\,(\mathbf{u}+\mathbf{v})^{\circ 4} - 8(\mathbf{u}-\mathbf{v})^{\circ 4} - (\mathbf{u}+2\mathbf{v})^{\circ 4} + (\mathbf{u}-2\mathbf{v})^{\circ 4} \quad (28)$$

where **u** and **v** are not collinear.
☞ This is the *maximal rank* of 4th order tensors of dimension 2.

Lack of closeness of $\mathcal{Y}_r$ (2)

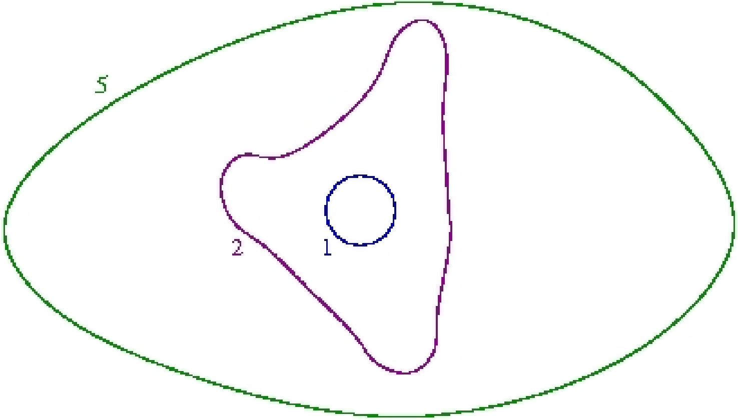
Successive sets $\mathcal{Y}_r = \{\mathbf{T} : \ rank(\mathbf{T}) \leq r\}$



- A tensor sequence in $\mathcal{Y}_r$ can converge to a limit in $\mathcal{Y}_{r+h}$

Lack of closeness of $\mathcal{Y}_r$ (2)

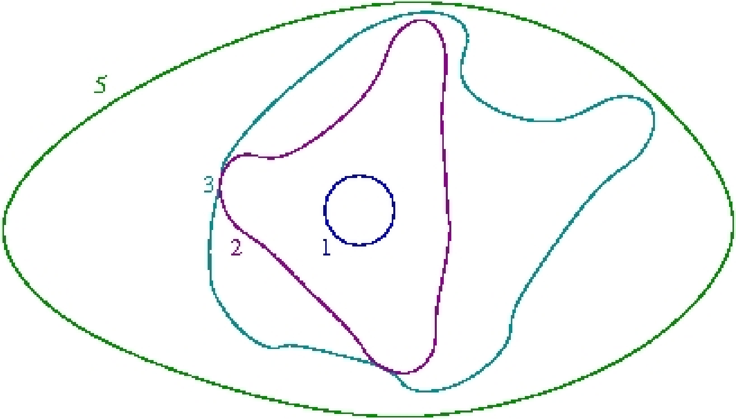
Successive sets $\mathcal{Y}_r = \{\mathbf{T} : \ rank(\mathbf{T}) \leq r\}$



- A tensor sequence in $\mathcal{Y}_r$ can converge to a limit in $\mathcal{Y}_{r+h}$

Lack of closeness of $\mathcal{Y}_r$ (2)

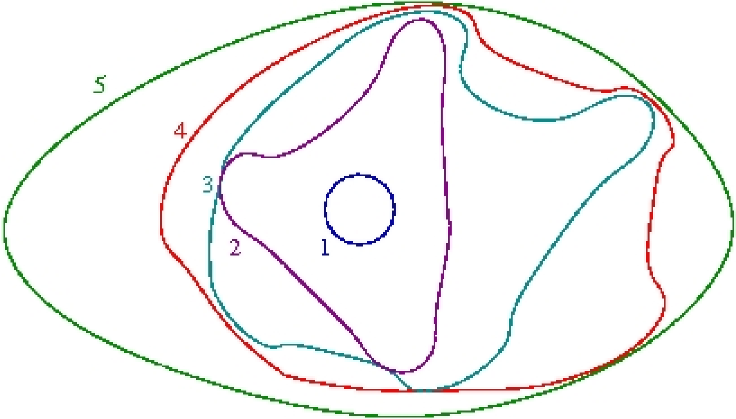
Successive sets $\mathcal{Y}_r = \{\mathbf{T} : \; rank(\mathbf{T}) \leq r\}$



- A tensor sequence in $\mathcal{Y}_r$ can converge to a limit in $\mathcal{Y}_{r+h}$

Lack of closeness of $\mathcal{Y}_r$ (2)

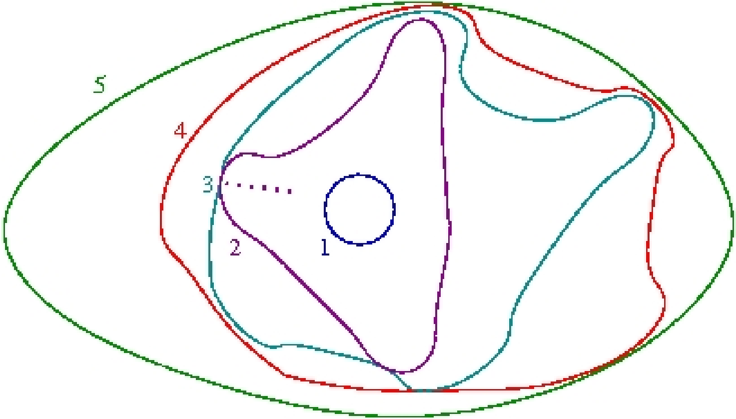
Successive sets $\mathcal{Y}_r = \{\mathbf{T} : \; rank(\mathbf{T}) \leq r\}$



- A tensor sequence in $\mathcal{Y}_r$ can converge to a limit in $\mathcal{Y}_{r+h}$

Lack of closeness of $\mathcal{Y}_r$ (2)

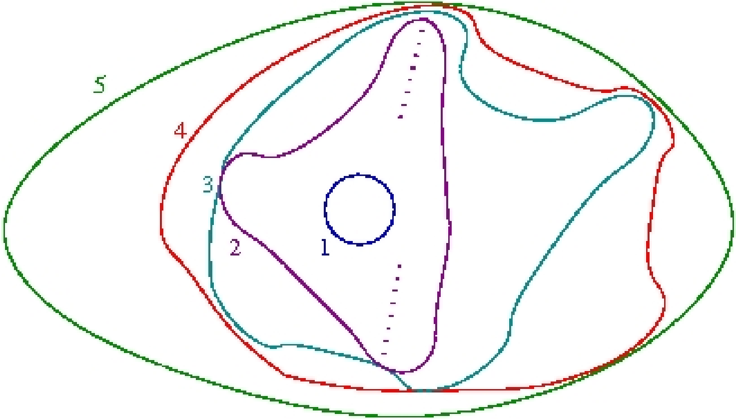
Successive sets $\mathcal{Y}_r = \{\mathbf{T} : \ rank(\mathbf{T}) \leq r\}$



- A tensor sequence in $\mathcal{Y}_r$ can converge to a limit in $\mathcal{Y}_{r+h}$

Lack of closeness of $\mathcal{Y}_r$ (2)

Successive sets $\mathcal{Y}_r = \{\mathbf{T} : \ rank(\mathbf{T}) \leq r\}$



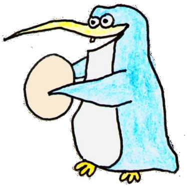
- A tensor sequence in $\mathcal{Y}_r$ can converge to a limit in $\mathcal{Y}_{r+h}$

Genericity

- **Formal definition** r is a typical rank if (density argument with Zariski):

  $$\overline{\mathcal{Z}}_r \text{ is the whole space}$$

- **Formal definition** Generic rank is *the typical rank when unique*
- In $\mathbb{C}$ a typical rank is unique, and hence generic
- For given values of order d and dimension K, the smallest typical rank in $\mathbb{R}$ coincides with the generic rank in $\mathbb{C}$



Existence of the generic rank in $\mathbb{C}$

- LEMMA The series of $\overline{\mathcal{Y}}_k$ is strictly increasing for $k \leq \overline{R}$ and then constant:

$$\overline{\mathcal{Y}}_1 \subsetneq \overline{\mathcal{Y}}_2 \subsetneq \dots \subsetneq \overline{\mathcal{Y}}_{\overline{R}} = \overline{\mathcal{Y}}_{\overline{R}+1} = \dots \mathcal{T}$$

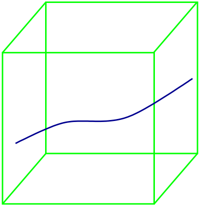
  which guarantees the existence of a unique $\overline{R}$

- PROPOSITION For tensors in $\mathbb{C}$
  If $r_1 < r_2 < \overline{R} < r_3 \leq R$, then

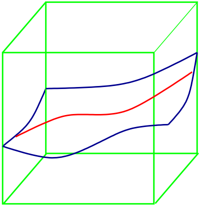
$$\overline{\mathcal{Z}}_{r_1} \subset \overline{\mathcal{Z}}_{r_2} \subset \overline{\mathcal{Z}}_{\overline{R}} \supset \overline{\mathcal{Z}}_{r_3} \supseteq \overline{\mathcal{Z}}_R \qquad (29)$$

➽ Proves that $\overline{R}$ is the generic rank in $\mathbb{C}$

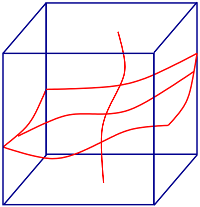
Generic rank in $\mathbb{C}$



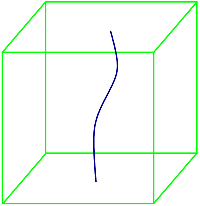
$\mathcal{Z}_1$



$\mathcal{Z}_2 = \mathcal{Y}_2 - \mathcal{Z}_1$



$$\begin{aligned} \mathcal{Z}_3 &= \mathcal{Y}_3 - \mathcal{Z}_1 - \mathcal{Z}_2 \\ &= \mathcal{T} - \mathcal{Z}_1 - \mathcal{Z}_2 - \mathcal{Z}_4 \end{aligned}$$



$\mathcal{Z}_4 = \mathcal{Y}_4 - \mathcal{Y}_3$

Numerical computation of the Generic Rank (1)

Mapping

$$\{\mathbf{u}(\ell),\ 1 \le \ell \le r\} \xrightarrow{\varphi} \sum_{\ell=1}^{r} \mathbf{u}(\ell)^{\circ\, d}$$

$$\{\mathbb{C}^K\}^r \xrightarrow{\varphi} \mathcal{S}$$

Rank

The rank of the Jacobian of φ equals $\dim(\bar{\mathcal{Z}}_r)$, and hence D for large enough r.

➽ The smallest r for wich $rank(\mathrm{Jacobian}(\varphi)) = D$ is $\bar{R}$.

Numerical computation of the Generic Rank (2)

Example 30: 3rd order symmetric tensors

$$\{\mathbf{u}(\ell),\, 1 \leq \ell \leq r\} \xrightarrow{\varphi} \mathbf{T} = \sum_{\ell=1}^{r} \mathbf{u}(\ell)^{\circ\, 3}$$

$\mathbf{T}$ has coordinate vector: $\sum_{\ell=1}^{r} \mathbf{u}(\ell) \otimes \mathbf{u}(\ell) \otimes \mathbf{u}(\ell)$. Hence the Jacobian of φ is the $r\, n \times n^3$ matrix:

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_n \otimes \mathbf{u}^{\mathsf{T}}(1) \otimes \mathbf{u}^{\mathsf{T}}(1) & + & \mathbf{u}(1)^{\mathsf{T}} \otimes \mathbf{I}_n \otimes \mathbf{u}^{\mathsf{T}}(1) & + & \mathbf{u}(1)^{\mathsf{T}} \otimes \mathbf{u}(1)^{\mathsf{T}} \otimes \mathbf{I}_n \\ \mathbf{I}_n \otimes \mathbf{u}^{\mathsf{T}}(2) \otimes \mathbf{u}^{\mathsf{T}}(2) & + & \mathbf{u}(2)^{\mathsf{T}} \otimes \mathbf{I}_n \otimes \mathbf{u}^{\mathsf{T}}(2) & + & \mathbf{u}(2)^{\mathsf{T}} \otimes \mathbf{u}(2)^{\mathsf{T}} \otimes \mathbf{I}_n \\ \dots & + & \dots & + & \dots \\ \mathbf{I}_n \otimes \mathbf{u}^{\mathsf{T}}(r) \otimes \mathbf{u}^{\mathsf{T}}(r) & + & \mathbf{u}(r)^{\mathsf{T}} \otimes \mathbf{I}_n \otimes \mathbf{u}^{\mathsf{T}}(r) & + & \mathbf{u}(r)^{\mathsf{T}} \otimes \mathbf{u}(r)^{\mathsf{T}} \otimes \mathbf{I}_n \end{bmatrix}$$

and

$$\begin{cases} \text{rank}\{\mathbf{J}\} = dim(\mathrm{Im}(\varphi)) \\ \bar{R} = Min\{r : \ \mathrm{Im}\{\varphi\} = \mathcal{S}\} \end{cases}$$

Numerical computation of the Generic Rank (3)

The symmetric rank is generically:

d \ K	2	3	4	5	6	7	8
3	2	4	5	**8**	10	12	15
4	3	**6**	**10**	**15**	21	30	42

$$\bar{R}_s \geq \frac{1}{K}\binom{K+d-1}{d}$$

Bold: exceptions to the ceil rule: $\bar{R}_s \geq \lceil \frac{1}{K}\binom{K+d-1}{d} \rceil$ [CM96]

Uniqueness of CanD

Number of solutions

- The fiber of solutions has dimension

$$F(n) = K\,\bar{R} - \binom{K+d-1}{d}$$

d \ K	2	3	4	5	6	7	8
3	0	2	0	**5**	4	0	0
4	1	**3**	**5**	**5**	0	0	6

➽ 0 means a finite number of solutions

Hirschowitz theorem

From Alexander-Hirschowitz theorem (cf. appendix), one can deduce [CGLM08]:

THEOREM For $d > 2$, the generic rank of a dth order symmetric tensor of dimension K is always equal to the lower bound

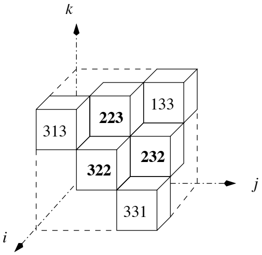
$$\bar{R}_s = \left\lceil \frac{\binom{K+d-1}{d}}{K} \right\rceil \tag{30}$$

except for the following cases:
$(d, K) \in \{(3, 5), (4, 3), (4, 4), (4, 5)\}$, for which it should be increased by 1.

➽ Only a *finite number* of exceptions !

Classification of ternary cubics

$\mathcal{GI}-$orbit	$\omega(p)$
x^3	1
$x^3 + y^3$	2
$x^2 y$	3
$x^3 + 3\, y^2 z$	4
$x^3 + y^3 + 6\, xyz$	4
$x^3 + 6\, xyz$	4
$a\,(x^3 + y^3 + z^3) + 6b\, xyz$	4 (**generic**)
$xz^2 + y^2 z$	5



George Salmon (1819-1904)

Other tensors

- Numerical computation of the Generic Rank
- Uniqueness of the CanD
- Tensors with particular symmetries
- Link with polynomials

Numerical computation of the Generic Rank (1)

Mapping

$$\{\mathbf{u}(\ell), \mathbf{v}(\ell), \dots, \mathbf{w}(\ell),\ 1 \leq \ell \leq r\} \xrightarrow{\varphi} \sum_{\ell=1}^{r} \mathbf{u}(\ell) \circ \mathbf{v}(\ell) \circ \dots \circ \mathbf{w}(\ell)$$

$$\{\mathbb{C}^{n_1} \circ \dots \circ \mathbb{C}^{n_d}\}^r \xrightarrow{\varphi} \mathcal{A}$$

➠ The smallest r for wich $rank(\mathrm{Jacobian}(\varphi)) = D$ is the generic rank, $\bar{R}$.

Numerical computation of the Generic Rank (2)

Example 31: 3rd order non symmetric tensors

$$\{\mathbf{a}(\ell), \mathbf{b}(\ell), \mathbf{c}(\ell)\} \xrightarrow{\varphi} \mathbf{T} = \sum_{\ell=1}^{r} \mathbf{a}(\ell) \circ \mathbf{b}(\ell) \circ \mathbf{c}(\ell)$$

$\mathbf{T}$ has coordinate vector: $\sum_{\ell=1}^{r} \mathbf{a}(\ell) \otimes \mathbf{b}(\ell) \otimes \mathbf{c}(\ell)$. Hence the Jacobian of φ is the $r(n_1 + n_2 + n_3) \times n_1 n_2 n_3$ matrix:

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{n_1} & \otimes & \mathbf{b}^{\mathsf{T}}(1) & \otimes & \mathbf{c}^{\mathsf{T}}(1) \\ \mathbf{I}_{n_1} & \otimes & \dots & \otimes & \dots \\ \mathbf{I}_{n_1} & \otimes & \mathbf{b}^{\mathsf{T}}(r) & \otimes & \mathbf{c}^{\mathsf{T}}(r) \\ \mathbf{a}(1)^{\mathsf{T}} & \otimes & \mathbf{I}_{n_2} & \otimes & \mathbf{c}^{\mathsf{T}}(1) \\ \dots & \otimes & \mathbf{I}_{n_2} & \otimes & \dots \\ \mathbf{a}(r)^{\mathsf{T}} & \otimes & \mathbf{I}_{n_2} & \otimes & \mathbf{c}^{\mathsf{T}}(r) \\ \mathbf{a}(1)^{\mathsf{T}} & \otimes & \mathbf{b}(1)^{\mathsf{T}} & \otimes & \mathbf{I}_{n_3} \\ \dots & \otimes & \dots & \otimes & \mathbf{I}_{n_3} \\ \mathbf{a}(r)^{\mathsf{T}} & \otimes & \mathbf{b}(r)^{\mathsf{T}} & \otimes & \mathbf{I}_{n_3} \end{bmatrix} \text{ and } \begin{cases} \text{rank}\{\mathbf{J}\} = dim(\mathrm{Im} \\ \bar{R} = Min\{r : \ \mathrm{Im}\{\varphi \end{cases}$$

Numerical computation of the Generic Rank (3)

Example 32: Tensors of order d with dimensions all equal to K

d \ K	2	3	4	5	6	7
3	2	**5**	7	10	14	19
4	4	9	20	37	62	97

$$\bar{R} \geq \frac{K^d}{Kd - d + 1}$$

Bold: exceptions to the ceil rule $\bar{R} = \lceil \frac{K^d}{Kd-d+1} \rceil$

Uniqueness of CanD

Number of solutions

- **Example 33: 3rd order with dimensions n_ℓ**
  $F(n_1, n_2, n_3) = (n_1 + n_2 + n_3 - 2)\,\bar{R} - n_1\, n_2\, n_3$
- **Example 34: dth order with equal dimensions, K**
  $F(n) = (Kd - d + 1)\bar{R} - K^d$

d \ K	2	3	4	5	6	7
3	0	**8**	6	5	8	18
4	4	0	4	4	6	24

➽ For generic/typical values, almost always infinitely many CanD's

Numerical computation of the Generic Rank (4)

Example 35: 3rd order tensors with unequal dim. N_ℓ [CtB08] [CtB06]

N_3		2				3			4	
N_2		2	3	4	5	3	4	5	4	5
N_1	2	2,3	3	4	4	3,4	4	5	4,5	5
	3	3	3,4	4	5	5	**5**	5,6	**6**	**6**
	4	4	4	4,5	5	5	**6**	**6**	**7**	**8**
	5	4	5	5	5,6	5,6	**6**	**8**	**8**	**9**
	6	4	6	6	6	6	**7**	**8**	**8**	**10**
	7	4	6	7	7	7	**7**	**9**	**9**	**10**
	8	4	6	8	8	8	8,9	**9**	**10**	**11**
	9	4	6	8	9	9	9	**9**	**10**	**12**

- There are exceptions to the ceil rule $\bar{R} = \lceil \frac{\prod_\ell N_\ell}{\sum_\ell (N_\ell - 1) + 1} \rceil$
- **Bold**: values that have not yet been proved theoretically

Third order tensors with symmetric slices

Example 36: Typical ranks for $N_1 \times N_2 \times N_2$ arrays, with $N_2 \times N_2$ symmetric slices.

