Tensor problems in Engineering

Pierre Comon

July 18, 2008

Contents I

Contents II

Lecture 1/3

Linear statistical model

$$\mathbf{y} = \mathbf{A}\,\mathbf{s} + \mathbf{b} \tag{1}$$

with
$$\begin{cases} \mathbf{y}: & K \times 1 \text{ random} \\ \mathbf{s}: & P \times 1 \text{ random with } \textit{stat. independent} \text{ entries} \\ \mathbf{A}: & K \times P \text{ deterministic} \\ \mathbf{b}: & \text{errors (may be removed for } P \text{ large enough)} \end{cases}$$

Taxinomy

- $K \geq P$: "over-determined"
  can be reduced to a square $P \times P$ regular mixture
  - **A** orthogonal or unitary
  - **A** square invertible
- $K < P$: "under-determined"
  - **A** rectangular with pairwise lin. independent columns

Goals

In the stochastic framework: solely from realizations of observation vector **y**,

- Estimate matrix **A**: Blind identification
- Estimate realizations of the "source" vector **s**: Blind separation/extraction/equalization

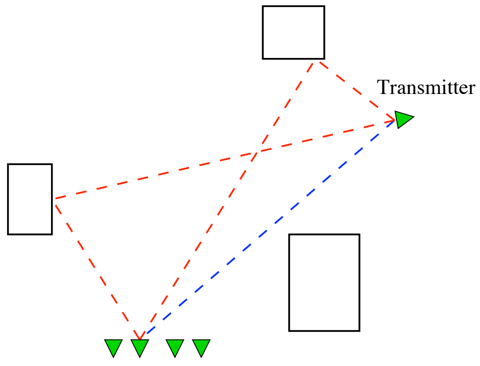
In the deterministic framework:

- Decompose the data tensor (cf. subsequent course)

Application areas for symmetric tensors

1. Telecommunications (Cellular, Satellite, Military),
2. Radar, Sonar,
3. Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...
4. Speech,
5. Machine Learning,
6. Control...

Example: Antenna Array Processing (1)



Example: Antenna Array Processing (2)

Modeling the signals received on an array of antennas generally leads to a *matrix decomposition*:

$$T_{ijp} = \sum_q \sum_\ell a_{iq\ell} \sum_k h_{q\ell kp} s_{kqj}$$

i:	space	k:	symbol #	**a**: receiver geometry
j:	time	q:	user #	**h**: global channel impulse response
		ℓ:	path #	**s**: Transmitted signal

But in the presence of additional *diversity*, a tensor can be constructed, thanks to new variable p

Example: Antenna Array Processing (3)

New variable p can represent:

- Oversampling (sample index),
- Spreading code (chip index),
- Frequency (multicarrier),
- Geometrical invariance (subarray index),
- Polarization...

Warning: tensor should not have proportional matrix slices (degeneration)

Link with tensors

In the stochastic framework:

- use of the characteristic function
- use of cumulants

The obtained tensor enjoys *symmetry properties*
$\longrightarrow$ another motivation to study symmetric tensors

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: $\Phi_{\mathbf{x}}(\mathbf{t}) \stackrel{\text{def}}{=} \mathrm{E}\{e^{\jmath\, \mathbf{t}^{\mathsf{T}}\mathbf{x}}\} = \int_{\mathbf{u}} e^{\jmath\, \mathbf{t}^{\mathsf{T}}\mathbf{u}}\, dF_{\mathbf{x}}(\mathbf{u})$

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 $\Phi(\mathbf{t}) \neq 0$

Characteristic functions (cont'd)

- Properties of the 2nd Characteristic function (cont'd):
  - if a c.f. $\Psi_x(t)$ is a polynomial, then its degree is at most 2 and x is Gaussian *(Marcinkiewicz'1938)* [Luka70]
  - if (x, y) statistically independent, then

$$\Psi_{x,y}(u, v) = \Psi_x(u) + \Psi_y(v) \tag{2}$$

Proof.

$$\begin{aligned}\Psi_{x,y}(u, v) &= \log[\mathrm{E}\{\exp(ux + vy)\}] \\ &= \log[\mathrm{E}\{\exp(ux)\}\, \mathrm{E}\{\exp(vy)\}].\end{aligned}$$

□

Problem posed in terms of Characteristic Functions

- If s_p independent and $\mathbf{x} = \mathbf{A}\,\mathbf{s}$, we have the *core equation*:

$$\Psi_x(\mathbf{u}) = \sum_p \Psi_{s_p}\left(\sum_q u_q\, A_{qp}\right) \tag{3}$$

Proof.

- Plug $\mathbf{x} = \mathbf{A}\,\mathbf{s}$, in definition of Ψ_x and get

$$\Phi_x(\mathbf{u}) \stackrel{\text{def}}{=} \mathrm{E}\{\exp(\mathbf{u}^{\mathsf{T}}\mathbf{A}\,\mathbf{s})\} = \mathrm{E}\{\exp(\sum_{p,q} u_q\, A_{qp}\, s_p)\}$$

- Since s_p independent, $\Phi_x(\mathbf{u}) = \prod_p \mathrm{E}\{\exp(\sum_q u_q\, A_{qp}\, s_p)\}$
- Taking the log concludes.

□

Problem: Decompose a mutlivariate function into a sum of univariate ones

Darmois-Skitovich theorem (1953)

Theorem

Let s_i be statistically *independent* random variables, and two linear statistics:

$$y_1 = \sum_i a_i\, s_i \quad \text{and} \quad y_2 = \sum_i b_i\, s_i$$

If y_1 and y_2 are statistically independent, then random variables s_k for which $a_k\, b_k \neq 0$ are Gaussian.

NB: holds in both $\mathbb{R}$ or $\mathbb{C}$

Sketch of proof

Let characteristic functions

$$\begin{aligned}\Psi_{1,2}(u,v) &= \log \mathrm{E}\{\exp(\jmath\, y_1\, u + \jmath\, y_2\, v)\}\\ \Psi_k(w) &= \log \mathrm{E}\{\exp(\jmath\, y_k\, w)\}\\ \varphi_p(w) &= \log \mathrm{E}\{\exp(\jmath\, s_p\, w)\}\end{aligned}$$

1. Independence between s_p's implies:

$$\begin{aligned}\Psi_{1,2}(u,v) &= \textstyle\sum_{k=1}^P \varphi_k(u\, a_k + v\, b_k)\\ \Psi_1(u) &= \textstyle\sum_{k=1}^P \varphi_k(u\, a_k)\\ \Psi_2(v) &= \textstyle\sum_{k=1}^P \varphi_k(v\, b_k)\end{aligned}$$

2. Independence between y_1 and y_2 implies

$$\Psi_{1,2}(u,v) = \Psi_1(u) + \Psi_2(v)$$

Does not restrict generality to assume that $[a_k, b_k]$ not collinear. To simplify, assume also φ_p differentiable.