N_1 N_2	2	3	4	5
2	2,3	3,4	4,5	5,6
3	3	4	**6**	**7**
4	3	4,5	**6**	**8**
5	3	5,6	**7**	**9**
6	3	6	**7**	**9**
7	3	6	**7**	**10**
8	3	6	**8**	**10**
9	3	6	9,10	**11**
10	3	6	10	**11**

Bold: smallest typical ranks computed numerically.
Plain: known typical ranks; in $\mathbb{C}$, the smallest value is generic.

Tucker 3

- Definitions
- Properties
- Usefulness

Definition (1)

According to Ledyard R. Tucker (1910-2004), any dth order $I_1 \times I_2 \times \cdots \times I_d$ tensor $\mathbf{T}$ can be decomposed as [TUC66]:

$$\mathbf{T} = \mathbf{S} \underset{1}{\bullet} \mathbf{U}^{(1)} \underset{1}{\bullet} \mathbf{U}^{(2)} \dots \underset{1}{\bullet} \mathbf{U}^{(d)}$$

where $\mathbf{S}$ has *smaller dimensions* than $\mathbf{T}$ (or equal to), and $\mathbf{U}^{(\ell)}$ are semi-unitary, i.e. $\mathbf{U}^{(\ell)\mathsf{T}}\mathbf{U}^{(\ell)} = \mathbf{I}_{n_\ell}$, $n_\ell \leq I_\ell$.

- $\mathbf{S}$ is called the *core tensor*.
- This decomposition is referred to as *Tucker3* or as *HOSVD* [dLdMV00b] [SBG04].

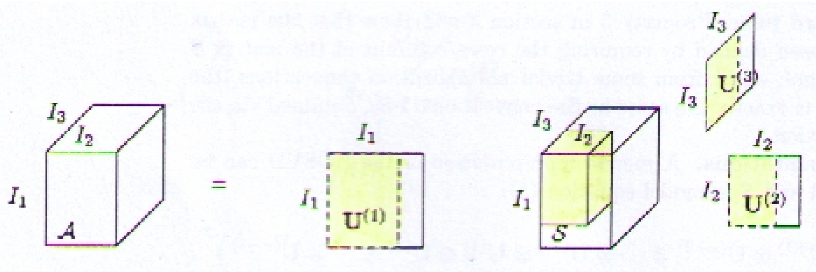
Two equivalent optimization problems

$$\operatorname*{Max}_{\mathbf{U}^{(1)},\mathbf{U}^{(2)}\ldots\mathbf{U}^{(d)}} ||\mathbf{T} \underset{1}{\bullet} \mathbf{U}^{(1)\mathsf{T}} \underset{1}{\bullet} \mathbf{U}^{(2)\mathsf{T}} \ldots \underset{1}{\bullet} \mathbf{U}^{(d)\mathsf{T}}||^2$$

$$\operatorname*{Min}_{\mathbf{U}^{(1)},\mathbf{U}^{(2)}\ldots\mathbf{U}^{(d)}} ||\mathbf{T} - \mathbf{S} \underset{1}{\bullet} \mathbf{U}^{(1)} \underset{1}{\bullet} \mathbf{U}^{(2)} \ldots \underset{1}{\bullet} \mathbf{U}^{(d)}||^2$$

Definition (2)

Other writing with unitary matrices $\mathbf{U}^{(\ell)}$. In that case, the core tensor $\mathbf{S}$ has same dimensions as $\mathbf{T}$ but is padded with zeros [dLdMV00b]:



Properties (1)

The dth order core tensor can be imposed to be quite particular

- All its $d-1$st order subtensors obtained by fixing one index are *all orthogonal* (w.r.t. scalar product induced by Frobenius norm); there are d of them.
- Entries of the core tensor can be sorted in such a way that for every mode ℓ:

$$||\mathbf{S}_{i_\ell=1}|| \geq ||\mathbf{S}_{i_\ell=2}|| \geq \dots ||\mathbf{S}_{i_\ell=I_\ell}||$$

- These norms may be viewed as ℓ-mode singular values.
- When $\mathbf{T}$ is a matrix, so is $\mathbf{S}$, and all-orthogonality can be satisfied only when $\mathbf{S}$ is diagonal. The sequence of norms $\sigma_i = ||S_{:i}|| = ||S_{i:}||$ are then the singular values.

Properties (2)

- ℓ-mode singular vectors can be computed as singular vectors of the ℓ-mode unfolding matrix; hence an easy computation
- The ℓ-mode singular values are uniquely defined
- When ℓ-mode singular values are different, corresponding ℓ-mode singular vectors are unique up to a unit-modulus scale factor
- For any fixed mode ℓ, the sum of all mode-ℓ squared singular values yields $||\mathbf{T}||^2$

Usefulness

- The nesting of ℓ-mode singular values & vectors allows to easily find the best approximate of a tensor of lower ℓ-mode rank by truncation of the HOSVD [dLdMV00c].
- May be applied to noise reduction
- May reduce subsequent computational complexity (dimension reduction)
- May be used as a pre-processing before the CanD calculation

Other decompositions

- Exact decompositions (if not truncated):
  - CanD
  - Tucker3 – HOSVD
- Approximate decompositions:
  - Diagonalization by orthogonal transform
  - Diagonalization by invertible transform

Conclusions on Tensors

- Still open problems
- Efficient numerical algorithms lacking
- Several ways of extending SVD to tensors
- Very powerful, and numerous application areas

Polynomial interpolation

Alexander-Hirschowitz Theorem [AH92] [AH95] Let $\mathcal{L}(d, m)$ be the space of hypersurfaces of degree at most d in m variables. This space is of dimension $D(m, d) \stackrel{\text{def}}{=} \binom{m+d}{d} - 1$.

THEOREM Denote $\{p_i\}$ K given distinct points in the complex projective space $\mathbb{P}^m$. The dimension of the linear subspace of hypersurfaces of $\mathcal{L}(d, m)$ having multiplicity at least 2 at every point p_i is:

$$D(m, d) - K(m + 1)$$

except for the following cases:

- $d = 2$ and $2 \leq K \leq m$
- $d \geq 3$ and $(m, d, K) \in \{(2, 4, 5), (3, 4, 9), (4, 1, 14), (4, 3, 7)\}$

In other words, there are a *finite number* of exceptions.

Part IV

Algorithms for static mixtures

Contents of part IV

Overview

- Introduction
- Algorithms based on pair sweeping (CoM1, CoM2)
  Link with tensor diagonalization
- Algorithms based on matrix slices (JADE, STOTD)
- Algorithms based on Deflation (FastICA, RobustICA, SAUD)
- Finite alphabet inputs (APF, MAP, ILSP...)

References

What we have seen so far

- Cumulants can measure independence at a given order
- Cumulants form a (symmetric) tensor object
- Tensors may have a rank larger than dimensions, even generically
- We have well-founded optimization criteria. Some of them amount to *approximately* diagonalizing a tensor.

Hypotheses

- Mixture is over-determined
- The rank of the signal cumulant tensor is equal (at most) to its dimension
- The mixture may be given by the CanD of the signal cumulant tensor
- Noise & measurement errors yield a measured cumulant tensor that has generic rank

Performance measure

How to test performances of algorithms in computer simulations?

- Difficulty because of the $\mathbf{\Lambda P}$ indeterminacy
- **Identification:** Gap between $\mathbf{F H}$ and matrix of the form $\mathbf{\Lambda P}$
- **Source extraction:** SINR (Signal to Interference plus Noise): needs exhaustive search for best $\mathbf{\Lambda P}$

Example of Gap

This gap does not need a combinatorial search, because it is $\mathbf{\Lambda}\,\mathbf{P}$-invariant [COM94a]:

$$\begin{aligned}\varepsilon(\mathbf{A},\hat{\mathbf{A}}) &= \sum_i \Big|\sum_j |\mathbf{D}_{ij}| - 1\,\Big|^2 + \Big|\sum_j |\mathbf{D}_{ij}|^2 - 1\,\Big| \\ &+ \sum_j \Big|\sum_i |\mathbf{D}_{ij}| - 1\,\Big|^2 + \Big|\sum_i |\mathbf{D}_{ij}|^2 - 1\,\Big|\end{aligned}$$

where $\mathbf{D} = \mathbf{A}^{-1}\hat{\mathbf{A}}$

Properties

- $\varepsilon\{\mathbf{A}\mathbf{\Lambda}\mathbf{P},\hat{\mathbf{A}}\} = \varepsilon\{\mathbf{A},\hat{\mathbf{A}}\} = \varepsilon\{\mathbf{A},\hat{\mathbf{A}}\,\mathbf{\Lambda}^{-1}\mathbf{P}\}$
- $\varepsilon\{\mathbf{A},\hat{\mathbf{A}}\} = 0 \Leftrightarrow \hat{\mathbf{A}} = \mathbf{A}\mathbf{\Lambda}\mathbf{P}$

Algorithms based on pair sweeping

- Block vs Adaptive
- Closed-form solutions in dimension 2, for various contrasts
- Sweeping of all pairs
- Complexity and convergence

Numerical Algorithms

What problem are they supposed to solve?

- Are we given a single block of data?
- Are we observing a sequence of blocks, or a series of samples?
- Must we update the solution at every block, or at every sample?

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*

Block vs Adaptive

- Increase power of DSP
- Limitations of time-recursive Adaptive Algorithms
  - Convergence time of optimization algorithm
  - Convergence time of moment estimators
  - Local extrema harder to handle
- Coherence time sometimes limited
  (e.g. GSM: 900MHz, 190km/h, $T_c \approx 2ms \approx 300$ symbols)
- Well matched to block transmission (TDMA)
- Better exploitation of data
  (uniform weight, resistance to loss in synchro, time reversal)

Solution of the 2-dimensional problem

- Assume data x have been standardized into $\tilde{\mathbf{x}}$.
- Then one looks for an estimate $\mathbf{z}$ of the source vector $\mathbf{s}$ as:

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

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

$$\mathbf{Q} = \begin{pmatrix} \cos\beta & \sin\beta\, e^{\jmath\varphi} \\ -\sin\beta\, e^{-\jmath\varphi} & \cos\beta \end{pmatrix} = \frac{1}{\sqrt{1+\theta\theta^*}} \begin{pmatrix} 1 & \theta \\ -\theta^* & 1 \end{pmatrix} \tag{31}$$

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

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 $\mathbf{P}\,\mathbf{\Lambda}$ 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 (31) fixes this indeterminacy, so that only 2 solutions remain

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

Invariance & Indeterminacy (2)

- Remark that $\mathbf{Q}[\theta]$ and $\mathbf{Q}[-1/\theta^*]$ are $\mathbf{P\Lambda}$-related:

$$\mathbf{Q}[-\frac{1}{\theta^*}] = \mathbf{Q}[\theta] \left(\begin{array}{cc} 0 & -e^{\jmath\varphi} \\ e^{-\jmath\varphi} & 0 \end{array} \right)$$

- Thus, rational function Υ satisfies $\Upsilon[-\frac{1}{\theta^*}] = \Upsilon[\theta]$.
- Consequently if θ_o is stationary point of Υ, so is $-1/\theta_o^* \Rightarrow$ Stationary points are roots of a polynomial $\omega(\xi)$ in $\xi \stackrel{\text{def}}{=} \theta - 1/\theta^*$
- Idea of algorithm:
  - Compute coefficients of ω from cumulants of $\tilde{\mathbf{x}}$
  - Compute roots ξ_o of ω
  - Root $\theta^2 - \xi_o\,\theta - 1$ in order to get $(\theta_o, -1/\theta_o^*)$.

Solution for contrast $\Upsilon_{2,3}$ in $\mathbb{R}$ (1)

- Contrast $\Upsilon_{2,3}$ is defined as:

$$\Upsilon_{2,3} = \mathrm{Cum}\{z_1, z_1, z_1\}^2 + \mathrm{Cum}\{z_2, z_2, z_2\}^2 \stackrel{\mathrm{def}}{=} (\kappa_{111})^2 + (\kappa_{222})^2$$

- Yet, by *multilinearity* of cumulants:

$$\kappa_{iii} = \sum_{jk\ell} Q_i j\, Q_{ik}\, Q_{i\ell}\, \gamma_{jk\ell},\ \gamma_{jk\ell} \stackrel{\mathrm{def}}{=} \mathrm{Cum}\{\tilde{x}_j, \tilde{x}_k, \tilde{x}_\ell\}$$

- Then $\Upsilon_{2,3}$ is a degree-6 polynomial in $(\cos\beta, \sin\beta)$, or a rational function in the tangent θ:

$$\psi_3(\theta) = (\theta + \frac{1}{\theta})^{-3} \sum_{i=1}^{3} a_i\ \left(\theta^i - (-\theta)^{-i}\right)$$

Solution for contrast $\Upsilon_{2,3}$ in $\mathbb{R}$ (2)

- Denote $\xi = \theta - 1/\theta$.
Because of the invariance under transformation $\theta \to -1/\theta$, stationary points are roots of a *very simple* polynomial:

$$\omega_3(\xi) = d_2\,\xi^2 + d_1\,\xi - 4\,d_2$$

where $d_1 = a_1/3 - a_3$, and $d_2 = a_2/6$
and:

$$\begin{aligned} a_3 &= \gamma_{111}^2 + \gamma_{222}^2, \\ a_2 &= 6\,(\gamma_{122}\,\gamma_{222} - \gamma_{111}\,\gamma_{112}), \\ a_1 &= 9\,(\gamma_{122}^2 + \gamma_{112}^2) + 6\,(\gamma_{112}\,\gamma_{222} + \gamma_{111}\,\gamma_{122}) \end{aligned}$$

- **Conclusion:** solution obtainable *algebraically* from estimates of cumulants $\gamma_{jk\ell} \stackrel{\text{def}}{=} \mathrm{Cum}\{\tilde{x}_j, \tilde{x}_k, \tilde{x}_\ell\}$ [COM94b].

Another solution for contrast $\Upsilon_{2,3}$ in $\mathbb{R}$ (2)

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

$$\mathbf{u} \stackrel{\text{def}}{=} [\cos 2\beta, \;\; \sin 2\beta]^{\mathsf{T}} \tag{32}$$

and

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

with [dLdMV01]:

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

Solution for contrast $\Upsilon_{2,4}$ in $\mathbb{R}$

- Now take $\Upsilon_{2,4} \stackrel{\text{def}}{=} (\kappa_{1111})^2 + (\kappa_{2222})^2$
- This contrast is a degree-8 polynomial $(\cos\beta, \sin\beta)$.
  Denote again $\xi = \theta - 1/\theta$. Then it is a rational function in ξ:

$$\psi_4(\xi) = (\xi^2 + 4)^{-2} \sum_{i=0}^{4} b_i\, \xi^i$$

- Then its stationary points are roots of a polynomial of degree 4:

$$\omega_4(\xi) = \sum_{i=0}^{4} c_i\, \xi^i$$

  whose roots are thus obtainable *algebraically*
  (e.g. via *Ferrari*'s technique).
- Coefficients b_i and c_i are given in [COM94b] as functions of $\gamma_{ijk\ell}$

Solution for contrast $\Upsilon_{1,4}$ in $\mathbb{R}$

Same approach feasible, but easier because absence of squares
$\Rightarrow$ Here another easier-accessible approach

- **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.

Solution for contrast $\Upsilon_{1,4}$ in $\mathbb{C}$

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

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

with

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

and

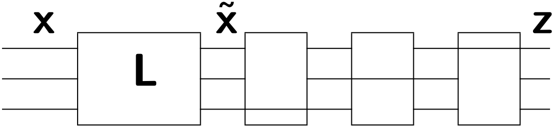
$$\mathbf{B} \;=\; \begin{pmatrix} \gamma_{1111} + \gamma_{2222} & \Re\{\delta\} & -\Im\{\delta\} \\ \Re\{\delta\} & 2\gamma_{12}^{12} + \Re\{\gamma_{22}^{11}\} & \Im\{\gamma_{22}^{11}\} \\ -\Im\{\delta\} & \Im\{\gamma_{22}^{11}\} & 2\gamma_{12}^{12} - \Re\{\gamma_{22}^{11}\} \end{pmatrix};$$

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

Conclusion: unexpectedly *entirely algebraic*! [COM01]

Jacobi Sweeping

Cyclic sweeping with fixed ordering: Example in dimension $P = 3$



Carl Jacobi, 1804-1851

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!

How do we compute tensor entries then?