3. Hence $\sum_{k=1}^P \varphi_p(u\, a_k + v\, b_k) = \sum_{k=1}^P \varphi_k(u\, a_k) + \varphi_k(v\, b_k)$
   Trivial for terms for which $a_k b_k = 0$.
   From now on, restrict the sum to terms $a_k b_k \neq 0$
4. Write this at $u + \alpha/a_P$ and $v - \alpha/b_P$:

$$\sum_{k=1}^P \varphi_k\left(u\, a_k + v\, b_k + \alpha(\frac{a_k}{a_P} - \frac{b_k}{b_P})\right) = f(u) + g(v)$$

5. Subtract to cancel Pth term, divide by α, and let $\alpha \to 0$:

$$\sum_{k=1}^{P-1} (\frac{a_k}{a_P} - \frac{b_k}{b_P})\, \varphi_k^{(1)}(u\, a_k + v\, b_k) = f^{(1)}(u) + g^{(1)}(v)$$

for some *univariate functions* $f^{(1)}(u)$ and $g^{(1)}(u)$.

Conclusion: We have one term less

6. Repeat the procedure $(P-1)$ times and get:

$$\prod_{j=2}^{P} (\frac{a_1}{a_j} - \frac{b_1}{b_j})\, \varphi_1^{(P-1)}(u\, a_1 + v\, b_1) = f^{(P-1)}(u) + g^{(P-1)}(v)$$

7. Hence $\varphi_1^{(P-1)}(u\, a_1 + v\, b_1)$ is linear, as a sum of two univariate functions ($\varphi_1^{(P)}$ is a constant because $a_1 b_1 \neq 0$).
8. Eventually φ_1 is a polynomial.
9. Lastly invoke Marcinkiewicz theorem to conclude that s_1 is Gaussian.
10. Same is true for any φ_p such that $a_p b_p \neq 0$: s_p is Gaussian.

NB: also holds if φ_p not differentiable

Equivalent representations

Let $\mathbf{y}$ admit two representations

$$\mathbf{y} = \mathbf{A}\, \mathbf{s} \quad \text{and} \quad \mathbf{y} = \mathbf{B}\, \mathbf{z}$$

where $\mathbf{s}$ (resp. $\mathbf{z}$) have independent components, and $\mathbf{A}$ (resp. $\mathbf{B}$) have pairwise noncollinear columns.

- These representations are *equivalent* if every column of $\mathbf{A}$ is proportional to some column of $\mathbf{B}$, and vice versa.
- If all representations of $\mathbf{y}$ are equivalent, they are said to be *essentially unique* (permutation & scale ambiguities only).

Identifiability & uniqueness theorems

Let $\mathbf{y}$ be a random vector of the form $\mathbf{y} = \mathbf{A}\,\mathbf{s}$, where s_p are independent, and $\mathbf{A}$ has non pairwise collinear columns.

- **Identifiability theorem** $\mathbf{y}$ can be represented as $\mathbf{y} = \mathbf{A}_1\,\mathbf{s}_1 + \mathbf{A}_2\,\mathbf{s}_2$, where $\mathbf{s}_1$ is non Gaussian, $\mathbf{s}_2$ is Gaussian independent of $\mathbf{s}_1$, and $\mathbf{A}_1$ is essentially unique.
- **Uniqueness theorem** If in addition the columns of $\mathbf{A}_1$ are linearly independent, then the distribution of $\mathbf{s}_1$ is unique up to scale and location indeterminacies.

Remark 1: if $\mathbf{s}_2$ is 1-dimensional, then $\mathbf{A}_2$ is also essentially unique
Remark 2: the proofs are not constructive [KagaLR73]

Example of uniqueness

Let s_i be independent with no Gaussian component, and b_i be independent Gaussian. Then the linear model below is identifiable, but $\mathbf{A}_2$ is not essentially unique whereas $\mathbf{A}_1$ is:

$$\begin{pmatrix} s_1 + s_2 + 2\,b_1 \\ s_1 + 2\,b_2 \end{pmatrix} = \mathbf{A}_1\,\mathbf{s} + \mathbf{A}_2 \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{A}_1\,\mathbf{s} + \mathbf{A}_3 \begin{pmatrix} b_1 + b_2 \\ b_1 - b_2 \end{pmatrix}$$

with

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},\ \mathbf{A}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \mathbf{A}_3 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Hence the distribution of $\mathbf{s}$ is essentially unique.
But $(\mathbf{A}_1,\,\mathbf{A}_2)$ not equivalent to $(\mathbf{A}_1,\,\mathbf{A}_3)$.

Example of non uniqueness

Let s_i be independent with no Gaussian component, and b_i be independent Gaussian. Then the linear model below is identifiable, but the distribution of $\mathbf{s}$ is not unique because a 2×4 matrix cannot be full column rank:

$$\begin{pmatrix} s_1 + s_3 + s_4 + 2\,b_1 \\ s_2 + s_3 - s_4 + 2\,b_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1 \\ s_2 \\ s_3 + b_1 + b_2 \\ s_4 + b_1 - b_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1 + 2\,b_1 \\ s_2 + 2\,b_2 \\ s_3 \\ s_4 \end{pmatrix}$$

with

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$$

Definition of Cumulants

- Moments:

$$\mu_r \overset{\text{def}}{=} \mathrm{E}\{x^r\} = (-\jmath)^r\, \left.\frac{\partial^r \Phi(t)}{\partial t^r}\right|_{t=0}$$

- Cumulants:

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

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

$$\log\,\Phi_x(t) = \Psi_x(t)$$

- Needs existence. Counter example: Cauchy

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

First Cumulants

- $\mathcal{C}_{(2)}$ is the variance: $\mathcal{C}_{(2)} = \mu_{(2)} - \mu_{(1)}^2$
- For zero-mean r.v.: $\mathcal{C}_{(3)} = \mu_{(3)}$, and $\mathcal{C}_{(4)} = \mu_{(4)} - 3\,\mu_{(2)}^2$
- 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: Zero-mean Gaussian

- Moments

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

In particular:

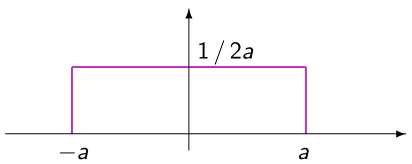
$$\mu_{(4)} = 3\mu_{(2)}^2, \quad \mu_{(6)} = 15\mu_{(2)}^3$$

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

Examples of Cumulants (2) S

Example: Uniform

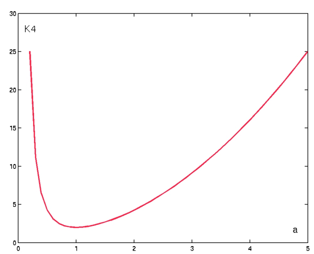
- uniformly distributed in $[-a, +a]$ with probability $\frac{1}{2a}$
- Moments: $\mu_{(2k)} = \frac{a^{2k}}{2k+1}$
- 4th order Cumulant: $\mathcal{C}_{(4)} = \frac{a^4}{5} - 3\,\frac{a^4}{9} = -2\,\frac{a^4}{15}$
- Kurtosis: $\mathcal{K}_{(4)} = -\frac{6}{5}$.



Examples of Cumulants (3) S

Example: 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 (bound)



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

Multivariate Cumulants

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

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

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

Properties of Cumulants

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

$$\begin{aligned} \text{Cum}\{\alpha X, Y, .., Z\} &= \alpha\, \text{Cum}\{X, Y, .., Z\} \qquad (4) \\ \text{Cum}\{X_1 + X_2, Y, .., Z\} &= \text{Cum}\{X_1, Y, .., Z\} + \text{Cum}\{X_2, Y, .., Z\} \end{aligned}$$

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

$$\text{Cum}\{X_1, X_2, .., X_r\} = 0$$

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

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

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

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

Problem posed in terms of Cumulants

Input-output relations If $\mathbf{y} = \mathbf{A}\,\mathbf{s}$, where s_p are independent, then multi-linearity of cumulants yields:

$$C_{\mathbf{y},ijk..\ell} = \sum_{p=1}^{P} A_{ip}\, A_{jp}\, A_{kp} .. A_{\ell p}\; C_{\mathbf{s},ppp..p} \qquad (5)$$

Can one identify $\mathbf{A}$ form tensor $\mathbf{C_y}$?

Remark

- Tensor $C_{\mathbf{y}}$ does not contain all the information whereas the c.f (3) did.
- Possibility to choose cumulant order(s)

Over-determined mixtures

In that framework, the statistical model involves *at most* as many sources as the dimension of the observation space:

$$\mathbf{y} = \mathbf{A}\,\mathbf{s}, \text{ with } K \stackrel{\text{def}}{=} dim\{\mathbf{y}\} \ge dim\{\mathbf{s}\} \stackrel{\text{def}}{=} P$$

That is, $\mathbf{A}$ admits a left inverse.

Warning:

- In practice, $C_{\mathbf{y}}$ or $\Psi_{\mathbf{y}}(\mathbf{u})$ are estimated from noisy measurements, so that (5) or (3) are never exactly satisfied if $K \ge P$: they become *approximations*

Essential uniqueness

- Over-determined mixtures are equivalent iff they are related by scale-permutation: $\mathbf{A} = \mathbf{B}\,\boldsymbol{\Lambda}\,\mathbf{P}$
- Hence in the absence of noise, the source random variables can be recovered up to scale and permutation as:

$$\mathbf{z} = \boldsymbol{\Lambda}\,\mathbf{P}\,\mathbf{s}$$

This is an inherent indeterminacy of the problem.

Direct vs Inverse

Two formulations in terms of cumulants:

1. Direct: look for $\mathbf{A}$ so as to fit eq. (5):

$$\min_{\mathbf{A}} ||C_{\mathbf{y},ijk..\ell} - \sum_{p=1}^{P} A_{ip}\, A_{jp}\, A_{kp}..\, A_{\ell p}\; C_{\mathbf{s},ppp..p}||^2$$

i.e. decompose $C_{\mathbf{y}}$ into a sum of P rank-one terms

2. Inverse: look for $\mathbf{B}$:

$$\min_{\mathbf{B}} \sum_{mnp..q \neq ppp..p} |\sum_{ijk..\ell} C_{\mathbf{y},ijk..\ell} B_{im}\, B_{jn}\, B_{kp}..\, B_{\ell q}|^2$$

i.e. try to diagonalize $C_{\mathbf{y}}$ by linear transform, $\mathbf{z} = \mathbf{B}\mathbf{y}$

Divide to conquer

Difficulty: many unknowns, in real or complex field

1. 1st idea: address a sequence of problems of smaller dimension instead of a single one in larger dimension.
2. 2nd idea: decompose $\mathbf{A}$ into two factors, $\mathbf{A} = \mathbf{L}\,\mathbf{Q}$, and compute $\mathbf{L}$ so as to *exactly* standardize the data. Look for the best $\mathbf{Q}$ in a second stage.

➠ Both are sub-optimal, but of practical value.

Pairwise vs. Mutual independence (1)

- Components s_k of $\mathbf{s}$ are *pairwise independent* $\Leftrightarrow$ Any pair of components (s_k, s_ℓ) are mutually independent.
- In general, pairwise and mutual independence *are not* equivalent
- But...

Pairwise vs. Mutual independence (2)

Example: Pairwise but not Mutual independence

- 3 mutually independent binary sources, $x_i \in \{-1, 1\}$, $1 \le i \le 3$
- Define $x_4 = x_1 x_2 x_3$. Then x_4 is also binary, *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 binary $\Rightarrow$
  $p(x_2\, x_3 = b/a).p(x_1 = a) = \frac{1}{2} \cdot \frac{1}{2}$
- NB: in particular, $\mathrm{Cum}\{x_1, x_2, x_3, x_4\} = 1 \neq 0$

Pairwise vs. Mutual independence (3)

Corollary of Darmois's theorem for over-determined mixtures:

Let $\mathbf{z} = \mathbf{C}\,\mathbf{s}$, where s_i are independent r.v., and assume either

- all s_i are non Gaussian and $\mathbf{C}$ is invertible, or
- at most one s_i is Gaussian and $\mathbf{C}$ is orthogonal/unitary

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 invertible and $\mathbf{P}$ permutation

Proof... $\boxed{1} \rightarrow \boxed{3}$ by contradiction

Standardization with covariance

- Let $\mathbf{x}$ be a zero-mean r.v. with covariance matrix:

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

and let $\mathbf{L}$ be any square root of $\mathbf{\Gamma}$:

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

Then $\tilde{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{L}^{-1}\mathbf{x}$ is a *standardized random variable*, i.e. it has unit covariance.