Jacobi Sweeping for tensors

Sweeping a $3 \times 3 \times 3$ tensor [COM89]

$$
\begin{pmatrix} {\color{red}X} & x & x \\ x & x & x \\ x & x & . \end{pmatrix}
\begin{pmatrix} {\color{red}X} & x & x \\ x & . & x \\ x & x & x \end{pmatrix}
\begin{pmatrix} . & x & x \\ x & x & x \\ x & x & x \end{pmatrix}
$$

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

$$
\begin{pmatrix} x & x & x \\ x & x & x \\ x & x & . \end{pmatrix}
\begin{pmatrix} x & x & x \\ x & . & x \\ x & x & {\color{red}X} \end{pmatrix}
\begin{pmatrix} . & x & x \\ x & x & x \\ x & x & {\color{red}X} \end{pmatrix}
$$

$$
\left.\begin{array}{rl} {\color{red}X}: & \text{maximized} \\ x: & \text{minimized} \\ .: & \text{unchanged} \end{array}\right\} \text{by last Givens rotation}
$$

Two possible updates of **T**

After processing *every* pair, one can:

- Update based on *multilinearity*:

$$T_{ij..k} \longleftarrow \sum_{pq..r} Q_{ip}\, Q_{jq} \,..Q_{kr}\; T_{pq..r}$$

  requires an initial computation of **T**

- Update of *observations* themselves

$$\mathbf{X} \longleftarrow \mathbf{Q}\,\mathbf{X}$$

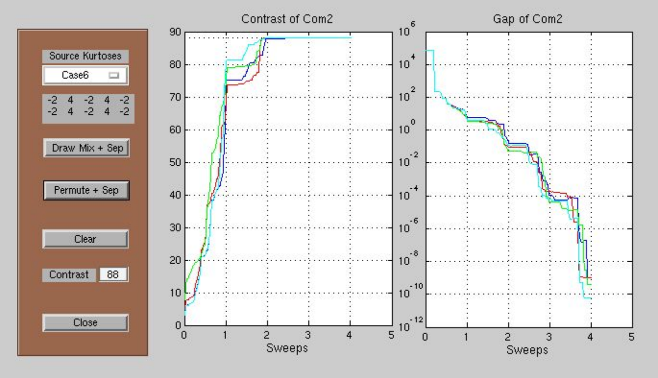
  and then

$$T_{ij..k} = \mathrm{Cum}\{x_i, x_j, ..x_k\}$$

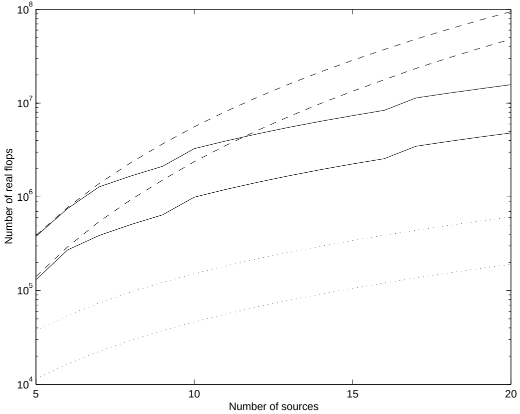
The best choice (i.e. least costly) depends on data length and dimensions.

Influence of ordering

With update based on multilinearity.



Complexity



10^8
10^7
10^6
10^5
10^4
Number of real flops
5
10
15
20
Number of sources

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 38)

Interpretation in terms of tensor diagonalization

Explanation for order 3 tensors

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

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

- *Theorem:* if $\mathbf{Q}$ is unitary, then $\Omega \stackrel{\text{def}}{=} \sum_{ijk} |G_{ijk}|^2$ is constant independent of $\mathbf{Q}$
  Proof: uses $\sum_p Q_{ip} Q_{jp} = \delta_{ij}$
- *Corollary:* Maximize $\Upsilon_{3,2} \Leftrightarrow$ minimize all non diagonal entries

Hence: “Tensor Diagonalization”

Tensor diagonalization

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

Why?

...

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

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, $r > 2$

Algorithms based on matrix slices

- JADE contrast
- JADE algorithm
- STOTD recursion on the order
- Other

Tensors as Linear Operators

Overview

- Linear Operator Ω acting on square matrices:

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

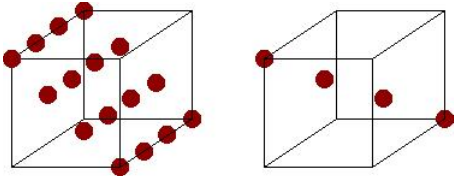
  admits eigen-matrices $\mathbf{N}(p)$, $1 \leq p \leq P^2$.

- 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

Joint Approximate Diagonalization (JAD)

Back to tensor diagonalization

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



Matrix slices diagonalization $\neq$ Tensor diagonalization

Real symmetric tensors

Definition (reminder)
G is real symmetric iff:

$$G_{ij..k} = G_{\sigma(ij..k)}$$

for all permutation σ

Two equivalent writings (order 3)

Lemma 1 Let $\mathbf{U}$ be an orthogonal real matrix, relating two 3rd order real symmetric tensors $\mathbf{G}$ and $\mathbf{g}$, then

$$\sum_{ik} G_{iik}^2 = \sum_r ||\mathbf{Diag}(\mathbf{U}^\mathsf{T}\mathbf{M}(r)\,\mathbf{U})||^2$$

where $\mathbf{M}(r)$ are symmetric matrix slices of $\mathbf{g}$: $M_{pq}(r) \stackrel{\mathrm{def}}{=} g_{pqr}$

Proof...

Theorem One can prove that $\mathcal{J} \stackrel{\mathrm{def}}{=} \sum_{ik} |G_{iik}|^2$ is a *contrast*. (at least 2 indices are equal)

Hermitian tensors

Definition

G is complex hermitian iff it is of even order, and enjoys the symmetries:

- $G_{ij..k}^{pq..r} = G_{\sigma(ij..k)}^{pq..r}$
- $G_{ij..k}^{pq..r} = G_{ij..k}^{\sigma(pq..r)}$
- $G_{ij..k}^{pq..r} = \left(G_{pq..r}^{ij..k} \right)^*$

for any permutation σ.

Two equivalent writings (order 4)

Lemma 2 Let $\mathbf{U}$ be a unitary matrix relating two complex hermitian tensors of even order 4, $\mathbf{G}$ and $\mathbf{g}$, then

$$\sum_{ik\ell} |G_{i\ell}^{ik}|^2 = \sum_{rs} ||\mathbf{Diag}(\mathbf{U}^{\mathsf{H}}\mathbf{M}(r,s)\,\mathbf{U})||^2$$

where $\mathbf{M}(r,s)$ are hermitian matrix slices of $\mathbf{g}$: $M_{pq}(r,s) \stackrel{\text{def}}{=} g_{ps}^{qr}$

Proof...

Theorem One can prove that $\mathcal{J} \stackrel{\text{def}}{=} \sum_{ik\ell..mn} |G_{i\ell..n}^{ik..m}|^2$ is a *contrast*. (only 2 indices are equal)

JADE as an approximation of $\Upsilon_{\alpha,4}$

Lemma 3 denote the EVD $\mathbf{g} = \sum_p \lambda_p \mathbf{N}(p)\mathbf{N}(p)^{\mathsf{H}}$, i.e. $g_{jkrs} = \sum_p \lambda_p N_{jk}(p)\, N_{rs}(p)$, then 3rd writing:

$$\mathcal{J}_{2,4} = \sum_{p=1}^{P^2} \lambda_p^2 \, ||\mathbf{diag}(\mathbf{U}^{\mathsf{H}}\, \mathbf{N}(p)\, \mathbf{U})||^2$$

Second approximation: Keep only the most significant eigen-matrices, $p \leq P$, which amounts to maximizing:

$$\mathcal{J}^E_{\alpha,4} \stackrel{\text{def}}{=} \sum_{p=1}^{P} \lambda_p^\alpha \, ||\mathbf{diag}(\mathbf{U}^{\mathsf{H}}\, \mathbf{N}(p)\, \mathbf{U})||^2$$

- Hence the name of Joint Approximate Diagonalization of Eigenmatrices (JADE).
- $\mathcal{J}_{\alpha,4}$ can be seen as an *approximation* of $\Upsilon_{\alpha,4}$.

Implementation of JADE with pair sweeping

Algebraic solution in dim 2

- Goal is to maximize the diagonal terms of $\mathbf{Q}^{\mathsf{H}}\mathbf{N}(r)\,\mathbf{Q}$
- Denote $\mathbf{N}(r) = \begin{pmatrix} a_r & b_r \\ c_t & d_r \end{pmatrix}$ and
  $\mathbf{Q} = \begin{pmatrix} \cos\theta & \sin\theta\, e^{\jmath\varphi} \\ -\sin\theta\, e^{-\jmath\varphi} & \cos\theta \end{pmatrix}$
- Then this amounts to maximizing w.r.t. (θ, φ): $\mathbf{v}^{\mathsf{T}}\,\Re(\mathbf{G}^{\mathsf{H}}\mathbf{G})\mathbf{v}$ where

$$\mathbf{G}^{\mathsf{H}}\mathbf{G} = \sum_r \begin{bmatrix} a_r - d_r \\ b_r + c_r \\ \jmath(c_r - b_r) \end{bmatrix}^* [a_r - d_r,\ b_r + c_r,\ \jmath(c_r - b_r)]$$

  and $\mathbf{v} = [\cos 2\theta,\ \sin 2\theta\cos\varphi,\ \sin 2\theta\sin\varphi]^{\mathsf{T}}$
- Thus, solution is the dominant eigenvector of a (real) symmetric matrix

Lower order simultaneous diagonalization (1)

Extend the idea: Slicing decreases the order

- Similarly, one can try to diagonalize a 4th order tensor $\mathbf{T} = [\gamma_{ijk\ell}]$ by jointly diagonalizing 3rd order slices $\mathbf{T}(\ell)$ (STOTD) [dLdMV01]
- Algorithm: Each Givens rotation is obtained again by maximizing a quadratic form $\mathbf{u}^{\mathsf{T}}\mathbf{B}\,\mathbf{u}$
- Noise reduction possibility: replace slices by a family of 3rd order tensors forming a basis of the map $\mathbb{C}^K \to \mathbb{C}^{K\times K\times K}$ (consider the 4th order tensor as a linear map; basis obtained by SVD)

Lower order simultaneous diagonalization (2)

In the real case, **B** is given as in slide 186 by:

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

with [dLdMV01]:

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

Diagonalization algorithms

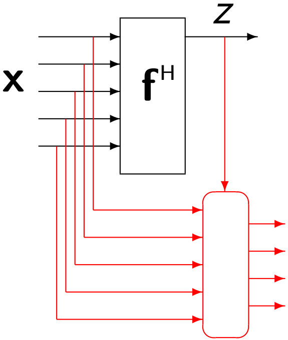
Obtain a diagonal tensor or diagonal slices:

- by orthogonal transforms [dL78] [CS93] [Com92]
- by invertible transforms [AFS07] [YER02] [?] [PHA01] [LAT06]
- by rectangular transforms [PAA99] [VO06] [COM04a] [NL06]

Algorithms based on Deflation

- Principle: Joint extraction vs Deflation
- Unitary adaptive deflation
- A so-called fixed point: FastICA
- RobustICA
- Deflation without spatial prewhitening, algebraic deflation
- Discussion on MISO criteria

Joint extraction vs Deflation



Deflation:

- Advantage: (a) reduced complexity at each stage, (b) simpler to understand
- Drawbacks: (i) accumulation of regression errors, limitation of number of extracted sources, (ii) possibly larger final complexity

Adaptive algorithms

Deflation by Kurtosis Gradient Ascent

Again same idea

After standardization, it is equivalent to maximize 4*th* order moment criterion, $\mathcal{M}_{z\,(4)} = \mathrm{E}\{|z|^4\}$, whose gradient is:

$$\nabla\mathcal{M} = 4\,\mathrm{E}\{\mathbf{x}\,(\mathbf{f}^{\mathsf{H}}\mathbf{x})(\mathbf{x}^{\mathsf{H}}\mathbf{f})^2\}$$

Overview

- Fixed step gradient on anglular parameters: [DL95]
- Locally optimal step gradient on filter taps: FastICA [HYV97]
- Globally optimal step gradient on filter taps: RobustICA [COM02a]
- Semi-Algebraic Unitary Deflation (SAUD) [COM05]

Adaptive algorithms

Adaptive implementation

- Fully adaptive solutions (update at every sample arrival) nowadays little useful
- Always easy to devise fully adaptive, or block-adaptive solutions form from block semi-algebraic algorithms (but reverse is not true!)

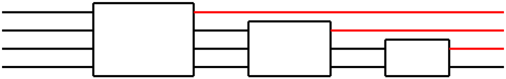
Unitary adaptive deflation (1)

- **Extraction**
  - To extract the first source, find a unitary matrix $\mathbf{U}$ so as to maximize the kurtosis of the first output
  - Matrix $\mathbf{U}$ can be iteratively determined by a sequence of Givens rotations
  - At each step, determine the best angle of the Givens rotation, e.g. by a gradient ascent [DL95]

  NB: only $P-1$ Givens rotations are involved
- **Deflation**
  - After convergence, the first output is extracted, and the $P-1$ remaining outputs of $\mathbf{U}$ can be processed in the same way

Unitary adaptive deflation (2)

At stage k, $\mathbf{Q} = \begin{pmatrix} \mathbf{q}_k^{\mathsf{H}} \\ \mathbf{Q}_k^{\mathsf{H}} \end{pmatrix}$ is unitary of size $P-k+1$, and only its first row is used to extract source k, $1 \leq k \leq P-1$



$$\begin{pmatrix} \mathbf{q}_1^{\mathsf{H}} \\ \mathbf{Q}_1^{\mathsf{H}} \end{pmatrix} \quad \begin{pmatrix} \mathbf{q}_2^{\mathsf{H}} \\ \mathbf{Q}_2^{\mathsf{H}} \end{pmatrix} \quad \begin{pmatrix} \mathbf{q}_3^{\mathsf{H}} \\ \mathbf{Q}_3^{\mathsf{H}} \end{pmatrix}$$

A so-called fixed point: FastICA (1)

- Any gradient ascent of a function $\mathcal{M}_\rho = \mathrm{E}\{\rho(\mathbf{f}^\mathsf{H}\mathbf{x})\}$ under unit-norm constraint $||\mathbf{f}||^2 = 1$ admits the Lagrangian formulation:

$$\mathrm{E}\{\mathbf{x}\dot{\rho}(\mathbf{f}^\mathsf{H}\mathbf{x})\} = \lambda\,\mathbf{f}$$

- **Convergence:** when $\nabla\mathcal{C}$ and $\mathbf{f}$ collinear (and *not* when gradient is null, because of constraint $||\mathbf{f}|| = 1$).
- **Remark:** It is *not* a fixed point algorithm, contrary to what had been claimed in [HYV97], because λ is not known!
- One can take $\rho(z) = |z|^4$

A so-called fixed point: FastICA (2)

Details of the algorithm proposed in [HYV99] in the real field; only difference compared to [TUG97] is fixed step size.

- **Gradient:** $\nabla \mathcal{M} = 4\,\mathrm{E}\{\mathbf{x}\,(\mathbf{f}^{\mathsf{T}}\mathbf{x})^3\}$
- **Hessian:** $12\,\mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\,(\mathbf{f}^{\mathsf{T}}\mathbf{x})^2\}$
- **Heavy approximation** of Hyvarinen [HYV99]:

$$\mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\,(\mathbf{f}^{\mathsf{T}}\mathbf{x})^2\} \approx \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\}\,\mathrm{E}\{(\mathbf{f}^{\mathsf{T}}\mathbf{x})^2\}$$

  - If $\mathbf{x}$ standardized and $\mathbf{f}$ unit norm, then Hessian equals Identity.
  - This yields an approximate Newton iteration: *a mere fixed step gradient!*

    $\mathbf{f} \leftarrow \mathbf{f} - \frac{1}{3}\,\mathrm{E}\{\mathbf{x}\,(\mathbf{f}^{\mathsf{T}}\mathbf{x})^3\}$ or $\mathbf{f} \leftarrow \mathrm{E}\{\mathbf{x}\,(\mathbf{f}^{\mathsf{T}}\mathbf{x})^3\} - 3\,\mathbf{f}$

    $\mathbf{f} \leftarrow \mathbf{f}/||\mathbf{f}||$

FastICA: weaknesses

This is a mere fixed step-size projected gradient algorithm, inheriting problems such as:

- Saddle points (slow/ill convergence)
- Flat areas (slow convergence)
- Local maxima (ill convergence)

NB: slow convergence may mean high complexity to reach the solution, or stopping iterations before reaching convergence (depends on stopping criterion).

Polynomial rooting

Theorem (1830). A polynomial of degree higher than 4 cannot in general be rooted algebraically in terms of a finite number of additions, subtractions, multiplications, divisions, and radicals (root extractions).

Niels Abel, 1802-1829

Evariste Galois 1811-1832

How to fix most drawbacks: RobustICA

Principle: Cheap exhaustive Line Search of a criterion $\mathcal{J}$

- Look for *absolute maximum* in the gradient direction (1-dim search)
- *Not costly* when criteria are polynomials or rational functions of low degree (same as AMiSRoF: polynomial to root, but here *at most of degree 4*)
- Applies to Kurtosis Maximization (KMA), Constant-Modulus (CMA), Constant-Power (CPA) Algorithms...