- In practice, use SVD to face ill-conditioning or singularity: $\tilde{\mathbf{x}}$ may have then a smaller dimension.

Pre-processing with SVD

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

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

  where "Principal Components" $\tilde{x}_i$ are *uncorrelated*.
  ith column $\mathbf{u}_i$ of $\mathbf{U}$ is called "ith PC loading vector"
- Two –among many– possible calculations from a $K \times N$ realization matrix $\mathbf{X}$:
  - EVD of sample covariance $\mathbf{R}_x$: $\mathbf{R}_x = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^{\mathsf{H}}$, $\tilde{\mathbf{x}} = \mathbf{U}^{\mathsf{H}}\,\mathbf{x}$
  - Sample estimate by SVD: $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}}$, $\tilde{\mathbf{X}} = \mathbf{\Sigma}\mathbf{V}^{\mathsf{H}}$

  The latter is more accurate, and avoids the calculation of the covariance.

Lecture 2/3

Orthogonal decomposition

If $\mathbf{Q}$ orthogonal, the two problems are equivalent:

1. Direct:

$$\min_{\mathbf{Q},\mathbf{\Lambda}} ||\mathcal{C}_{ijk\ell} - \sum_{p=1}^{P} Q_{ip}\, Q_{jp}\, Q_{kp}\, Q_{\ell p}\, \Lambda_{pppp}||^2$$

2. Inverse: $\min_{\mathbf{Q},\mathbf{\Lambda}} ||\sum_{ijk\ell} Q_{ip}\, Q_{jq}\, Q_{kr}\, Q_{\ell s}\, \mathcal{C}_{ijk\ell} - \Lambda_{pppp}\, \delta_{pqrs}||^2$ or

$$\textit{e.g.} \quad \max_{\mathbf{Q}} \sum_p |\sum_{ijk\ell} Q_{ip}\, Q_{jp}\, Q_{kp}\, Q_{\ell p}\, \mathcal{C}_{ijk\ell}|^2$$

Proof. The Frobenius norm is invariant under orthogonal change of coordinates. □

What we have seen so far

- The joint 2nd characteristic function of $\mathbf{x} = \mathbf{A}\mathbf{s}$ contains all the information. The problem consists of decomposing it into a sum of univariate characteristic functions.
- Cumulant tensors contain part of the information, but may suffice. The problem reduces to decomposing one of them.
- Identification of an invertible mixture generally leads to an approximation problem
- 1st idea: in an inverse approach, one can process pairwise
- 2nd idea: two-stage by decomposing $\mathbf{A} = \mathbf{L}\,\mathbf{Q}$. First compute $\mathbf{L}$ via an exact standardization of $\mathbf{x}$. Second compute the best orthogonal matrix $\mathbf{Q}$.
- Drawback: one puts an infinite weight on 2nd order statistics

Estimation of the orthogonal matrix

Assume we have realizations of the standardized r.v. $\tilde{\mathbf{x}} = \mathbf{L}^{-1}\mathbf{x}$. We have to:

1. Choose an optimization criterion to maximize, e.g. based on the cumulant tensor of $\mathbf{z}$.
2. Devise a numerical algorithm, e.g. proceeding pairwise

Contrast criteria

1 Optimization criteria

Let $\mathbf{s}$ have stat. independent components, and $\mathbf{z} = \mathbf{Q}\,\mathbf{s}$. A good "contrast" criterion $\Upsilon(\mathbf{Q})$ should:

- be maximal if and only if the components of $\mathbf{z}$ are independent, ie. iff $\mathbf{Q}$ is *trivial* (scale-permutation): $\mathbf{Q} = \mathbf{\Lambda}\,\mathbf{P}$,
- decrease if $\mathbf{Q}$ is not trivial.

This can be summarized by:

$$\Upsilon(\mathbf{Q}) \leq \Upsilon(\mathbf{I}), \text{ with equality iff } \mathbf{Q} = \mathbf{\Lambda}\,\mathbf{P} \tag{6}$$

Examples of contrasts

Denote $\mathbf{C}$ the 4th order cumulant tensor of $\mathbf{z}$. If at most 1 source has a null kurtosis, then

- $\Upsilon_{CoM}(\mathbf{Q}) \stackrel{\text{def}}{=} \sum_i C_{iiii}^2$ (maximize diagonal entries)
- $\Upsilon_{STO}(\mathbf{Q}) \stackrel{\text{def}}{=} \sum_{ij} C_{iiij}^2$ (jointly diagonalize 3rd order slices)
- $\Upsilon_{JAD}(\mathbf{Q}) \stackrel{\text{def}}{=} \sum_{ijk} C_{iijk}^2$ (jointly diagonalize matrix slices)

are "contrast" criteria.

Examples of contrasts (cont'd)

Proof. We need to prove (6), that is, $\Upsilon(\mathbf{Q}) \leq \Upsilon(\mathbf{I})$. Denote $\mathbf{D}$ the (diagonal) cumulant tensor of $\mathbf{s}$. First use multi-linearity (5) of cumulant tensors: $C_{ijk\ell} = \sum_p Q_{ip}\,Q_{jp}\,Q_{kp}\,Q_{\ell p}\,D_p$. Then:

$$\begin{aligned}\Upsilon_{JAD} &\stackrel{\text{def}}{=} \sum_{ijk} C_{iijk}^2 = \sum_{ijk} \Big|\sum_p Q_{ip}^2\,Q_{jp}\,Q_{kp}\,D_p\Big|^2 \\ &= \sum_i \sum_{pq} Q_{ip}^2\,Q_{iq}^2 \sum_j Q_{jp}\,Q_{jq} \sum_k Q_{kp}\,Q_{kq}\,D_p\,D_q\end{aligned}$$

and because $\mathbf{Q}$ is orthogonal

$$\Upsilon_{JAD} = \sum_i \sum_p Q_{ip}^4\,D_p^2$$

Now $Q_{ij} \leq 1$ yields $\Upsilon_{JAD} \leq \sum_i \sum_p Q_{ip}^2\,D_p^2 = \sum_i D_i^2 \stackrel{\text{def}}{=} \Upsilon_{JAD}(\mathbf{I})$

This proves inequality (6) for Υ_{JAD}.

Proof. (cont'd) On the other hand, clearly

$$\Upsilon_{CoM} \leq \Upsilon_{STO} \leq \Upsilon_{JAD}$$

so that we have also proved the inequality (6) for Υ_{CoM} and Υ_{STO}. Now if equality holds, all inequalities above are equalities, and in particular

$$\sum_i \sum_p (Q_{ip}^2 - Q_{ip}^4)\,D_p^2 = 0$$

Thus $|Q_{ip}| \in \{0, 1\}$ by L^p-norm inequality, for all i and for all p such that $D_p \neq 0$. Yet, at most one D_p is null by hypothesis, and $\mathbf{Q}$ is orthogonal. Thus $\mathbf{Q}$ must be a signed permutation.

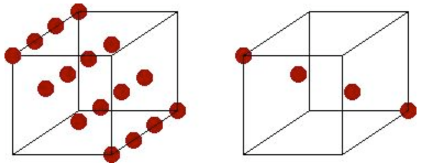
$\square$

Examples of contrasts (cont'd)

Interpretation.

- $\Upsilon_{CoM} \leq \Upsilon_{STO} \leq \Upsilon_{JAD}$ shows that Υ_{CoM} is more sensitive
- Υ_{STO} and Υ_{JAD} are less discriminant since they also maximize non diagonal terms

Example for $4 \times 4 \times 4$ tensors



Matrix slices diagonalization $\neq$ Tensor diagonalization

Algorithms: Jacobi pair Sweeping (1)

2 Pairwise processing

Split the orthogonal matrix into a product of plane *Givens* rotations:

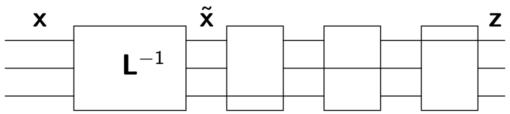
$$\mathbf{G}[i,j] \stackrel{\text{def}}{=} \frac{1}{\sqrt{1+\theta^2}} \begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}$$

acting in the subspace defined by (z_i, z_j).

➡ the dimension has been reduced to 2, and we have a single unknown, θ, that can be imposed to lie in $(-1, 1]$.

Algorithms: Jacobi pair Sweeping (2)

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

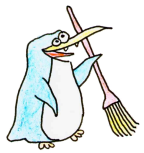


Carl Jacobi, 1804-1851

Algorithms: Jacobi pair Sweeping (3)

Sweeping a $3 \times 3 \times 3$ symmetric tensor

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



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

Algorithms: Jacobi pair Sweeping (4)

- Criteria Υ_{CoM}, Υ_{STO} and Υ_{JAD}, are rational functions of θ, and their absolute maxima can be computed algebraically.
- To prove this, consider the elementary 2×2 problem

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

denote $C_{ijk\ell}$ the cumulants of $\mathbf{z}$, and

$$\mathbf{G} \stackrel{\text{def}}{=} \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \stackrel{\text{def}}{=} \frac{1}{\sqrt{1+\theta^2}} \begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}$$

Solution for Υ_{CoM}

- We have $\Upsilon_{CoM} \stackrel{\text{def}}{=} (C_{1111})^2 + (C_{2222})^2$
- Denote $\xi = \theta - 1/\theta$. Then it is a rational function in ξ:

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

- Its stationary points are roots of a polynomial of degree 4:

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

obtainable *algebraically* via Ferrari's technique. Coefficients b_i and c_i are given functions of cumulants of $\tilde{\mathbf{x}}$.
- θ is obtained from ξ by rooting a 2nd degree trinomial.

Solution for Υ_{JAD}

- Goal: maximize squares of diagonal terms of $\mathbf{G}^{\mathsf{H}}\mathbf{N}(r)\,\mathbf{G}$, where matrix slices are denoted $\mathbf{N}(r) = \begin{pmatrix} a_r & b_r \\ c_r & d_r \end{pmatrix}$ and are cumulants of $\tilde{\mathbf{x}}$
- Let $\mathbf{v} \stackrel{\text{def}}{=} [\cos 2\beta,\ \sin 2\beta]^{\mathsf{T}}$. Then this amounts to maximizing the quadratic form $\mathbf{v}^{\mathsf{T}}\,\mathbf{M}\mathbf{v}$ where

$$\mathbf{M} \stackrel{\text{def}}{=} \sum_r \begin{bmatrix} a_r - d_r \\ b_r + c_r \end{bmatrix} [a_r - d_r,\ b_r + c_r]$$

- Thus, 2β is given by the dominant eigenvector of $\mathbf{M}$
- and $\mathbf{G}$ is obtained by rooting a 2nd degree trinomial.

Solution for Υ_{STO}

- Goal: maximize squares of diagonal terms of 3rd order tensors $\mathbf{T}[\ell]_{pqr} \stackrel{\text{def}}{=} \sum_{ijk} G_{pi}\,G_{qj}\,G_{rk}\,C_{ijk\ell}$
- If $\mathbf{G} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}$ then denote $\mathbf{v} \stackrel{\text{def}}{=} [\cos 2\beta,\ \sin 2\beta]^{\mathsf{T}}$.
- Angle 2β is given by vector $\mathbf{v}$ maximizing a quadratic form $\mathbf{v}^{\mathsf{T}}\mathbf{B}\mathbf{v}$, where $\mathbf{B}$ is 2×2 symmetric and contains sums of products of cumulants of $\tilde{\mathbf{x}}$
- θ is obtained from ξ by rooting a 2nd degree trinomial.

First conclusions

- Pair sweeping can be executed thanks to the equivalence between pairwise and mutual independence
- The cumulant tensor can be diagonalized iteratively via a Jacobi-like algorithm
- For each pair, there is a closed-form solution for the optimal Givens rotation (absolute maximimum of the contrast criterion).

Questions:

- But what about *global* convergence?
- Pairwise processing holds valid if model is exact

Stationary points: symmetric matrix case

- Given a matrix $\mathbf{m}$ with components m_{ij}, it is sought for an orthogonal matrix $\mathbf{Q}$ such that Υ_2 is maximized:

$$\Upsilon_2(\mathbf{Q}) = \sum_i M_{ii}^2; \quad M_{ij} = \sum_{p,q} Q_{ip} Q_{jq}\, m_{pq}.$$

- Stationary points of Υ_2 satisfy for any pair of indices $(q, r), q \neq r$:

$$M_{qq} M_{qr} = M_{rr} M_{qr}$$

- Next, $d^2\Upsilon_2 < 0 \Leftrightarrow M_{qr}^2 < (M_{qq} - M_{rr})^2$, which proves that
  - $M_{qr} = 0$, $\forall q \neq r$ yields a maximum
  - $M_{qq} = G_{rr}$, $\forall q, r$ yields a minimum
  - Other stationary points are saddle points

Stationary points: symmetric tensor case

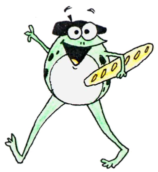
- Similarly, one can look at relations characterizing local maxima of criterion Υ

$$\begin{aligned} T_{qqqq} T_{qqqr} - T_{rrrr} T_{qrrr} &= 0, \\ 4T_{qqqr}^2 + 4T_{qrrr}^2 - (T_{qqqq} - \frac{3}{2}\, T_{qqrr})^2 & \\ -(T_{rrrr} - \frac{3}{2}\, T_{qqrr})^2 &< 0. \end{aligned}$$

  for any pair of indices $(p, q), p \neq q$.
- As a conclusion, contrary to Υ_2 in the matrix case, Υ might have theoretically spurious local maxima in the tensor case (order > 2).