This yields corresponding Optimal-Step (OS) algorithms:
OS-KMA, OS-CMA, OS-CPA... [ZC08] [ZC05]

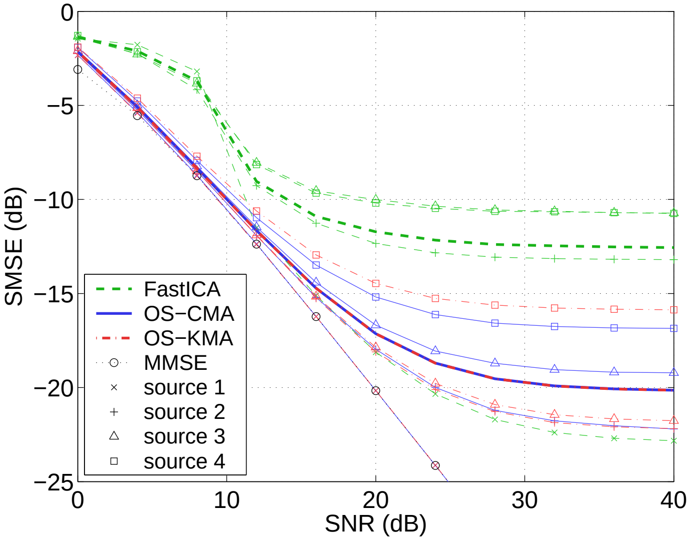
RobustICA

Algorithm

- compute coefficients of polynomial $\frac{\partial}{\partial\mu}\mathcal{J}(\mathbf{f}+\mu\nabla)$ for fixed $\mathbf{f}$ and ∇
- compute all its roots $\{\mu_i\}$
- select μ_{opt} among those roots, which yields the absolute maximum
- set $\mathbf{f} \leftarrow \mathbf{f} + \mu_{opt}\nabla$

[ZC07]

RobustICA vs FastICA



Semi-Algebraic Unitary Deflation

CoM1 [COM01]

```
Loop on sweeps
  for i = 1 to P − 1
    for j = i to P
    Algebraic 2 × 2 separ.
    end
  end
end
Extraction
```

SAUD [ACX07]

```
for i = 1 to P − 1
  Loop on sweeps
    for j = i to P
    Algebraic 2 × 2 separ.
    end
  end
  Extraction
end
```

Equivalence between KMA and CMA

- Recall the 2 criteria:

$$\Upsilon_{KMA} = \frac{\mathrm{Cum}\{z, z, z^*, z^*\}}{[\mathrm{E}\{\,|z|^2\}]^2}, \quad \mathcal{J}_{CMA} = \mathrm{E}\{[\,|z|^2 - R\,]^2\}$$

- Assume 2nd Order circular sources: $\mathrm{E}\{s^2\} = 0$
- Then KMA and CMA are equivalent
  Proof.

Discussion on Deflation (MISO) criteria

Let $z \stackrel{\text{def}}{=} \mathbf{f}^{\mathsf{H}}\mathbf{x}$. Criteria below stationary iff differentials $\dot{\mathbf{p}}$ and $\dot{\mathbf{q}}$ are collinear:

- **Ratio:** $\displaystyle \operatorname*{Max}_{\mathbf{f}} \frac{p(\mathbf{f})}{q(\mathbf{f})}$
  Example: **Kurtosis**, with $p = \mathrm{E}\{|z|^4\} - 2\mathrm{E}\{|z|^2\}^2 - |\mathrm{E}\{z^2\}|^2$ and $q = \mathrm{E}\{|z|^2\}^2$
- **Difference:** $\displaystyle \operatorname*{Min}_{\mathbf{f}} p(\mathbf{f}) - \alpha\, q(\mathbf{f})$
  Example: **Constant Modulus**, with $p = \mathrm{E}\{|z|^4\}$ and $q = 2\,a\,\mathrm{E}\{|z|^2\} - a^2$ or **Constant Power**, with $q = 2\,a\,\Re(\mathrm{E}\{z^2\}) - a^2$
- **Constrained:** $\displaystyle \operatorname*{Max}_{q(\mathbf{f})=1} p(\mathbf{f})$
  Example: **Cumulant**, with $p = \mathrm{E}\{|z|^4\} - 2\mathrm{E}\{|z|^2\}^2 - |\mathrm{E}\{z^2\}|^2$
  Example: **Moment**, with $p = \mathrm{E}\{|z|^4\}$, if standardized and with either $q = ||\mathbf{f}||^2$ or $q = \mathrm{E}\{|z|^2\}^2$

Finite alphabets

- Back to contrast criteria: APF
- Approximation of the MAP estimate
- Semi-Algebraic Blind Extraction algorithm: AMiSRoF
- Blind Extraction by ILSP
- Convolutive model
- Presence of Carrier Offset (in Digital Communications)

Contrast for discrete inputs (1)

- **Hypothesis H5** The sources take their value in a finite aphabet $\mathcal{A}$ defined by the roots in $\mathbb{C}$ of some polynomial $q(z) = 0$
- **Theorem** [COM04b]
  Under **H5**, the following is a contrast over the set $\mathcal{H}$ of invertible $P \times P$ FIR filters.

$$\Upsilon(\mathbf{G}; \mathbf{z}) \stackrel{\text{def}}{=} -\sum_n \sum_i |q(z_i[n])|^2 \tag{33}$$

APF: Algebraic Polynomial Fitting

Contrast for discrete inputs (2)

- For given alphabet $\mathcal{A}$, denote $\mathcal{G}$ the set of numbers c such that $c\,\mathcal{A} \subseteq \mathcal{A}$.
- **Lemma 1** Trivial filters satisfying **H5** are of the form:

$$\mathbf{P}\,\mathbf{D}[z]$$

with $\mathbf{D}[z]$ diagonal and $D_{pp}[z] = c_p\, z^n$, for some $n \in \mathbb{Z}$ and some $c_p \in \mathcal{G}$.

- Because $\mathcal{A}$ is finite, any $c \in \mathcal{G}$ must be of unit modulus, and we must have $c\,\mathcal{A} = \mathcal{A}, \forall c \in \mathcal{G}$.
Also any $c \in \mathcal{G}$ has an inverse c^{-1} in $\mathcal{G}$.

Contrast for discrete inputs (3)

Sketch of proof of the theorem. We prove the 3 properties of slide 68:

- $\forall \mathbf{T} \in \mathcal{T}$, $\Upsilon(\mathbf{T}; \mathbf{x}) = \Upsilon(\mathbf{I}; \mathbf{x})$
- $\forall \mathbf{G} \in \mathcal{H}$, $\forall \mathbf{s} \in \mathcal{S}$, set of independent sources in $\mathcal{A}$, $\Upsilon(\mathbf{G}; \mathbf{s}) \leq \Upsilon(\mathbf{I}; \mathbf{s})$
- $\forall \mathbf{G} \in \mathcal{H}$, $\forall \mathbf{s} \in \mathcal{S}$, equality $\Upsilon(\mathbf{G}; \mathbf{s}) = \Upsilon(\mathbf{I}; \mathbf{s}) \Rightarrow \mathbf{G}$ trivial. The proof needs the lemma
  - **Lemma 2** Let $\mathcal{A}$ be $\{a_k, 1 \leq k \leq d\} \neq \{0\}$. If $\sum_{i=1}^{L} c_i\, a_{\sigma(i)} \in \mathcal{A}$ for all mappings σ from $\{1, \ldots, L\}$ to $\{1, \ldots, d\}$, then only one $c_i \neq 0$.
  - The proof of this lemma needs sources to be *sufficiently exciting*, e.g. that all binary states are present.

Contrast for discrete inputs (4)

Idea of the proof of Lemma 2

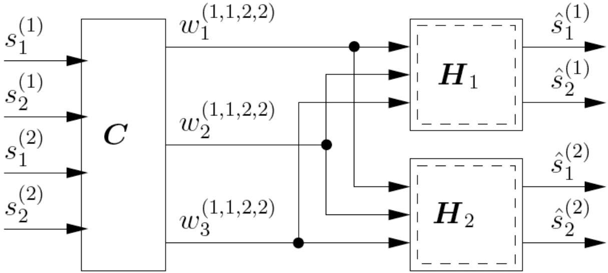
- Assume that for some $\mathbf{c} \in \mathbb{C}^L$, we have $\mathbf{x}^\mathsf{T}\mathbf{c} \in \mathcal{A}$ for all $\mathbf{x} \in \mathcal{A}^L$.
- Then $\mathbf{c}$ must be trivial:
  Non trivial vectors $\mathbf{c}$ may generate symbols that lie outside the convex
  hull of $\mathcal{A}$, or between the two closest symbols.

Contrast for discrete inputs (4)

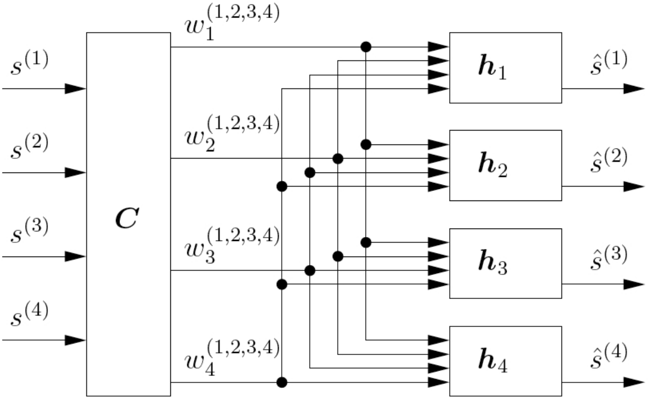
Advantages

- The previous contrast allows to separate *correlated* sources
- But it needs all sources to have the same (known) alphabet
- If sources have *different* alphabets, one can extract sources in parallel with different criteria: *Parallel Extraction* [RZC05]
- By deflation with different criteria, one can extract more sources than sensors: *Parallel Deflation* [RZC05]

Parallel Deflation

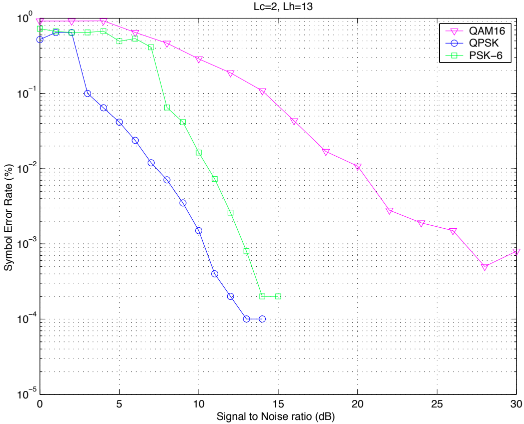


Parallel Extraction



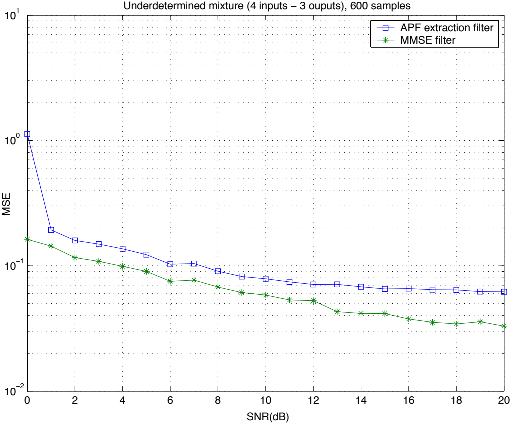
Parallel extraction

Parallel extraction of 3 sources (QPSK, QAM16, PSK6), from a 3-sensor length-3 random Gaussian channel [RZC05]



APF extraction

Parallel Deflation from a mixture of 4 sources (2 QPAK and 2 QAM16) received on 3 sensors. Extraction of a QPSK source in figure, compared to MMSE [RZC05]



MAP estimate

- Optimal solution

$$(\hat{\mathbf{H}}, \hat{\mathbf{s}})_{MAP} = Arg \operatorname*{Max}_{\mathbf{H}\,\mathbf{s}} p_{s|x,H}(\mathbf{x}, \mathbf{s}, \mathbf{H})$$

- If $s_p \in \mathcal{A}$, and if noise is Gaussian, then

$$(\hat{\mathbf{H}}, \hat{\mathbf{s}})_{MAP} = Arg \operatorname*{Min}_{\mathbf{H},\, \mathbf{s} \in \mathcal{A}^P} ||\mathbf{x} - \mathbf{H}\,\mathbf{s}||^2$$

- Less costly to search (inverse filter when it exists)

$$(\hat{\mathbf{F}}, \hat{\mathbf{s}})_{MAP} = Arg \operatorname*{Min}_{\mathbf{F},\, \mathbf{s} \in \mathcal{A}^P} ||\mathbf{F}\mathbf{x} - \mathbf{s}||^2$$

- or by deflation:

$$(\hat{\mathbf{f}}, \hat{\mathbf{s}})_{MAP} = Arg \operatorname*{Min}_{\mathbf{f},\, \mathbf{s} \in \mathcal{A}^P} ||\mathbf{f}^{\mathsf{H}}\mathbf{x} - \mathbf{s}||^2 \tag{34}$$

Approximation of the MAP estimate

For alphabet of constant modulus, MAP criterion (34) is asymptotically equivalent (for large samples of size T) to [GC98]:

$$\Upsilon_T(\mathbf{f}) = \frac{1}{T}\sum_{t=1}^{T}\prod_{j=1}^{card\mathcal{A}} |\mathbf{f}^{\mathsf{H}}\mathbf{x}[t] - a_j[t]|^2$$

where $a_j[t] \in \mathcal{A}$
We have transformed an exhaustive serach into a *polynomial* alphabet fit

Algorithm AMiSRoF

Absolute Minimimum Search by Root Finding [GC98]

- Initialize $\mathbf{f} = \mathbf{f}_o$
- For $k = 1$ to $k$max, and while $|\mu_k| >$threshold, do
  - Compute gradient $\mathbf{g}_k$ and Hessian $\mathbf{H}_k$ at $\mathbf{f}_{k-1}$
  - Compute a search direction $\mathbf{v}_k$, e.g. $\mathbf{v}_k = \mathbf{H}_k{}^{-1}\mathbf{g}_k$
  - Normalize $\mathbf{v}_k$ to $||\mathbf{v}_k|| = 1$
  - Compute the *absolute* minumum μ_k of the rational function in μ:

$$\Phi(\mu) \stackrel{\text{def}}{=} \Upsilon_T(\mathbf{f}_{k-1} + \mu\,\mathbf{v}_k)$$

  - Set $\mathbf{f}_k = \mathbf{f}_{k-1} + \mu_k\,\mathbf{v}_k$

Algorithm ILSP

Iterative Least-Squares with Projection [TVP96]

- Assumes that components $s_i[n] \in \mathcal{A}$, known alphabet
- Assumes columns of $\mathbf{H}$ belong to a known *array manifold*
- Initialize $\mathbf{H}$, and start the loop
  - Compute LS estimate of matrix $\mathbf{S}$ in equation $\mathbf{X} = \mathbf{H}\,\mathbf{S}$
  - Project $\mathbf{S}$ onto $\mathcal{A}$
  - Compute LS estimate of $\mathbf{H}$ in equation $\mathbf{X} = \mathbf{H}\,\mathbf{S}$
  - Project $\mathbf{H}$ onto the array manifold

Part V

Algorithms for convolutive mixtures

Contents

Here limited to over-determined mixtures

- Blind equalization,
  - Modeling, Carrier offset
  - Contrast criteria
  - Algorithms (Pajod, subspace, linear prediction...)
- Blind identification
  - Cumulant matching
  - Algebraic approaches
  - Subspace techniques
  - ARMA mixtures

References

Blind Equalization

- Modeling of Dynamic Mixtures
- Contrast-based
  - MISO Deflation
  - Para-unitary
- SIMO channel
  - subspace
  - mutually referenced
  - Linear prediction
- MIMO
- Matched Filter after Blind Identification

SISO Modeling

- Sequence of symbols $s[k]$ at a rate $1/T_s$
- Overall channel $h(t)$, containing transmit&receive filters and propagation
- received process $x(t) = \sum_{k\in\mathbb{Z}} h(t - k\,T_s)\,s[k]$
- If sampled at a rate $1/T$:

$$x[n] = \sum_{k\in\mathbb{Z}} h(n\,T - k\,T_s)\,s[k]$$

- If sampled exactly at symbol rate, we get a *discrete convolution*:

$$x[n] = \sum_{k\in\mathbb{Z}} h[n-k]\,s[k]$$

with $h[m] \stackrel{\text{def}}{=} h(m\,T)$

MIMO Modeling

In practice, one often assumes the approximation of discrete convolutive FIR:

$$\mathbf{x}[n] = \sum_{k=0}^{L} \mathbf{H}[k]\,\mathbf{s}[n-k] + \mathbf{v}[k] \quad (35)$$

Either:

- **Blind Identification**
  Estimate the finite matrix sequence $\mathbf{H}[k]$, or
- **Blind Equalization**
  Estimate a FIR filter $\mathbf{F}[\ell]$, $0 \leq \ell \leq L'$

Carrier offset (1)

In practical contexts of Blind Techniques, carrier frequency might be unacurately estimated

- In the SISO case, this yields

$$x[n] = \sum_k h[n-k]\, s[k]\, e^{\jmath\, k\, \delta}$$

- An *equivalent writing* is

$$x[n] = e^{\jmath\, n\, \delta} \sum_k h'[n-k]\, s[k]$$

where $h'[m] \stackrel{\text{def}}{=} h[m]\, e^{\jmath\, m\, \delta}$.

- alphabet fitting at the output may be limited by the presence of this Carrier residual. But Blind Equalization is still feasible.