Problem P2

1. At each step, a plane rotation is computed and yields the global maximum of the objective Υ restricted to one variable
2. There is no proof that the sequence of successive plane rotations yields the global maximum, in the general case (tensors that are not necessarily diagonalizable)
3. Yet, no counter-example has been found since 1991



Other algorithms for orthogonal diagonalization

The 3 previous criteria summarize most ways to address tensor approximate diagonalization.
But the orthogonal matrix can be treated differently, e.g. via *other parameterizations*

- express an orthogonal matrix as the exponential of a skew-symmetric matrix: $\mathbf{Q} = \exp \mathbf{S}$
- or use the Cayley parameterization: $\mathbf{Q} = (\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S})^{-1}$,
- ...

Joint Approximate Diagonalization S

The idea is to consider the symmetric tensor of dimension K and order d as a collection of K^{d-2} symmetric $K \times K$ matrices

This is the *Joint Approximate Diagonalization* (JAD) problem

Bibliographical comments S

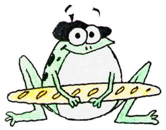
- Orthogonal diagonalization of symmetric tensors:
  - JAD in 2 modes: [DeLe78] ($\mathbb{R}$), [CardS93] ($\mathbb{C}$), with matrix exponential [TanaF07]
  - JAD 2 modes with positive definite matrices: [Flur86]
  - JAD in 3 modes: [DelaDV01]
  - direct diago without slicing (pairs): [Como92] [Como94]
- Orthogonal diagonalization of non symmetric tensors:
  - ALS type [Kroo83] [Kier92]
  - ALS on pairs: [MartV08] [SoreC08]
  - JAD in 2 modes ($\mathbb{R}$): [Pesq01]
  - Matrix exponential [SorelCD08]

Approximate diagonalization by invertible transform

- Joint Approximate Diagonalization (JAD) of a collection of:
  - symmetric matrices
  - symmetric diagonally dominant matrices
  - symmetric positive definite matrices
  - for a collection of matrices, not necessarily symmetric (2 invertible transforms)
  - ...
- Direct approaches without slicing the tensor into a collection of matrices: algorithms devised for *underdetermined mixtures* apply (cf. subsequent lecture)

Problem P3

- Ill-poseness of optimization over set of invertible matrices
- Possible solution:
  impose a constraint like $\det \mathbf{A} \geq \eta > 0$?



Survey of 4 algorithms

1. Iterative algorithm based on a probabilistic criterion
2. An algorithm of ALS type: ACDC
3. An algorithm with a multiplicative update, valid if $\mathbf{A}$ is diagonally dominant
4. An algorithm based on Joint triangularization

Probabilistic approach (1) S

Let $\mathbf{T}(q)$ be a collection of symmetric *positive semidefinite* matrices
Look for $\mathbf{B}$ such that $\mathbf{M}(q) \stackrel{\text{def}}{=} \mathbf{B}\,\mathbf{T}(q)\,\mathbf{B}^{\mathsf{T}}$ are as diagonal as possible.

1. Choose criterion to maximize $\Upsilon \stackrel{\text{def}}{=} \sum_q \alpha_q \log \frac{\det \mathbf{M}(q)}{\det \mathrm{Diag}\{\mathbf{M}(q)\}}$
   We have $\Upsilon \leq 0$ from Hadamard's inequality.
   - Avoids singularity
   - Linked to Maximum Likelihood
2. Use multiplicative update as $\mathbf{B}^{(\ell+1)} = \mathbf{U}\,\mathbf{B}^{(\ell)}$
3. Criterion after update: $\Upsilon = \sum_q \alpha_q \log \frac{\det \mathbf{M}(q)\, \det^2 \mathbf{U}}{\det \mathrm{Diag}\{\mathbf{U}\mathbf{M}(q)\mathbf{U}^{\mathsf{T}}\}}$

Probabilistic approach (2) S

4. Variation of Υ during one update:
   $\sum_q \alpha_q \left[2 \log \det \mathbf{U} - \log \det \mathrm{Diag}\{\mathbf{U}\mathbf{M}(q)\mathbf{U}^{\mathsf{T}}\} + \log \det \mathrm{Diag}\mathbf{M}(q)\right]$
5. Update two rows at a time, i.e. $\mathbf{U}$ is equal to Identity except for entries $(i,i), (i,j), (j,i), (j,j)$.
   By concavity of log, get a lower bound on variation:
   $$\sum_q \alpha_q \left[2 \log \det \mathbf{U} - \log(\mathbf{U}\mathbf{P}\mathbf{U}^{\mathsf{T}})_{11} - \log(\mathbf{U}\mathbf{Q}\mathbf{U}^{\mathsf{T}})_{22}\right]$$
   where $\mathbf{P}$ and $\mathbf{Q}$ are the 2×2 matrices:
   $\mathbf{P} = \sum_q \frac{p_q}{M(q)_{ii}} \mathbf{M}(q)[i,j]$ and $\mathbf{Q} = \sum_q \frac{p_q}{M(q)_{jj}} \mathbf{M}(q)[i,j]$
6. Maximize this bound instead. This leads to rooting a *2nd degree trinomial*. Sweep all the pairs in turns

Alternate Least Squares

1. Two writings of the criterion:

$$\Upsilon = \sum_q ||\mathbf{T}(q) - \mathbf{B}\boldsymbol{\Lambda}(q)\mathbf{B}^{\mathsf{H}}||^2$$

$$\Upsilon = \sum_q ||\mathbf{t}(q) - \mathcal{B}\,\boldsymbol{\lambda}(q)||^2$$

2. Stationary values for Diag$\boldsymbol{\Lambda}(q)$: $\boldsymbol{\lambda}(q) = \{\mathcal{B}^{\mathsf{H}}\mathcal{B}\}^{-1}\mathcal{B}^{\mathsf{H}}\,\mathbf{t}(q)$
3. Stationary value for each column $\mathbf{b}[\ell]$ of matrix $\mathbf{B}$ is the dominant eigenvector of the Hermitean matrix

$$\mathbf{P}[\ell] = \frac{1}{2}\sum_q \lambda_\ell(q)\{\tilde{\mathbf{T}}[q;\ell]^{\mathsf{H}} + \tilde{\mathbf{T}}[q;\ell]\}$$

where $\tilde{\mathbf{T}}[q;\ell] \stackrel{\text{def}}{=} \mathbf{T}(q) - \sum_{n\neq\ell}\lambda_n(q)\,\mathbf{b}[n]\mathbf{b}[n]^{\mathsf{H}}$.

4. ALS: calculate $\boldsymbol{\Lambda}(q)$ and $\mathbf{B}$ *alternately* Use LS solution when matrices are singular.

Diagonally dominant matrices (1)

One wishes to minimize iteratively $\sum_q ||\mathbf{T}(q) - \mathbf{A}\,\boldsymbol{\Lambda}_q\,\mathbf{A}^{\mathsf{T}}||^2$
Assume $\mathbf{A}$ is strictly diagonally dominant: $|A_{ii}| > \sum_{j\neq i}|A_{ij}|$
(cf. Levy-Desplanques theorem)

1. Initialize $\mathbf{A} = \mathbf{I}$
2. Update $\mathbf{A}$ multiplicatively as $\mathbf{A} \leftarrow (\mathbf{I} + \mathbf{W})\mathbf{A}$, where $\mathbf{W}$ is zero-diagonal
3. Compute the best $\mathbf{W}$ assuming that it is small and that $\mathbf{T}(q)$ are almost diagonal (first order approximation)

Diagonally dominant matrices (2)

Computational details:

- We have: $\mathbf{T}^{(\ell+1)}(q) \leftarrow (\mathbf{I} + \mathbf{W})\mathbf{T}^{(\ell)}(q)(\mathbf{I} + \mathbf{W})^{\mathsf{T}}$
- $\mathbf{T}^{(\ell)}(q) \stackrel{\text{def}}{=} (\mathbf{D} - \mathbf{E})$, where $\mathbf{D}$ is diagonal, and $\mathbf{E}$ zero-diagonal
- If $\mathbf{W}$ and $\mathbf{E}$ are small:
  $\mathbf{T}^{(\ell+1)}(q) \approx \mathbf{D}(q) + \mathbf{W}\,\mathbf{D}(q) + \mathbf{D}(q)\,\mathbf{W}^{\mathsf{T}} - \mathbf{E}(q)$
- Hence minimize, wrt $\mathbf{W}$:

$$\sum_q \sum_{i\neq j} |W_{ij}\,D_{jj}(q) + W_{ji}\,D_{ii}(q) - E_{ij}(q)|^2$$

- This is of the form $\min_{\mathbf{w}} ||\mathbf{J}\,\mathbf{w} - \mathbf{e}||^2$, where $\mathbf{J}$ is sparse: $\mathbf{J}^{\mathsf{T}}\mathbf{J}$ is block diagonal

Thus one gets at each iteration a collection of *decoupled* 2×2 linear systems

Joint triangularization of matrix slices

1. From $\mathbf{T}$, determine a collection of matrices (e.g. matrix slices), $\mathbf{T}(q)$, satisfying $\mathbf{T}(q) = \mathbf{A}\,\mathbf{D}(q)\,\mathbf{B}^{\mathsf{T}}$, $\mathbf{D}(q) \stackrel{\text{def}}{=} \text{diag}\{C_{q,:}\}$.
2. Compute the Generalized Schur decomposition
   $\mathbf{Q}\,\mathbf{T}(q)\,\mathbf{Z} = \mathbf{R}(q)$, where $\mathbf{R}(q)$ are upper-triangular
3. Since $\mathbf{Q}^{\mathsf{T}}\,\mathbf{R}(q)\,\mathbf{Z}^{\mathsf{T}} = \mathbf{A}\,\mathbf{D}(q)\,\mathbf{B}^{\mathsf{T}}$, matrices $\mathbf{R}' \stackrel{\text{def}}{=} \mathbf{Q}\,\mathbf{A}$ and $\mathbf{R}'' \stackrel{\text{def}}{=} \mathbf{B}^{\mathsf{T}}\,\mathbf{Z}$ are upper triangular, and can be assumed to have a unit diagonal. Hence $\mathbf{R}'$ and $\mathbf{R}''$ can be computed by solving from the bottom to the top the triangular system, two entries R'_{ij} and R''_{ij} at a time:

$$\mathbf{R}(q) = \mathbf{R}'\,\mathbf{D}(q)\,\mathbf{R}''$$

4. Compute $\mathbf{A} = \mathbf{R}'\,\mathbf{Q}^{\mathsf{T}}$ and $\mathbf{B}^{\mathsf{T}} = \mathbf{R}''\,\mathbf{Z}^{\mathsf{T}}$
5. Compute matrix $\mathbf{C}$ from $\mathbf{T}$, $\mathbf{A}$ and $\mathbf{B}$ by solving the over-determined linear system $\mathbf{C}\cdot\{(\mathbf{B}\odot\mathbf{A})^{\mathsf{T}}\} = \mathbf{T}_{K\times JI}$

Bibliographical comments

For collection of symmetric matrices:

- maximizes iteratively a lower bound to the decrease on a probabilistic objective [Pham01]
- alternately minimize $\sum_q ||\mathbf{T}(q) - \mathbf{A}\,\mathbf{\Lambda}(q)\,\mathbf{A}^{\mathsf{T}}||^2$ wrt $\mathbf{A}$ and $\mathbf{\Lambda}(q)$ [Yere02]. See also: [Li07] [VollO06]
- minimizes iteratively $\sum_q ||\mathbf{T}(q) - \mathbf{A}\,\mathbf{\Lambda}(q)\,\mathbf{A}^{\mathsf{T}}||^2$ under the assumption that $\mathbf{A}$ is diagonally dominant [Zieh04] .