Carrier offset (2)

- In the MIMO case, the carrier offset cannot be pulled into the channel anymore, unless all sources have *the same* carrier offset
- In fact on sensor k:

$$x_k[n] = \sum_\ell \sum_p H_{kp}[n-\ell]\, s_p[\ell]\, e^{\jmath\, \ell\, \delta_p}$$

or

$$x_k[n] = \sum_\ell \sum_p e^{\jmath\, n\, \delta_p}\, H'_{kp}[n-\ell]\, s_p[\ell]$$

with $H'_{kp}[m] \stackrel{\text{def}}{=} H_{kp}[m]\, e^{\jmath\, m\, \delta_p}$

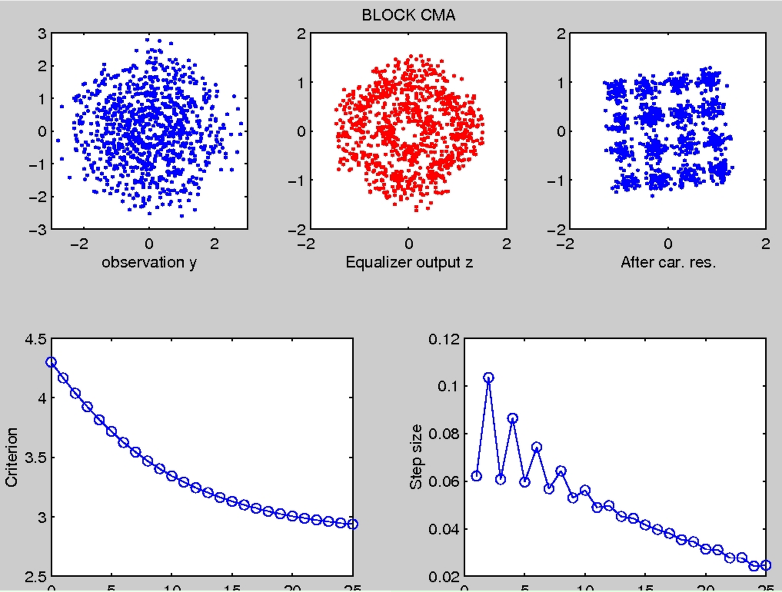
- Thus blind equalization is not possible anymore before carrier residual mitigation

Carrier offset (3)

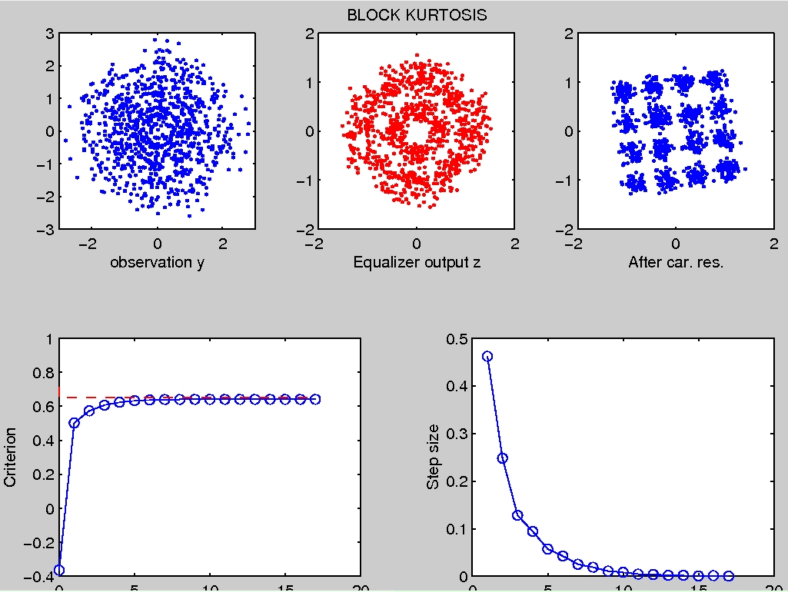
Summary

- SISO case: BE and CO can be permuted
- MIMO case: BE and CO cannot generally be permuted

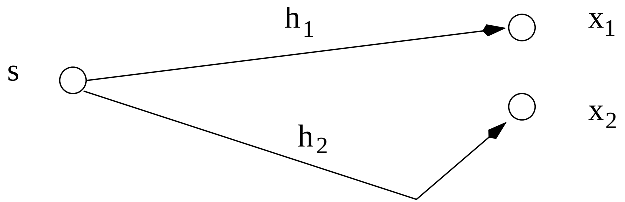
Carrier offset (3)



Carrier offset (4)



SIMO mixture with diversity $K = 2$ (1)



Disparity condition:

$$h_1[z] \wedge h_2[z] = 1 \Rightarrow x_1[z] \wedge x_2[z] = s[z]$$

Bézout:

$$\begin{aligned} \exists v_1[z],\ v_2[z] \,/ \quad v_1[z]\,h_1[z] + v_2[z]\,h_2[z] &= 1 \\ \Rightarrow \ v_1[z]\,x_1[z] + v_2[z]\,x_2[z] &= s[z] \end{aligned}$$

Thus

FIR filter $\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$ admits the FIR inverse $\mathbf{v} = \begin{pmatrix} v_1, & v_2 \end{pmatrix}$.

SIMO mixture with diversity $K = 2$ (2)

Theorem

If two polynomials $p(z) = \sum_{i=0}^{m} a_i z^i$ and $q(z) = \sum_{i=0}^{n} b_i z^i$ are prime, then the resultant below is non zero:

$$\mathcal{R}(p,q) = \begin{vmatrix} a_0 & \dots & a_m & 0 & \dots \\ 0 & \ddots & & \ddots & 0 \\ 0 & 0 & a_0 & \dots & a_m \\ b_0 & \dots & b_n & 0 & \dots \\ 0 & \ddots & & \ddots & 0 \\ 0 & 0 & b_0 & \dots & b_n \end{vmatrix} \stackrel{\text{def}}{=} \det \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}$$

Use of time diversity

Time diversity

If channel bandwidth exceeds symbol rate $\frac{1}{T_s}$ (excess bandwidth), then a sampling faster than $\frac{1}{T_s}$ brings extra information on channel. [?] [?]

How to build a SIMO channel from a SISO?

- sample twice faster: $x[k] = x(k\,T_s/2)$
- denote odd samples $x_1[k] = x[2k+1]$, and even samples $x_2[k] = x[2k]$
- then

$$\begin{pmatrix} x_1[k] \\ x_2[k] \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix} \mathbf{s}[k] \stackrel{\text{def}}{=} \mathbf{H}\,\mathbf{s}[k]$$

Matrix $\mathbf{H}$ is full rank (well conditioned) if sufficient *excess bandwidth*

Mutually Referenced Equalizers (1)

- Recall the compact modeling of equation (42) slide 279:

$$\mathbf{X}(n) = \mathcal{H}_T \, \mathbf{S}(n)$$

Then observe that if $\mathcal{H}_T$ is column shaped and full rank (here $T + L + 1$):

$$\exists \mathbf{V} \ : \quad \mathbf{V}^{\mathsf{H}} \mathcal{H}_T = \mathbf{I}$$

- Each row of $\mathbf{V}^{\mathsf{H}}$ defines an equalizer $\mathbf{v}_i^{\mathsf{H}}$, deduced from each other by a delay [?]:

$$\mathbf{v}_j^{\mathsf{H}} \, \mathbf{X}(n-i) = \mathbf{v}_i^{\mathsf{H}} \, \mathbf{X}(n-j) = s(n-i-j)$$

Mutually Referenced Equalizers (2)

- The equations $\mathrm{E}\{|\mathbf{v}_k^{\mathsf{H}}\mathbf{X}(n) - \mathbf{v}_{k+1}^{\mathsf{H}}\mathbf{X}(n+1)|^2\} = 0$ for $0 \leq k \leq T+L$ yield:

$$\mathcal{V}^{\mathsf{H}}\,\mathcal{R}\,\mathcal{V} = 0 \quad \text{with}$$

$$\mathcal{V} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{T+L} \end{pmatrix}, \text{ and } \mathcal{R} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{R}(0) & -\mathbf{R}(1)^{\mathsf{H}} & 0 & \dots \\ -\mathbf{R}(1) & 2\,\mathbf{R}(0) & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & 2\,\mathbf{R}(0) & -\mathbf{R} \\ \vdots & 0 & \ddots & -\mathbf{R}(1) & \mathbf{R} \end{pmatrix}$$

and $\mathbf{R}(k) \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{X}(n+k)\,\mathbf{X}(n)^{\mathsf{H}}\}$.

- Thus, take $\mathcal{V}$ as being the dominated eigenvector, and extract $\mathbf{v}_k$ from it
- In practice, necessary to add a constraint to avoid $\mathbf{v}_k \in \mathrm{null}(\mathcal{H}^{\mathsf{H}})$

Contrast criteria (1)

Proofs derived in the static case hold true in the convolutive case, e.g. family of contrasts of slides 73-74

Proofs...

Contrast criteria (2)

- But also possible to devise new families of contrasts for para-unitary equalizers after prewhitening [?] [?]. For instance:

$$\Upsilon(\mathbf{y}) = \sum_{i} \sum_{j\,p} \sum_{k\,q} |\mathrm{Cum}\{y_i[n],\, y_i[n],\, y_j[n-p],\, y_k[n-q]\}|^2 \tag{36}$$

- In the above, one can conjugate any of the variables y_ℓ's
- Holds true for almost any cumulants of order ≥ 3
- Only two indices need to be identical with same delay

Proof Based on the property that, for para-unitary filters $\mathbf{G}$:

$$\mathbf{y}[n] \stackrel{\text{def}}{=} \sum_{t} \mathbf{G}[n-t]\, s[t] \;\Rightarrow\; \Upsilon(\mathbf{y}) = \sum_{i} \sum_{\ell\, t} |G_{i\ell}[t]|^4\, |\kappa_\ell|^2$$

MISO Dynamic Extractor: Deflation

- Fixed step gradient Deflation [TUG97]
- Optimal Line search along a descent direction, OS-KMA [?] [ZC05]

PAJOD (1)

- Technique applied *after* space-time prewhitening
- Then one looks for a *para-unitary* equalizer, by maximizing the contrast

$$\mathcal{J}_{2,r}(\mathbf{y}) = \sum_{\mathbf{b}} \sum_{\boldsymbol{\beta}} ||\mathbf{Diag}\{\mathcal{H}^{\mathsf{H}}\, \mathbf{M}(\mathbf{b}, \boldsymbol{\beta})\, \mathcal{H}\}||^2$$

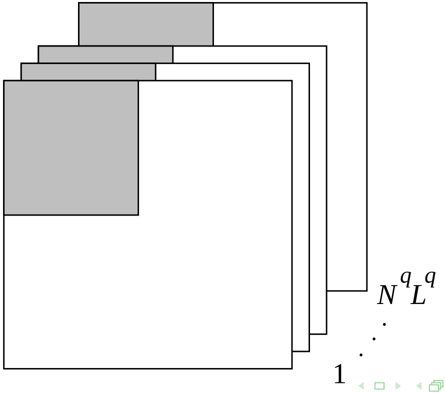
  Matrix $\mathcal{H}$ is now defined differently, and is semi-unitary. Matrices $\mathbf{M}(\cdot)$ contain cumulants of whitened observations

- Contrast (36) is maximized again by a sweeping technique

PAJOD (2)

PAJOD: Partial Approximate Joint Diagonalization of matrix slices

One actually attempts to diagonalize only a portion of the tensor



MIMO Blind Equalization

- linear prediction after BI [?]
- linear prediction [?] [?] [?]
- subspace [?] [?] [?] [?] [?]
- identifiability issues by subspace techniques [?] [?]

Equalization after prior Blind Identification

Assume channel $\mathbf{H}[z]$ has been identified, with:
$\mathbf{x}[n] = \mathbf{H}[z] \star \mathbf{s}[z] + \mathbf{v}[z]$
An estimate of $\mathbf{s}[z]$ is obtained with $\mathbf{F}[z] \star \mathbf{x}[z]$.
Possible equalizers $\mathbf{F}[z]$:

- Zero-Forcing: $\mathbf{F}[z] = \mathbf{H}[z]^{-1}$
- Matched Filter: $\mathbf{F}[z] = \mathbf{H}[1/z^*]^{\mathsf{H}}$
  (used in MLSE; optimal if channel AWGN; maximizes output SNR)
- Minimum Mean Square Error (MSE):
  $\mathbf{F}[z] = (\mathbf{H}[z]\mathbf{H}[1/z^*]^{\mathsf{H}} + \mathbf{R}_v[z])^{-1}\mathbf{H}[1/z^*]^{\mathsf{H}}$

➠ One can insert soft or hard decision to stabilize the inverse, or to reduce noise, e.g. decision Feedback Equalizers (DFE).

Overview

- Interest
- MA identifiability (second order vs hOS)
- SISO: Cumulant matching
- MIMO: Cumulant matching and linear prediction (non monic MA)
- Algebraic approaches, Quotient Ring
- SIMO: Subspace approaches
- MIMO: Subspace, IIR, ...

BE vs BI

If sources $s_p[k]$ are discrete, it is:

- rather easy to define a BE optimization criterion in order to match an output alphabet
- difficult to exploit a source alphabet in BI

Example the property of constant modulus of an alphabet is mainly used in Blind Equalization: CMA (Constant Modulus Algorithm)

Interest of Blind Identification

- When the mixture does not have a stable inverse
  ➽ When may want to control stability by soft/hard decision in a Feedback Equalizer
- When sources are not of interest (e.g. channel characteristics, localization only)

SISO Cumulant matching (1)

- Consider first the **SISO** version of (35):

$$x[n] = \sum_{k=0}^{L} h[k]\, s[n-k] + v[k]$$

where $v[k]$ is Gaussian stationary, and $s[n]$ is 4th order white stationary.

- Then, by the multilinearity property of cumulants (slide 52):

$$C_x(i,j) \stackrel{\text{def}}{=} \mathrm{Cum}\{x[t+i], x[t+j], x[t+L], x[t]\} = h[i]\, h[j]\, h[L]\, h[0]\, c_s$$

with $c_s \stackrel{\text{def}}{=} \mathrm{Cum}\{s[n], s[n], s[n], s[n]\}$.

- By substitution of the unknown $h[L]\, h[0]\, c_s$, one gets a whole family of equations [?] [?]:

$$h[i]\, h[j]\, C_x(k,\ell) = h[k]\, h[\ell]\, C_x(i,j),\ \forall i,j,k,\ell \qquad (37)$$

SISO Cumulant matching (2)

- A solution to the subset of (37) for which $j = \ell$ can be easily obtained:

$$h[i]\,C_x(k,j) - h[k]\,C_x(i,j) = 0,\ \ 0 \le i < k \le L,\ 0 \le j \le L \tag{38}$$

- This is a *linear system* of $L(L+1)^2/2$ equations in $L+1$ unknowns
$\Rightarrow$ **Least Square (LS) solution**, up to a scale factor (e.g. $h(0) = 1$).
- Since 4th order only, asymptotically (for large samples) insensitive to Gaussian noise.
- Total Least Squares (TLS) solution possible as well

MIMO Cumulant matching (1)

Inteterminacy

- Scale (scalar) factor for SISO, but $\mathbf{\Lambda P}$ factor for MIMO

Reduction to a monic model [COM94b]

- If $\mathbf{H}[0]$ is invertible, (35) can be rewritten as

$$\mathbf{y}[n] = \mathbf{H}[0]\,\mathbf{s}[n], \tag{39}$$

$$\mathbf{x}[n] = \sum_{k=0}^{L} \mathbf{B}[k]\,\mathbf{y}[n-k] + \mathbf{w}[k] \tag{40}$$

where $\mathbf{B}[k] \stackrel{\text{def}}{=} \mathbf{H}[k]\mathbf{H}[0]^{-1}$.

- Because $\mathbf{B}[0] = \mathbf{I}$, MA model (40) is said to be *monic*.
- *Indeterminacy* is only in (39), which is solved by ICA if $\mathbf{s}[n]$ is spatially white

MIMO Cumulant matching (2)

Kronecker notation

- Store 4th order cumulant tensors in vector form:

$$\mathbf{c}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} \stackrel{\text{def}}{=} \mathbf{vec}\{\mathrm{Cum}\{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\}\}$$

- Then, we have the property (where $*$ denotes term-wise product):

$$\mathbf{c}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} = \mathrm{E}\{\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{d}\} - \mathrm{E}\{\mathbf{a}\otimes\mathbf{b}\}\otimes\mathrm{E}\{\mathbf{c}\otimes\mathbf{d}\} - \mathrm{E}\{\mathbf{a}\otimes\mathrm{E}\{\mathbf{b}$$
$$-\mathrm{E}\{\mathbf{a}\otimes\mathbf{1}_\beta\otimes\mathbf{c}\otimes\mathbf{1}_\delta\} * \mathrm{E}\{\mathbf{1}_\alpha\otimes\mathbf{b}\otimes\mathbf{1}_\gamma\otimes\mathbf{d}\}$$

MIMO Cumulant matching (3)

Assume monic MA model (40) where $\mathbf{s}[n]$ white in time and L fixed, and denote

$$\mathbf{c_x}(i,j) \stackrel{\text{def}}{=} \mathbf{vec}\{\mathrm{Cum}\{\mathbf{x}[t+i], \mathbf{x}[t+j], \mathbf{x}[t+L], \mathbf{x}[t]\}\}$$

- Then we can prove [**?**]:

$$C_{\mathbf{x}}(i,j) = C_{\mathbf{x}}(0,j)\,\mathbf{B}[i]^{\mathsf{T}}, \;\; \forall j,\; 0 \le j \le L$$

  where $C_{\mathbf{x}}(i,j) \stackrel{\text{def}}{=} \mathbf{Unvec}_P\,(\mathbf{c_x})$ is $P^3 \times P$
- For every fixed i, $\mathbf{B}[i]$ is obtained by solving the system of $(L+1)P^4$ equations in P^2 unknowns in LS sense:

$$[\mathbf{I}_P \otimes C_{\mathbf{x}}(0,j)]\,\mathbf{vec}\{\mathbf{B}[i]\} = \mathbf{c_x}(i,j) \qquad (41)$$

MIMO Cumulant matching (4)

- **Summary of the algorithm**
  - Choose a maximum L
  - Estimate cumulants of observation, $\mathbf{c}_{\mathbf{x}}(i,j)$ for $i,j \in \{0,\dots,L\}$
  - Solve the $(L+1)$ systems (41) in $\mathbf{B}[i]$
  - Compute the residue $\mathbf{y}[t]$ *(Linear Prediction)*
  - Solve the ICA problem $\mathbf{y}[t] = \mathbf{H}[0]\,\mathbf{s}[t]$
- Weaknesses
  - $\mathbf{H}[0]$ must be invertible
  - FIR model (35) needs to have a stable inverse

Algebraic Blind identification (1)

Types of discrete source studied

- BPSK: $b[k] \in \{-1, 1\}$, $i.i.d.$
- MSK: $m[k+1] = \jmath\, m[k]\, b[k]$
- QPSK: $p[k] \in \{-1, -\jmath, 1, \jmath\}$, $i.i.d.$
- $\frac{\pi}{4}$-DQPSK: $d[k+1] = e^{\jmath\pi/4}\, d[k]\, p[k]$
- 8-PSK: $q[k] \in \{e^{\jmath n\pi/4},\ n \in \mathbb{Z}\}$, i.i.d.
- etc...

Algebraic Blind identification (2)

Input/Output relations:

- For $s[k]$ **BPSK**: $\mathrm{E}\{x[n]\,x[n-\ell]\} = s[0]^2 \sum_{m=0}^{L} h[m]\,h[m+\ell]$
- For $s[k]$ **MSK**:
  $\mathrm{E}\{x[n]\,x[n-\ell]\} = s[0]^2 \sum_{m=0}^{L} (-1)^m\,h[m]\,h[m+\ell]$
- For $s[k]$ **QPSK**:
  $\mathrm{E}\{x[n]^2\,x[n-\ell]^2\} = s[0]^2 \sum_{m=0}^{L} h[m]^2\,h[m+\ell]^2$
- etc..

Algebraic Blind identification (3)

Principle:

- Compute all roots of the polynomial system in $h[n]$.
  For instance for MSK sources and a channel of length 2 [**?**]:

$$
\begin{aligned}
h[0]^2 - h[1]^2 + h[2] &= \beta_0 \\
h[0]\, h[1] - h[1]\, h[2] &= \beta_1 \\
h[0]\, h[2] &= \beta_2
\end{aligned}
$$

- Choose among these roots the one that best matches the I/O correlation:

$$
\mathrm{E}\{x[n]\, x[n-\ell]^*\} = \sum_{m=0}^{L} h[m]\, h[m+\ell]^*
$$

Algebraic Blind identification (4)

Theorem (Bezout) A polynomial system of degree d in N variables has either:

- infinitely many solutions
- no solution
- exactly d^N solutions (distinct or not)

Etienne Bézout, 1730-1783

Algebraic Blind identification (5)

- **Standard approaches**
  - Gröbner bases
- **Efficient solution of polynomial system**: Normal Forms
  There are two approaches, both working in the Quotient Ring modulo the Ideal defined by polynomial system:
  - Eigenvectors of the transposed multiplication matrix $\mathbf{M}_u^\mathsf{T}$ in the Quotient Ring
  - The Rational Univariate Representation (RUR) of eigenvalues of $\mathbf{M}_u$

  Main advantage: most (symbolic) calculations depend *only* on distribution of $s[n]$, and may thus be *stored in ROM* $\longrightarrow$ Limited numerical computations left depending on measurements.

SIMO mixture (1)

- **FIR of length L and dimension K:**

$$\mathbf{x}(n) = \sum_{i=1}^{L} \mathbf{h}(i) s(n-i) + \mathbf{b}(n)$$

with: $\mathrm{E}\{\mathbf{b}(m)\,\mathbf{b}(n)^{\mathsf{H}}\} = \sigma_b^2\,\mathbf{I}\,\delta(m-n)$

and $\mathrm{E}\{\mathbf{b}(m)\,s(n)^*\} = \mathbf{0}$

- **For T successive values:**

$$\begin{pmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-T) \end{pmatrix} = \begin{pmatrix} \mathbf{h}(0) & \mathbf{h}(1) & \dots & \mathbf{h}(L) & 0 & \dots & 0 \\ 0 & \mathbf{h}(0) & \dots & \dots & \mathbf{h}(L) & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{h}(0) & \mathbf{h}(1) & \dots & \mathbf{h}(L) \end{pmatrix} \begin{pmatrix} s(n) \\ \vdots \\ s(n-T-L \end{pmatrix}$$

Or in compact form:

$$\mathbf{X}(n:n-T) = \mathcal{H}_T\,\mathbf{S}(n:n-T-L) \qquad (42)$$

Here, $\mathcal{H}_T$ is of size $(T+1)K \times (T+L+1)$

SIMO mixture (2)

- **Condition of "column" matrix**
  $\mathcal{H}$ has strictly more rows than columns **iff**

$$(T+1)K > T+L+1$$

$$\Leftrightarrow T > L/(K-1) - 1 \Leftarrow T \geq L$$

  It suffices that T exceeds channel memory.

- **Disparity condition**
  Columns of $\mathcal{H}$ are linearly independent **iff**

$$\mathbf{h}[z] \neq \mathbf{0}, \; \forall z$$

- **Noise subspace**
  Under these 2 conditions, there exists a "noise subspace":

$$\exists \mathbf{v} \, / \quad \mathbf{v}^{\mathsf{H}} \mathcal{H}_T = \mathbf{0}$$

SIMO mixture (3)

Properties of vectors in the null space

- Since $\mathbf{R}_x \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{X}\,\mathbf{X}^{\mathsf{H}}\} = \mathcal{H}_T\,\mathcal{H}_T^{\mathsf{H}} + \sigma_b^2\,\mathbf{I}$,
  vectors $\mathbf{v}^{(p)}$ of noise space can be computed from $\mathbf{R}_x$:

$$\mathbf{R}_x\,\mathbf{v}^{(p)} = \sigma_b^2\,\mathbf{v}^{(p)}$$

- And since convolution is commutative:

$$\mathbf{v}^{(p)\mathsf{H}}\,\mathcal{H}_T = \mathbf{h}^{\mathsf{H}}\,\mathcal{V}^{(p)}$$

  where $\mathcal{V}^{(p)}$ block Töplitz, built on $\mathbf{v}^{(p)}$.

- Thus $\mathbf{h}^{\mathsf{H}} = [\mathbf{h}(0)^{\mathsf{H}},\ \mathbf{h}(1)^{\mathsf{H}},\ \dots\mathbf{h}(L)^{\mathsf{H}}]$ are obtained by computing the left singular vector common to $\mathcal{V}^{(p)}$.

SIMO mixture (4)

Summary of the SIMO Subspace Algorithm

- Choose $T \geq L$
- Compute $\mathbf{R}_x$, correlation matrix of size $(T+1)K$
- Compute the $d = T(K-1) + K - L - 1$ vectors $\mathbf{v}^{(p)}$ of the noise space
- Compute vector $\mathbf{h}$ minimizing the quadratic form

$$\mathbf{h}^{\mathsf{H}} \left[\sum_{p=1}^{d} \mathcal{V}^{(p)} \mathcal{V}^{(p)\mathsf{H}} \right] \mathbf{h}$$

- Cut $\mathbf{h}$ into $L+1$ slices $\mathbf{h}(i)$ of length K

Under the assumed hypotheses, the solution is unique up to a scalar scale factor [**?**]

SIMO mixture (5)

Summary of the SIMO Subspace Algorithm when $K = 2$

- Choose $T = L$. There is a single vector $\mathbf{v}$ in the noise space
- Compute $\mathbf{R}_x$, correlation matrix of size $(T+1)K$
- Compute the vector $\mathbf{v}$ of the noise space
- Cut $\mathbf{v}$ into $L+1$ slices $\mathbf{v}(i)$ of length $K = 2$
- Compute $\mathbf{h}(i) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{v}(i)$
  In fact $x_i = h_i \star s \Rightarrow h_2 \star x_1 - h_1 \star x_2 = 0$
  Approach called **SRM** *(Subchannel Response Matching)* [**?**] [**?**]

SISO Identifiability

- Second order statistics
  - $\alpha_\ell = \mathrm{E}\{x[n]\,x[n-\ell]^*\}$ ➽ allow to estimate $|h[m]|$
  - $\beta_\ell = \mathrm{E}\{x[n]\,x[n-\ell]\}$ ➽ allow to estimate $h[m]$ if $\mathrm{E}\{s^2\} \neq 0$
- Fourth order statistics ➽ many (polynomial) additional equations
  - $\gamma_{0jk\ell} = \mathrm{Cum}\{x[n], x[n-j], x[n-k], x[n-\ell]\}$
  - $\gamma_{0j}^{k\ell} = \mathrm{Cum}\{x[n], x[n-j], x[n-k]^*, x[n-\ell]^*\}$

If some sources are 2nd order circular, sample Statistics of order higher than 2 are mandatory, but *otherwise not* [?] !

SIMO Identifiability

With a receive diversity, (deterministic) identifiability conditions are weaker [?] [?]

- **Definition** A length-N input sequence $s[n]$ has P modes iff the Hankel matrix below is full row rank:

$$\begin{pmatrix} s[1] & s[2] & \dots & s[N-p+1]] \\ s[2] & s[3] & \ddots & s[N-p+2] \\ \vdots & \vdots & & \vdots \\ s[p] & s[p+1] & \dots & s[N] \end{pmatrix}$$

- **Theorem** A $K \times L$ FIR channel $\mathbf{h}$ is identifiable if:
  - Channels $h_k[z]$ do not have common zeros
  - The observation length of each $x_k[n]$ must be at least $L+1$
  - The input sequence should have at least $L+1$ modes (sufficiently exciting)

Subspace algorithm for MIMO mixtures

- Similarly to the SIMO case, we have the compact form:

$$\mathbf{X}(n) = \mathcal{H}_T\,\mathbf{S}(n) + \mathbf{B}(n)$$

where $\mathcal{H}$ is now built on matrices $\mathbf{H}(k), 1 \leq k \leq L$, and is of size $(T+1)K \times (T+L+1)P$.

- For large enough T, this matrix is "column shaped"
- Again $\mathbf{R}_x = \mathcal{H}_T\,\mathcal{H}_T^{\mathsf{H}} + \sigma_b^2\,\mathbf{I}$
- But now, vectors of the noise space caracterize $\mathbf{H}[z]$ only up to a constant post-multiplicative matrix $\Rightarrow$ *ICA must be used afterwards*
- Foundations of the MIMO subspace algorithm are more complicated [Loubaton'99]

In the MIMO case, HOS are in general mandatory.

SISO ARMA mixtures

What are the tools when the channel is IIR?

- In general, just consider it as a FIR (truncation) $\rightarrow$ already seen
- But also possible to assume presence of a recursive part
  - Define I/O relation: $\sum_{i=0}^{p} a_i\, x[n-i] = \sum_{j=0}^{q} b_j\, w[n-i]$ where $w[\cdot]$ is i.i.d. and $a_0 = b_0 = 1$
  - Second order $c_x(\tau) \stackrel{\text{def}}{=} \mathrm{E}\{x[n]\, x[n+\tau]\}$ can be used to identify a_k:
  $$\sum_{k=0}^{p} a_k\, c_x(\tau - k) = 0, \quad \forall \tau > q$$
  - Then compute the residue and identify b_ℓ with HOS (cf. slide 269)
- Also possible with HOS only for AR part [**?**]

MIMO ARMA mixtures (1)

Results of SISO case can be extended

- Take a K-dimensional ARMA model: Define I/O relation:

$$\sum_{i=0}^{p} \mathbf{A}_i \, \mathbf{x}[n-i] = \sum_{j=0}^{q} \mathbf{B}_j \, \mathbf{w}[n-i]$$

  where $w[\cdot]$ is i.i.d. and $\mathbf{A}_0 = \mathbf{I}$ and $\mathbf{B}_0$ inveritible

- For instance at order 4, AR identification is based on:
  - $\sum_{j=1}^{p} \mathbf{A}_j \, \bar{\mathbf{c}}_x(t, \tau - j) = -\bar{\mathbf{c}}_x(t, \tau)$, $\forall \tau > q$, $\forall t$
  - with $\bar{\mathbf{c}}_x(i, j) \stackrel{\text{def}}{=} \mathbf{Unvec}_K\left(\mathrm{Cum}\{\mathbf{x}[n], \mathbf{x}[n], \mathbf{x}[n+i], \mathbf{x}[n+j]\}\right)$

MIMO ARMA mixtures (2)

Limitations

- Sources need to be *linear processes*
- $\mathbf{B}_0$ needs to be invertible
- AR residuals need to be computed (MA filtering) to compute $\mathbf{B}_i$
- One can compute MA residuals (AR filtering) if input $\mathbf{s}[n]$ is requested $\longrightarrow$ but might be unstable

Part VI

Algorithms for under-determined mixtures

Back to Essential uniqueness

- Recall the general model (22) to fit (here 3rd order):

$$\varepsilon \stackrel{\text{def}}{=} ||\mathbf{T} - \sum_{q=1}^{\text{rank}\{\mathbf{T}\}} \mathbf{a}^{(q)} \circ \mathbf{b}^{(q)} \circ \mathbf{c}^{(q)}||^2 \qquad (43)$$

- For instance, if $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is solution, so is $(\mathbf{A}\mathbf{P}\boldsymbol{\Lambda}, \mathbf{B}\mathbf{P}\boldsymbol{\Delta}, \mathbf{C}\mathbf{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Delta}^{-1})$
- *Essential uniqueness*: uniqueness up to a common scale-permuation ambiguity.
- The scale indetermination can be fixed by introducing a diagonal tensor $\boldsymbol{\Delta}$ and imposing unit-norm columns in the matrices:

$$\mathbf{T} \approx \boldsymbol{\Delta} \underset{1}{\bullet} \mathbf{A} \underset{2}{\bullet} \mathbf{B} \underset{3}{\bullet} \mathbf{C}$$

Essential uniqueness

Sufficient condition The *Kruskal rank* of a matrix $\mathbf{A}$ is the maximum number k_A, such that any subset of k_A columns are linearly independent.

Kruskal's bound [KRU77] [SB00] [SS07] gives *sufficient conditions*. Essential uniqueness is ensured if the tensor rank R is below an upper bound:

- $2R + 2 \leq k_A + k_B + k_C$,
- or generically, for a tensor of order d and dimensions N_ℓ:

$$2R + d - 1 \leq \sum_{\ell=1}^{d} \min(N_\ell, R)$$

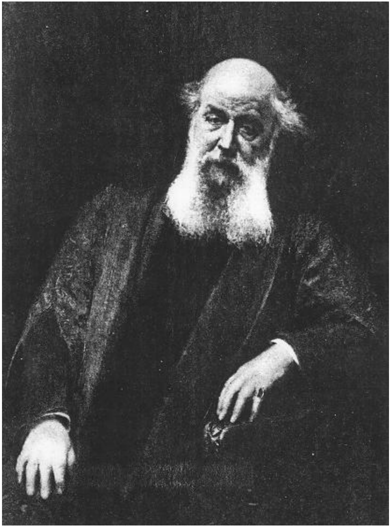
Essential uniqueness (cont'd)

Necessary and sufficient condition Essential uniqueness has been proved via *local identifiability*, under the condition that the rank is *sub-generic*:

$$\text{rank}\{\mathbf{T}\} \leq \left\lceil \frac{\prod_\ell N_\ell}{\sum_\ell (N_\ell - 1) + 1} \right\rceil$$

This condition is necessary and sufficient up to some exceptions, for which the maximal rank should be decreased by 1. The proof is numerical for the general case [CtB08], but algebraic in the symmetric case [CGLM08].

Questions: What algorithms, and under what conditions?



James Joseph Sylvester **(1814–1897)**

Binary case

Construction of the CanD (1)
Sylvester's theorem in $\mathbb{R}$

- A binary quantic $p(x,y) = \sum_{i=0}^{d} \gamma_i\, c(i)\, x^i\, y^{d-i}$ can be decomposed in $\mathbb{R}[x,y]$ into a sum of r powers as $p(x,y) = \sum_{j=1}^{r} \lambda_j\,(\alpha_j\, x + \beta_j\, y)^d$ if and only if the form

$$q_c(x,y) = \prod_{j=1}^{r}(\beta_j\, x - \alpha_j\, y) = \sum_{l=0}^{r} g_l\, x^l\, y^{r-l}$$

$$\text{satisfies} \quad \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_r \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{r+1} \\ \vdots & & & \vdots \\ \gamma_{d-r} & & \cdots & \gamma_d \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_r \end{bmatrix} = 0.$$

  and has distinct real roots.
- Valid even in non generic cases.

Construction of the CanD (2)

Sylvester's theorem in $\mathbb{C}$

A binary quantic $p(x,y) = \sum_{i=0}^{d} c(i)\,\gamma_i\,x^i\,y^{d-i}$ can be written as a sum of dth powers of r distinct linear forms:

$$p(x,y) = \sum_{j=1}^{r} \lambda_j\,(\alpha_j\,x + \beta_j\,y)^d, \tag{44}$$

if and only if **(i)** there exists a vector $\mathbf{g}$ of dimension $r+1$, with components g_ℓ, such that

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_r \\ \vdots & & & \vdots \\ \gamma_{d-r} & \cdots & \gamma_{d-1} & \gamma_d \end{bmatrix} \mathbf{g}^* = \mathbf{0}. \tag{45}$$

and **(ii)** the polynomial $q(x,y) \stackrel{\text{def}}{=} \sum_{\ell=0}^{r} g_\ell\,x^\ell\,y^{r-\ell}$ admits r distinct roots

Algorithm

- Start with $r = 1$ ($d \times 2$ matrix) and increase r until it looses its column rank

1	2
2	3
3	4
4	5
5	6
6	7
7	8

$\longrightarrow$

1	2	3
2	3	4
3	4	5
4	5	6
5	6	7
6	7	8

$\longrightarrow$

1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7
5	6	7	8

Symmetric tensors of larger dimension

We have seen the link with polynomials in slide 124. The idea is to extend Sylvester's algorithm to more than 2 variables.

- Xx
- Xx

xx Tsigaridas Mourrain Comon

Iterative algorithms

Continue to keep 3rd order case to illustrate the reasoning. Define

$$\mathbf{p} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{vec}\{\mathbf{A}^{\mathsf{T}}\} \\ \mathbf{vec}\{\mathbf{B}^{\mathsf{T}}\} \\ \mathbf{vec}\{\mathbf{C}^{\mathsf{T}}\} \end{bmatrix}, \quad \text{and the gradient } \mathbf{g} = \begin{bmatrix} \mathbf{g}_{\mathbf{A}} \\ \mathbf{g}_{\mathbf{B}} \\ \mathbf{g}_{\mathbf{C}} \end{bmatrix}$$

Gradient (1)

- Newton update rule: $\mathbf{p}(k+1) = \mathbf{p}(k) - \mathbf{H}(k)^{-1}\,\mathbf{g}(k)$
  Pure gradient: $\mathbf{p}(k+1) = \mathbf{p}(k) - \mu(k)\,\mathbf{g}(k)$
- Systematic step variation:
  - $\mu(k)$ constant if $\varepsilon(k) - \varepsilon(k+1) > 0.005\,\varepsilon(k)$
  - μ inreased via $\mu(k+1) = 1.1\,\mu(k)$ if $0 \leq \varepsilon(k) - \varepsilon(k+1) \leq 0.005\,\varepsilon(k)$
  - μ decreased via $\mu(k+1) = \mu(k)/2$ if $\varepsilon(k) < \varepsilon(k+1)$

Gradient (2)

Closed-form expressions of the gradients of ε (43):

$$\begin{aligned}
\mathbf{g_A} &= [\mathbf{I}_A \otimes (\mathbf{C}^{\mathsf{H}}\mathbf{C} \boxdot \mathbf{B}^{\mathsf{H}}\mathbf{B})]\mathbf{vec}\{\mathbf{A}^{\mathsf{T}}\} - [\mathbf{I}_A \otimes (\mathbf{C} \odot \mathbf{B})]\mathbf{vec}\{\mathbf{T}_{KJ\times I}\} \\
\mathbf{g_B} &= [\mathbf{I}_B \otimes (\mathbf{A}^{\mathsf{H}}\mathbf{A} \boxdot \mathbf{C}^{\mathsf{H}}\mathbf{C})]\mathbf{vec}\{\mathbf{B}^{\mathsf{T}}\} - [\mathbf{I}_A \otimes (\mathbf{A} \odot \mathbf{C})]\mathbf{vec}\{\mathbf{T}_{IK\times J}\} \\
\mathbf{g_C} &= [\mathbf{I}_C \otimes (\mathbf{B}^{\mathsf{H}}\mathbf{B} \boxdot \mathbf{A}^{\mathsf{H}}\mathbf{A})]\mathbf{vec}\{\mathbf{C}^{\mathsf{T}}\} - [\mathbf{I}_C \otimes (\mathbf{B} \odot \mathbf{A})]\mathbf{vec}\{\mathbf{T}_{JI\times K}\}
\end{aligned}$$

where $\boxdot$ denotes the elementwise product (Hadamard)

Quasi-Newton (1)

Define Jacobians with respect to matrices **A**, **B** and **C**, and the joint Jacobian:

$$\begin{aligned} \mathbf{J}_{\mathbf{A}} &= \mathbf{I}_A \otimes (\mathbf{C} \odot \mathbf{B}) \\ \mathbf{J}_{\mathbf{B}} &= \Pi_1 \left[\mathbf{I}_B \otimes (\mathbf{A} \odot \mathbf{C})\right] \\ \mathbf{J}_{\mathbf{C}} &= \Pi_2 \left[\mathbf{I}_C \otimes (\mathbf{B} \odot \mathbf{A})\right] \end{aligned}$$

where Π_i are appropriately chosen permutations, and

$$\mathbf{J} = [\mathbf{J}_{\mathbf{A}},\ \mathbf{J}_{\mathbf{B}},\ \mathbf{J}_{\mathbf{C}}]$$

Quasi-Newton (2)

- Quasi-Newton iteration:

$$\mathbf{p}(k+1) = \mathbf{p}(k) - [\mathbf{J}(k)^{\mathsf{H}}\mathbf{J}(k) + \mathbf{M}(k)]^{-1}\,\mathbf{g}(k)$$

where matrix $\mathbf{M}(k)$ is updated from $\mathbf{J}(k)$, $\mathbf{M}(k)$, $\mathbf{g}(k)$ and $\mathbf{p}(k)$.

- The Levenberg-Marquardt iteration takes the form:

$$\mathbf{p}(k+1) = \mathbf{p}(k) - [\mathbf{J}(k)^{\mathsf{H}}\mathbf{J}(k) + \lambda(k)\,\mathbf{I}]^{-1}\,\mathbf{g}(k)$$

where $\lambda(k)$ is updated according to a specific rule, depending on the quality of the approximation of the objective:

$$\varepsilon(\mathbf{p}+\delta) - \varepsilon(\mathbf{p}) \approx \delta^{\mathsf{H}}\mathbf{g} + \frac{1}{2}\,\delta^{\mathsf{H}}(\mathbf{J}^{\mathsf{H}}\mathbf{J} + \lambda\,\mathbf{I})\delta$$

Gradient algorithms for tensors with symmetries

In the presence of symmetries, the gradient takes a simpler form, given here for clarity in the case of 3rd order tensors, with symmetry in the first 2 modes, i.e. [?]:

$$T_{i,j,k} = T_{\sigma(i,j),k}$$

We have two matrices to determine, **A** and **C** since:

$$\varepsilon = ||\mathbf{T} - \sum_{q=1}^{R} \mathbf{a}(q) \circ \mathbf{a}(q) \circ \mathbf{c}(q)||^2$$

The gradient and the Jacobian are of the form

$$\begin{aligned} \mathbf{g} &= \begin{bmatrix} \mathbf{g}_A + \mathbf{g}_B \\ \mathbf{g}_C \end{bmatrix} \\ \mathbf{J} &= [\mathbf{J}_A + \mathbf{J}_B,\ \mathbf{J}_C] \end{aligned}$$

where **B** is set to $\mathbf{B} = \mathbf{A}$.

Other minimization algorithms

Algorithms using explicit expressions of the Hessian

- Newton: $\mathbf{p}(k+1) = \mathbf{p}(k) - \mathbf{H}(k)^{-1}\,\mathbf{g}(k)$
- Conjugate Gradient: e.g. the "Multilinear Engine" [PAA99]
- etc...

➽ More costly in terms of memory and complexity per iteration, but fewer iterations needed.

➽ Do not solve the problem of *local minima*

Compact writing of Objective

The objective function (43) can be written as:

$$\varepsilon = ||\mathbf{T}_{I\times KJ} - \mathbf{A}(\mathbf{C}\odot\mathbf{B})^{\mathsf{T}}||^2 \tag{46}$$

Advantage: compact writing of the best matrix **A**, for fixed **B** and **C**, since (46) in quadratic in **A** [HL94]:

$$\hat{\mathbf{A}} = \mathbf{T}_{I\times KJ}\cdot\{(\mathbf{C}\odot\mathbf{B})^{\mathsf{T}}\}^{\dagger}$$

where † denotes pseudo-inverse.

Similarly:

$$\begin{aligned} ||\mathbf{T}_{J\times IK} - \mathbf{B}(\mathbf{A}\odot\mathbf{C})^{\mathsf{T}}||^2 &\rightarrow \hat{\mathbf{B}} = \mathbf{T}_{J\times IK}\cdot\{(\mathbf{A}\odot\mathbf{C})^{\mathsf{T}}\}^{\dagger} \\ ||\mathbf{T}_{K\times JI} - \mathbf{C}(\mathbf{B}\odot\mathbf{A})^{\mathsf{T}}||^2 &\rightarrow \hat{\mathbf{C}} = \mathbf{T}_{K\times JI}\cdot\{(\mathbf{B}\odot\mathbf{A})^{\mathsf{T}}\}^{\dagger} \end{aligned}$$

Alternating Least Squares algorithm (1)

Start with arbitrary $\mathbf{B}(0)$ and $\mathbf{C}(0)$
For $k = 1 ... kmax$,

- $\mathbf{A}(k+1) = \mathbf{T}_{I\times KJ} \cdot \{(\mathbf{C}(k) \odot \mathbf{B}(k))^{\mathsf{T}}\}^{\dagger}$
- $\mathbf{B}(k+1) = \mathbf{T}_{J\times IK} \cdot \{(\mathbf{A}(k+1) \odot \mathbf{C}(k))^{\mathsf{T}}\}^{\dagger}$
- $\mathbf{C}(k+1) = \mathbf{T}_{K\times JI} \cdot \{(\mathbf{B}(k+1) \odot \mathbf{A}(k+1))^{\mathsf{T}}\}^{\dagger}$

Hence the ALS algorithm also needs that:

$$R \leq \min(JK, IK, IJ)$$

According to Kruskal [KRU89], this inequality is always satisfied.

Alternating Least Squares algorithm (2)

Another compact writing [COM04a]: jointly diagonalize slices of lower order:

$$\varepsilon = \sum_{i=1}^{I} ||\mathbf{T}[i] - \mathbf{B}\mathbf{\Lambda}[i]\mathbf{C}^{\mathsf{T}}||^2$$

where $\mathbf{\Lambda}[i] = \mathbf{Diag}\{A_{i1}, \dots A_{kR}\}$. Let $\mathbf{\Lambda}[i]$ denote the vecor containing the diagonal of $\mathbf{\Lambda}[i]$, and $\mathbf{t}[i] \stackrel{\text{def}}{=} \mathbf{vec}\{\mathbf{T}[i]\}$. Hence:

$$\varepsilon = \sum_{i} ||\mathbf{t}[i] - \sum_{q=1}^{R} \lambda_n[i]\ \mathbf{c}[n] \otimes \mathbf{b}[n]||^2 \stackrel{\text{def}}{=} \sum_{i} ||\mathbf{t}[i] - \mathbf{M}\,\boldsymbol{\lambda}[k]||^2 \quad (47)$$

Then stationary values are:

$$\mathbf{B} = \{\sum_{k} \mathbf{T}[k]\mathbf{C}\mathbf{\Lambda}[k]\}\{\sum_{\ell} \mathbf{\Lambda}[\ell]\mathbf{C}^{\mathsf{T}}\mathbf{C}\mathbf{\Lambda}[\ell]\}^{-1}$$

$$\mathbf{C} = \{\sum_{k} \mathbf{T}[k]^{\mathsf{T}}\mathbf{B}\mathbf{\Lambda}[k]\}\{\sum_{\ell} \mathbf{\Lambda}[\ell]\mathbf{B}^{\mathsf{T}}\mathbf{B}\mathbf{\Lambda}[\ell]\}^{-1}$$

ALS for symmetric tensors (1)

For clarity, take a symmetric tensor $\mathbf{T}$ of order 4:

- One can force symmetry in the iteration of page 307:
  Start with arbitrary $\mathbf{A}(0)$, $\mathbf{A}(1)$, $\mathbf{A}(2)$
  For $k = 2...kmax$,
  Soft forcing:
  $\mathbf{A}(k+1) = \mathbf{T}_{I\times I^3} \cdot \{(\mathbf{A}(k) \odot \mathbf{A}(k-1) \odot \mathbf{A}(k-2))^\mathsf{T}\}^\dagger$
  Hard forcing: $\mathbf{A}(k+1) = \mathbf{T}_{I\times I^3} \cdot \{(\mathbf{A}(k) \odot \mathbf{A}(k) \odot \mathbf{A}(k))^\mathsf{T}\}^\dagger$

Obviously applies at any order $d \geq 3$ [?].

ALS for symmetric tensors (2)

More tricky iteration based on compact writing of page 308. When $\mathbf{T}$ is real symmetric:

$$\varepsilon = \sum_i ||\mathbf{T}[i] - \mathbf{B}\boldsymbol{\Lambda}[i]\mathbf{B}^{\mathsf{T}}||^2 \stackrel{\text{def}}{=} \sum_i ||\mathbf{t}[i] - \mathbf{M}\,\boldsymbol{\lambda}[i]||^2$$

- One shows that [COM04a] [YER02]

$$\boldsymbol{\lambda}[i] = \{\mathbf{M}^{\mathsf{T}}\,\mathbf{M}\}^{-1}\mathbf{M}^{\mathsf{T}}\,\mathbf{t}[i]$$

and each column of $\mathbf{B}$ is the dominant eigenvector of the real symmetric matrix:

$$\mathbf{P}[\ell] = \frac{1}{2}\sum_k \lambda_\ell[k]\{\tilde{\mathbf{T}}[k;\ell]^{\mathsf{T}} + \tilde{\mathbf{T}}[k;\ell]\}$$

where $\tilde{\mathbf{T}}[k;\ell] \stackrel{\text{def}}{=} \mathbf{T}[k] - \sum_{n\neq\ell}\lambda_n[k]\mathbf{b}[n]\mathbf{b}[n]^{\mathsf{T}}$.

ALS drawbacks

1. Fairly slow convergence when reaching plateaux
2. May be stuck about local minima

ALS with extrapolation

Attempt to face the first drawback [BA98] [BRO97].

- Compute stationary values $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ as in page 306
- At every other iteration, set:

$$\begin{aligned}
\mathbf{A}(k+1) &= \hat{\mathbf{A}} + \mu(k)\,(\mathbf{A}(k) - \hat{\mathbf{A}}) \\
\mathbf{B}(k+1) &= \hat{\mathbf{B}} + \mu(k)\,(\mathbf{B}(k) - \hat{\mathbf{B}}) \\
\mathbf{C}(k+1) &= \hat{\mathbf{C}} + \mu(k)\,(\mathbf{C}(k) - \hat{\mathbf{C}})
\end{aligned}$$

  where one may take $\mu(k) = k^{1/3}$.
- and otherwise $\mathbf{A}(k+1) = \hat{\mathbf{A}}$, $\mathbf{B}(k+1) = \hat{\mathbf{B}}$ and $\mathbf{C}(k+1) = \hat{\mathbf{C}}$.

ALS with Enhanced Line Search (ELS)

Attempt to face both drawbacks [RCH08] [RC05]

- Compute stationary values $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ as in page 306
- At every other iteration, set:

$$\begin{aligned}
\mathbf{A}(k+1) &= \hat{\mathbf{A}} + \mu\,(\mathbf{A}(k) - \hat{\mathbf{A}}) \\
\mathbf{B}(k+1) &= \hat{\mathbf{B}} + \mu\,(\mathbf{B}(k) - \hat{\mathbf{B}}) \\
\mathbf{C}(k+1) &= \hat{\mathbf{C}} + \mu\,(\mathbf{C}(k) - \hat{\mathbf{C}})
\end{aligned}$$

  where $\mu = Arg \min_\mu ||\mathbf{T} - \mathbf{A}(k+1) \bullet \mathbf{B}(k+1) \bullet \mathbf{C}(k+1)||^2$.
- and otherwise $\mathbf{A}(k+1) = \hat{\mathbf{A}}$, $\mathbf{B}(k+1) = \hat{\mathbf{B}}$ and $\mathbf{C}(k+1) = \hat{\mathbf{C}}$.

NB: $\mu(k)$ is obtained by rooting a polynomial of degree 5. $\Rightarrow$ one gets the *absolute minimum* along the search direction $\Rightarrow$ increased capability to escape from local minima.

ELS applied to other iterative algorithms

The same principle applies to any iterative algorithm [?]:

- Compute a search direction $[\Delta\mathbf{A}, \Delta\mathbf{B}, \Delta\mathbf{C}]$, which can be the gradient $\mathbf{g}$, a direction $\mathbf{H}^{-1}\mathbf{g}$, or a difference $[\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}] - [\mathbf{A}(k), \mathbf{B}(k), \mathbf{C}(k)]$...
- Compute the 6 first coefficients of the 6th degree polynomial $\varepsilon(\mu)$, defined by replacing $[\mathbf{A}, \mathbf{B}, \mathbf{C}]$ by $[\mathbf{A} + \mu\,\delta\mathbf{A}, \mathbf{B} + \mu\,\delta\mathbf{B}, \mathbf{C} + \mu\,\delta\mathbf{C}]$
- Compute the 5 roots of its derivative
- Select the root μ_o yielding the smallest minumum of $\varepsilon(\mu)$
- Update: $\mathbf{A}(k+1) = \mathbf{A}(k) + \mu_o\,\delta\mathbf{A}$, $\mathbf{B}(k+1) = \mathbf{B}(k) + \mu_o\,\delta\mathbf{B}$, $\mathbf{C}(k+1) = \mathbf{C}(k) + \mu_o\,\delta\mathbf{C}$.

Can be executed at every iteration, or less often.

Definition of c.f.'s

Characteristic functions

First: $\Phi_x(\mathbf{u}) \stackrel{\text{def}}{=} \mathrm{E}\{\exp(\jmath\,\mathbf{u}^\mathsf{T}\mathbf{x})\}$

Second: $\Psi_x(\mathbf{u}) \stackrel{\text{def}}{=} \log \Phi_x(\mathbf{u})$

Generating functions

First: $\Phi_x(\mathbf{u}) \stackrel{\text{def}}{=} \mathrm{E}\{\exp(\mathbf{u}^\mathsf{T}\mathbf{x})\}$

Second: $\Psi_x(\mathbf{u}) \stackrel{\text{def}}{=} \log \Phi_x(\mathbf{u})$

Key property

If $\mathbf{s}$ has statistically independent components

$$\Psi_s(\mathbf{u}) = \sum_p \Psi_{s_p}(u_p)$$

Characteristic function of a linear mixture

- If s_p independent, $\mathrm{E}\{\prod_p f(s_p)\} = \prod_p \mathrm{E}\{f(s_p)\}$.
- Hence if $\mathbf{x} = \mathbf{H}\,\mathbf{s}$, then

$$\begin{aligned}\Phi_x(\mathbf{u}) \stackrel{\text{def}}{=} \mathrm{E}\{\exp(\mathbf{u}^{\mathsf{T}}\mathbf{H}\,\mathbf{s})\} &= \mathrm{E}\{\exp(\sum_{p,q} u_q\, H_{qp}\, s_p)\} \\ &= \prod_p \mathrm{E}\{\exp(\sum_q u_q\, H_{qp}\, s_p)\}\end{aligned}$$

- Thus we have the *core equation*:

$$\Psi_s(\mathbf{u}) = \sum_p \Psi_{s_p}\left(\sum_q u_q\, H_{qp}\right)$$

Putting the problem in tensor form (1)

Goal: Find a matrix $\mathbf{H}$ such that the $K-$variate function $\Psi_x(\mathbf{u})$ decomposes into a sum of P univariate functions $\psi_p \stackrel{\text{def}}{=} \Psi_{s_p}$.

- Assumption: functions ψ_p, $1 \leq p \leq P$ admit finite derivatives up to order r in a neighborhood of the origin.
- Then, Taking $r = 3$ as a working example:

$$\frac{\partial^3 \Psi_x}{\partial u_i \partial u_j \partial u_k}(\mathbf{u}) = \sum_{p=1}^{P} H_{ip}\, H_{jp}\, H_{kp}\ \ \psi_p^{(3)}(\sum_{q=1}^{K} u_q\, H_{qp})$$

Putting the problem in tensor form (1)

Several equivalent writings:

- A decomposition into a sum of rank-1 terms:

$$T_{ijk\ell} = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, B_{\ell p}$$

- A joint diagonalization of matrix slices via a common rectangular transform

$$\mathbf{T}[k,\ell] = \mathbf{H} \,\cdot\, \mathbf{Diag}\{\mathbf{H}(k,:)\}\, \mathbf{Diag}\{\mathbf{B}(\ell,:)\} \,\cdot\, \mathbf{H}^{\mathsf{T}}$$

- The cumulant tensor case: only one point $\mathbf{u} = 0$, i.e. $\ell = 1$ and matrix **B** disappears.

Putting the problem in tensor form (3)

Use of several orders simultaneously:

- Order 3:

$$T^{(3)}_{ijk\ell} = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, B_{\ell p}$$

- Order 4:

$$T^{(4)}_{ijkm\ell} = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, H_{mp}\, C_{\ell p}$$

- Orders 3 and 4:

$$T_{ijk\ell}[m] = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, D_{\ell p}[m]$$

with $D_{\ell p}[m] = H_{mp}\, C_{\ell p}$ and $D_{\ell p}[0] = B_{\ell p}$.

BIOME algorithms

- These algorithms work with a cumulant tensor of even order $2r > 4$
- We take the case $2r = 6$ for the presentation, and denote

$$\mathcal{C}_{ijk}^{\ell mn} \stackrel{\text{def}}{=} \mathrm{Cum}\{x_i,\, x_j,\, x_k,\, x_l^*,\, x_m^*,\, x_n^*\} \qquad (48)$$

- In that case, we have

$$\mathcal{C}_{x,\,ijk}^{\ell mn} = \sum_{p=1}^{P} H_{ip}\, H_{jp}\, H_{kp}\, H_{\ell p}^*\, H_{mp}^*\, H_{np}^*\, \Delta_p$$

where $\Delta_p^{(6)} \stackrel{\text{def}}{=} \mathrm{Cum}\{s_p,\, s_p,\, s_p,\, s_p^*,\, s_p^*,\, s_p^*\}$ denote the diagonal entries of a $P \times P$ diagonal matrix, $\mathbf{\Delta}^{(6)}$

Writing in matrix form

- Tensor $\mathcal{C}_x$ is of dimensions $K \times K \times K \times K \times K \times K$ and enjoys symmetries and Hermitian symmetries.
- Tensor $\mathcal{C}_x$ can be stored in a $K^3 \times K^3$ Hermitian matrix, $\mathbf{C}_x^{(6)}$, called the *hexacovariance*. With an appropriate storage of the tensor entries, we have

$$\mathbf{C}_x^{(6)} = \mathbf{H}^{\odot 3}\, \boldsymbol{\Delta}^{(6)}\, \mathbf{H}^{\odot 3\,\mathsf{H}} \tag{49}$$

- Because $\mathbf{C}_x^{(6)}$ is Hermitian, $\exists \mathbf{V}$ unitary, such that

$$(\mathbf{C}_x^{(6)})^{1/2} = \mathbf{H}^{\odot 3}\, (\boldsymbol{\Delta}^{(6)})^{1/2}\, \mathbf{V} \tag{50}$$

- **Idea:** Use an invariance property existing between blocks of $(\mathbf{C}_x^{(6)})^{1/2}$.

Using the invariance to estimate V

- Cut the $K^3 \times P$ matrix $(\mathbf{C}_x^{(6)})^{1/2}$ into K blocks of size $K^2 \times P$.
- Each of these blocks, $\mathbf{\Gamma}[n]$, satisfies:

$$\mathbf{\Gamma}[n] = (\mathbf{H} \odot \mathbf{H}^{\mathsf{H}})\mathbf{D}[n]\,(\mathbf{\Delta}^{(6)})^{1/2}\,\mathbf{V}$$

  where $\mathbf{D}[n]$ is the $P \times P$ diagonal matrix containing the nth row of $\mathbf{H}$, $1 \leq n \leq K$.
- Hence matrices $\mathbf{\Gamma}[n]$ share the same common right singular space
- **Algorithm:** compute the joint EVD of the $K(K-1)$ matrices

$$\mathbf{\Theta}[m,n] \stackrel{\text{def}}{=} \mathbf{\Gamma}[m]^{\dagger}\mathbf{\Gamma}[n]$$

  as: $\mathbf{\Theta}[m,n] = \mathbf{V}\,\mathbf{\Lambda}[m,n]\,\mathbf{V}^{\mathsf{H}}$.

Estimation of $\mathbf{H}$

Matrices $\mathbf{\Lambda}[m, n]$ cannot be used directly because $(\mathbf{\Delta}^{(6)})^{1/2}$ is unknown. But we use $\mathbf{V}$ to obtain the estimate of $\mathbf{H}^{\odot 3}$ up to a scale factor:

$$\widehat{\mathbf{H}^{\odot 3}} = (\mathbf{C}_x^{(6)})^{1/2}\,\mathbf{V} \tag{51}$$

Then several possibilities exist to get $\mathbf{H}$ from $\mathbf{H}^{\odot 3}$ [ACCF04]. The best is as follows:

- Build K^2 matrices $\mathbf{\Xi}[m]$ of size $K \times P$ form conjugates rows of $\widehat{\mathbf{H}^{\odot 3}}$
- From $\mathbf{\Xi}[m]$ find matrices $\mathbf{D}[m]$ and $\widehat{\mathbf{H}}$ in the LS sense:

$$xx$$

Conditions of identifiability

- Xx [ACCF04] [AFCC03]
- Xx

FOOBI algorithms

xx

XX

- Xx
- Xx

XX

- Xx VI
- Xx

Part VII

Conclusions

False beliefs

1. BSS always requires High-Order Statistics (HOS)
   $\longrightarrow$ *Second-order can (rarely) suffice*
2. Sources must be statistically independent
   $\longrightarrow$ *Correlated sources can be sometimes separated (e.g. Discrete/CM sources, Pairwise cumulants...)*
3. HOS are always required when sources are *i.i.d.*
   $\longrightarrow$ *Second-order BSS algorithms exist*
4. Even local maxima of a contrast function yield good solutions
   $\longrightarrow$ *sometimes local maxima correspond to bad solutions*
5. There should be at least as many sensors as sources: $K \geq P$ (sufficient diversity)
   $\longrightarrow$ *Underdetermined mixtures can be identified*

False beliefs (cont'd)

6. Perfect source extraction is impossible if $K < P$
 $\longrightarrow$ *Discrete sources can often be perfectly extracted from under-determined mixtures (insufficient diversity)*
7. Conditions of application of Parafac are mild
 $\longrightarrow$ *except when one dimension* $= 2$, *the typical rank always exceeds the Parafac bound for uniqueness*
8. Approximate a tensor by another of lower rank is as easy as for matrices
 $\longrightarrow$ *beside for rank 1, there is a lack of closeness*
9. The Constant Modulus (CM) property is the best way to handle PSK sources
 $\longrightarrow$ *The whole alphabet can be taken into account in order to define a contrast function*

Part VIII

References

L. ALBERA, P. COMON, P. CHEVALIER, and A. FERREOL.

Blind identification of underdetermined mixtures based on the hexacovariance.

In *ICASSP'04*, volume II, pages 29–32, Montreal, May 17–21 2004.

L. ALBERA, P. COMON, and H. XU.

SAUD, un algorithme d'ica par déflation semi-algébrique.

In *GRETSI*, Troyes, September 2007.

L. ALBERA, A. FERREOL, P. COMON, and P. CHEVALIER.

Sixth order blind identification of under-determined mixtures - BIRTH.

In *ICA'03*, Japan, April 1-4 2003.

B. AFSARI.

What can make joint diagonalization difficult.

In *Icassp'07*, Honolulu, Hawai, April 15-20 2007.

J. ALEXANDER and A. HIRSCHOWITZ.

La methode d'Horace eclatee: application a l'interpolation en degre quatre.

Invent. Math., 107(3):585–602, 1992.

J. ALEXANDER and A. HIRSCHOWITZ.

Polynomial interpolation in several variables.

J. Algebraic Geometry, 4(2):201–222, 1995.

R. BRO and C. A. ANDERSSON.

Improving the speed of multiway algorithms. part ii : Compression.

Chemom. Intell. Lab. Syst., 42:105–113, 1998.

G. M. BERGMAN.

Ranks of tensors and change of base field.

Journal of Algebra, 11:613–621, 1969.

A. BENVENISTE, M. GOURSAT, and G. RUGET.

Robust identification of a non-minimum phase system.

IEEE Trans. Auto. Contr., 25(3):385–399, June 1980.

J. W. BREWER.

Kronecker products and matrix calculus in system theory.

IEEE Trans. on Circuits and Systems, 25(9):114–122, September 1978.

R. BRO.

PARAFAC, tutorial and applications.

Chemom. Intel. Lab. Syst., 38:149–171, 1997.

A. CICHOCKI and S-I. AMARI.

Adaptive Blind Signal and Image Processing.

Wiley, New York, 2002.

J. D. CARROLL and J. J. CHANG.

Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-Young decomposition.

Psychometrika, 35(3):283–319, September 1970.

P. COMON, G. GOLUB, L-H. LIM, and B. MOURRAIN.

Symmetric tensors and symmetric tensor rank.

SIAM Journal on Matrix Analysis Appl., 2008.

to appear.

P. COMON and C. JUTTEN.

Séparation de Sources: Concepts de base et Analyse en Composantes Indépendantes, volume 1 of *Collection IC2*.

Hermes, February 2007.

ISBN 978-2-7462-1517-7.

D. COX, J. LITTLE, and D. O'SHEA.

Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.

Undergraduate Texts in Mathematics. Springer Verlag, New York, 1992.

2nd ed. in 1996.

P. COMON and B. MOURRAIN.

Decomposition of quantics in sums of powers of linear forms.

Signal Processing, Elsevier, 53(2):93–107, September 1996.
special issue on High-Order Statistics.

P. COMON, B. MOURRAIN, L-H. LIM, and G. GOLUB.
Genericity and rank deficiency of high order symmetric tensors.
In *ICASSP'06*, Toulouse, May 14-19 2006.

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.

P. Comon.
Independent Component Analysis.
In J-L. Lacoume, editor, *Higher Order Statistics*, pages 29–38. Elsevier, Amsterdam, London, 1992.

P. COMON.
Independent Component Analysis, a new concept ?
Signal Processing, Elsevier, 36(3):287–314, April 1994.

Special issue on Higher-Order Statistics.

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.

P. COMON.

Supervised classification, a probabilistic approach.

In Verleysen, editor, *ESANN-European Symposium on Artificial Neural Networks*, pages 111–128, Brussels, Apr 19-21 1995. D facto Publ.

invited paper.

P. COMON.

"Cumulant tensors".

IEEE SP Workshop on HOS, Banff, Canada, July 21-23 1997.

invited keynote address,
`www.i3s.unice.fr/~comon/plenaryBanff97.html`.

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.

P. COMON.

Independent component analysis, contrasts and convolutive mixtures.

In *Second IMA Conference on Mathematics in Communications*, pages 10–17, Lancaster, UK, Dec. 16-18, 2002.

invited.

P. COMON.

Tensor decompositions.

In J. G. McWhirter and I. K. Proudler, editors, *Mathematics in Signal Processing V*, pages 1–24. Clarendon Press, Oxford, UK, 2002.

P. COMON.

Canonical tensor decompositions.

Research Report RR-2004-17, I3S, June 17 2004.

P. COMON.

Contrasts, independent component analysis, and blind deconvolution.

Int. Journal Adapt. Control Sig. Proc., 18(3):225–243, April 2004.

special issue on Signal Separation: http://www3.interscience.wiley.com/cgi-bin/jhome/4508. Preprint: I3S Research Report RR-2003-06.

F. CASTELLS, J. RIETA, J. MILLET, and V. ZARZOSO.

Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias.

IEEE Trans. Biomedical Eng., 52(2):258–267, February 2005.

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.

P. COMON and J. ten BERGE.

Generic and typical ranks of three-way arrays.

Research Report ISRN I3S/RR-2006-29-FR, I3S, Sophia-Antipolis, France, September 4 2006.

P. COMON and J. ten BERGE.

Generic and typical ranks of three-way arrays.

In *Icassp'08*, Las Vegas, March 30 - April 4 2008.

J. de LEEUW.

A new computational method to fit the weighted euclidean distance model.

Psychometrika, 43(4):479–490, December 1978.

N. DELFOSSE and P. LOUBATON.

Adaptive blind separation of independent sources: a deflation approach.

Signal Processing, 45:59–83, 1995.

L. de LATHAUWER.

Signal Processing based on Multilinear Algebra.

Doctorate, Katholieke Universiteit Leuven, September 1997.

L. de LATHAUWER, B. de MOOR, and J. VANDEWALLE.

Fetal electrocardiogram extraction by blind source subspace separation.

IEEE Trans. Biomedical Engineering, 47(5):567–572, May 2000.

Special Topic Section on Advances in Statistical Signal Processing for Biomedicine.

L. de LATHAUWER, B. de MOOR, and J. VANDEWALLE.

A multilinear singular value decomposition.

SIAM Jour. Matrix Ana. Appl., 21(4):1253–1278, April 2000.

L. de LATHAUWER, B. de MOOR, and J. VANDEWALLE.
On the best rank-1 and rank-(R1,R2,...RN) approximation of high-order tensors.
SIAM Jour. Matrix Ana. Appl., 21(4):1324–1342, April 2000.

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.

O. GRELLIER and P. COMON.
Blind separation of discrete sources.
IEEE Signal Processing Letters, 5(8):212–214, August 1998.

R. A. HARSHMAN.
Foundations of the Parafac procedure: Models and conditions for an explanatory multimodal factor analysis.
UCLA Working Papers in Phonetics, 16:1–84, 1970.
http://publish.uwo.ca/ harshman.

S. HAYKIN.

Unsupervised Adaptive Filtering, volume 1.

Wiley, 2000.

series in Adaptive and Learning Systems for Communications, Signal Processing, and Control.

S. HAYKIN.

Unsupervised Adaptive Filtering, volume 2.

Wiley, 2000.

series in Adaptive and Learning Systems for Communications, Signal Processing, and Control.

F. L. HITCHCOCK.

Multiple invariants and generalized rank of a p-way matrix or tensor.

J. Math. and Phys., 7(1):39–79, 1927.

R. A. HARSHMAN and M. LUNDY.

PARAFAC: Parallel factor analysis.

Computational Stat. Data Analysis, 18:39–72, 1994.

T. D. HOWELL.

Global properties of tensor rank.

Linear Algebra and Applications, 22:9–23, 1978.

A. HYVÄRINEN.

A family of fixed-point algorithms for Independent Component Analysis.

In *ICASSP'97*, pages 3917–3920, Munich, Germany, April 20-24 1997.

A. HYVÄRINEN.

Fast and robust fixed-point algorithms for independent component analysis.

IEEE Trans. Neural Networks, 10(3):626–634, 1999.

C. JUTTEN and P. COMON.

Séparation de Sources: Au delà de l'aveugle, et applications, volume 2 of *Collection IC2*.

Hermes, February 2007.

ISBN 978-2-7462-1518-4.

A. M. KAGAN, Y. V. LINNIK, and C. R. RAO.
Characterization Problems in Mathematical Statistics.
Probability and Mathematical Statistics. Wiley, New York, 1973.

J. B. KRUSKAL.
Three-way arrays: Rank and uniqueness of trilinear decompositions.
Linear Algebra and Applications, 18:95–138, 1977.

J. B. KRUSKAL.
Rank, decomposition, and uniqueness for 3-way and n-way arrays.
In R. Coppi and S. Bolasco, editors, *Multiway Data Analysis*, pages 7–18. Elsevier Science, North-Holland, 1989.

M. KENDALL and A. STUART.
The Advanced Theory of Statistics, Distribution Theory, volume 1.
C. Griffin, 1977.

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.

E. LASKER.
Kanonische formen.
Mathematische Annalen, 58:434–440, 1904.

L. De LATHAUWER.
A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization.
SIAM Journal on Matrix Analysis, 28(3):642–666, 2006.

P. McCULLAGH.
Tensor Methods in Statistics.
Monographs on Statistics and Applied Probability. Chapman and Hall, 1987.

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.

D. NION and L. De LATHAUWER.

Levenberg-Marquardt computation of the block factor model for blind multi-user access in wireless communications.

In *14th European Signal Processing Conference (EUSIPCO)*, Florence, Italy, September 4-8 2006.

C. L. NIKIAS and A. P. PETROPULU.

Higher-Order Spectra Analysis.

Signal Processing Series. Prentice-Hall, Englewood Cliffs, 1993.

P. PAATERO.

The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model.

Journal of Computational and Graphical Statistics, 8(4):854–888, December 1999.

D. T. PHAM.

Joint approximate diagonalization of positive definite hermitian matrices.

SIAM Journal on matrix Analysis, 22(4):1136–1152, 2001.

M. RAJIH and P. COMON.

Enhanced line search: A novel method to accelerate Parafac.

In *Eusipco'05*, Antalya, Turkey, Sept. 4-8 2005.

M. RAJIH, P. COMON, and R. HARSHMAN.

Enhanced line search : A novel method to accelerate PARAFAC.

SIAM Journal on Matrix Analysis Appl., 2008.

to appear.

J. J. RIETA, F. CASTELLS, C. SANCHEZ, V. ZARZOSO, and J. MILLET.

Atrial activity extraction for atrial fibrillation analysis using blind source separation.

IEEE Trans. Biomedical Eng., 51(7):1176–1186, July 2004.

B. REZNICK.

Sums of even powers of real linear forms.

Memoirs of the AMS, 96(463), March 1992.

J. R. RUIZ-TOLOSA and E. CASTILLO.

From Vectors to Tensors.

Universitext. Springer, 2005.

L. ROTA, V. ZARZOSO, and P. COMON.

Parallel deflation with alphabet-based criteria for blind source extraction.

In *IEEE SSP'05*, Bordeaux, France, July, 17-20 2005.

N. D. SIDIROPOULOS and R. BRO.

On the uniqueness of multilinear decomposition of N-way arrays.

Jour. Chemo., 14:229–239, 2000.

A. SMILDE, R. BRO, and P. GELADI.

Multi-Way Analysis.

Wiley, 2004.

B. W. SILVERMAN.

Density Estimation for Statistics and Data Analysis.

Chapman and Hall, 1986.

A. STEGEMAN and N. D. SIDIROPOULOS.

On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition.

Linear Algebra and Appl., 420:540–552, 2007.

V. STRASSEN.

Rank and optimal computation of generic tensors.

Linear Algebra Appl., 52:645–685, July 1983.

J. J. SYLVESTER.

Sur une extension d'un thorme de Clebsch relatif aux courbes du quatrime degr.

Comptes Rendus de l'Acadmie des Sciences, 102:1532–1534, 1886.

L. R. TUCKER.

Some mathematical notes for three-mode factor analysis.

Psychometrika, 31:279–311, 1966.

J. K. TUGNAIT.

Identification and deconvolution of multichannel non-Gaussian processes using higher order statistics and inverse filter criteria.

IEEE Trans. Sig. Proc., 45:658–672, March 1997.

S. TALWAR, M. VIBERG, and A. PAULRAJ.

Blind estimation of multiple co-channel digital signals arriving at an antenna array: Part I, algorithms.

IEEE Trans. Sig. Proc., pages 1184–1197, May 1996.

R. VOLLGRAF and K. OBERMAYER.

Quadratic optimization for simultaneous matrix diagonalization.

IEEE Trans. Sig,. Proc, 54:3270–3278, September 2006.

E. K. WAKEFORD.

Apolarity and canonical forms.

In *Proceedings of the London Mathematical Society*, pages 403–410, 1918.

A. YEREDOR.

Non-orthogonal joint diagoinalization in the LS sense with application in blind source separation.

IEEE Trans. Sig. Proc., 50(7):1545–1553, 2002.

V. ZARZOSO and P. COMON.

Blind and semi-blind equalization based on the constant power criterion.

IEEE Trans. Sig. Proc., 53(11):4363–4375, November 2005.

V. ZARZOSO and P. COMON.

Comparative speed analysis of FastICA.

In *ICA'07*, London, UK, September 9-12 2007.

V. ZARZOSO and P. COMON.

Optimal step-size constant modulus algorithm.

IEEE Trans. Com., January 2008.

submitted in 2004.

V. ZIVOJNOVIC.
Higher-order statistics and Huber's robustness.
In *IEEE-ATHOS Workshop on Higher-Order Statistics*, pages 236–240, Begur, Spain, 12–14 June 1995.