For collection of non symmetric matrices:

- factor $\mathbf{A}$ into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].
- others: algorithms applicable to underdetermined case work here [AlbeFCC05]

Lecture 3/3

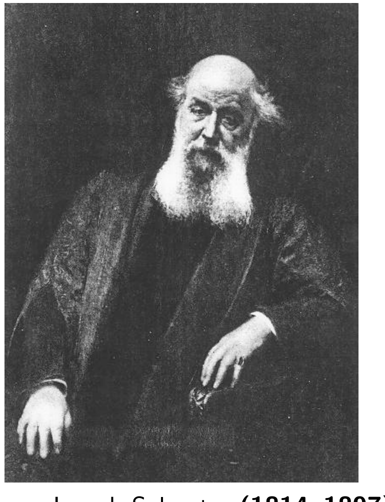
What we have seen so far

For over-determined mixtures:

- Solving the invertible problem is sufficient
- Orthogonal framework:
  - Fully use 2nd order statistics, and then decompose approximately the tensor under *orthogonal constraint*. Easier to handle singularity, but arbitrary.
  - Several (contrast) criteria and algorithms for orthogonal decomposition
- Invertible framework:
  - Decompose the higher order cumulant tensor directly under invertible constraint
  - Ill-posed
  - Several algorithms, mainly working with matrix slices

Principles & algorithms dedicated to under-determined mixtures may apply

Binary case



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

Sylvester's theorem

Sylvester's theorem in $\mathbb{R}$ (1886)

- A binary quantic $t(x_1, x_2) = \sum_{i=0}^{d} c(i)\,\gamma_i\, x_1^i\, x_2^{d-i}$ can be written in $\mathbb{R}[x_1, x_2]$ as a sum of dth powers of r distinct linear forms:

  $t(x_1, x_2) = \sum_{j=1}^{r} \lambda_j\,(\alpha_j\, x_1 + \beta_j\, x_2)^d$ if and only if:

  1. there exists a vector $\mathbf{g}$ of dimension $r+1$ such that

  $$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_r \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{r+1} \\ \vdots & & & \vdots \\ \gamma_{d-r} & & \cdots & \gamma_d \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_r \end{bmatrix} = 0. \qquad (7)$$

  2. $q(x_1, x_2) \stackrel{\text{def}}{=} \sum_{\ell=0}^{r} g_\ell\, x_1^\ell\, x_2^{r-\ell}$ has r distinct real roots

- Then $q(x_1, x_2) \stackrel{\text{def}}{=} \prod_{j=1}^{r}(\beta_j\, x_1 - \alpha_j\, x_2)$ yields the r forms
- Valid even in non generic cases.

Proof of Sylvester's theorem (1)

Lemma

- For homogeneous polynomials of degree d parameterized as $p(\mathbf{x}) \stackrel{\text{def}}{=} \sum_{|\mathbf{i}|=d} c(\mathbf{i})\,\gamma(\mathbf{i}; p)\,\mathbf{x}^{\mathbf{i}}$, define the *apolar scalar product*:

$$\langle p, q\rangle = \sum_{|\mathbf{i}|=d} c(\mathbf{i})\,\gamma(\mathbf{i}; p)\,\gamma(\mathbf{i}; q)$$

- Then $L(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{a}^{\mathsf{T}}\mathbf{x} \Rightarrow \langle p, L^d\rangle = \sum_{|\mathbf{i}|=d} c(\mathbf{i})\,\gamma(\mathbf{i}; p)\,\mathbf{a}^{\mathbf{i}} = p(\mathbf{a})$

Proof of Sylvester's theorem (2)

1. Assume the r distinct linear forms $L_j = \alpha_j x_1 + \beta_j x_2$ are given. Let $q(\mathbf{x}) \stackrel{\text{def}}{=} \prod_{j=1}^{r}(\beta_j x_1 - \alpha_j x_2)$. Then $q(\alpha_j, \beta_j) = 0$, $\forall j$.
2. Hence from lemma, $\forall m(\mathbf{x})$ of degree $d-r$, $\langle mq, L_j^d\rangle = mq(\mathbf{a}_j) = 0$, and $\langle mq, t\rangle = 0$.
3. Take for instance polynomials $m_\mu(\mathbf{x}) = x_1^\mu\, x_2^{d-r-\mu}$, $1 \le \mu \le d-r$, and denote g_ℓ coefficients of q:

$$\langle m_\mu q, t\rangle = 0 \;\Rightarrow\; \sum_{\ell=0}^{r} g_\ell\, \gamma_{\ell+\mu} = 0$$

   This is exactly (7) expressed in canonical basis
4. Roots of $q(x_1, x_2)$ are distinct real since forms L_j are.
5. Reasoning goes also backwards

Algorithm for rth order symmetric tensors of dimension 2

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

Decomposition of maximal rank: $x_1 x_2^{d-1}$

1. Maximal rank $r = d$ when (7) reduces to a 1-row matrix:
$$[0, 0, \ldots 0, 1, 0]\,\mathbf{g} = 0$$
2. Find (α_i, β_i) such that $q(x_1, x_1) = \prod_{j=1}^{d}(\beta_j x_1 - \alpha_j x_2)$ $\stackrel{\text{def}}{=} \sum_{\ell=0}^{d} g_\ell\, x_1^\ell\, x_2^{r-\ell}$ has d distinct roots
3. Take $\alpha_j = 1$. Then $g_{d-1} = 0$ just means $\sum \beta_j = 0$
Choose arbitrarily such distinct β_j's
4. Compute λ_j's by solving the Van der Monde linear system:
$$\begin{bmatrix} 1 & \ldots & 1 \\ \beta_1 & \ldots & \beta_d \\ \beta_1^2 & \ldots & \beta_d^2 \\ \vdots & \vdots & \vdots \\ \beta_1^d & \ldots & \beta_d^d \end{bmatrix} \boldsymbol{\lambda} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Rank of binary quantics

- A binary quantic of odd degree $2n+1$ has *generic rank* $n+1$
- A binary quantic of even degree $2n$ *generic rank* $n+1$
- A binary quantic of degree d may reach *maximal rank* d.
Orbit of maximal rank is $x_1\, x_2^{d-1}$.
- Sylvester's theorem allows to compute the decomposition, even in non generic cases

Problem P16

- In the super generic case, there are infinitely many decompositions
- In dimension larger than 2: simple ways to compute one such decomposition, as in binary case?

Alexander-Hirschowitz theorem

Theorem (1995) For $d > 2$, the generic rank of a dth order symmetric tensor of dimension K is *always* equal to the lower bound
$$\bar{R}_s = \left\lceil \frac{\binom{K+d-1}{d}}{K} \right\rceil \qquad (8)$$
except for the following cases:
$(d, K) \in \{(3, 5), (4, 3), (4, 4), (4, 5)\}$, for which it should be increased by 1 (i.e. only a *finite number* of exceptions, also called *defective* cases)

Values of the Generic Rank (1)

Symmetric tensors of order d and dimension K

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: $\bar{R}_s = \lceil \frac{1}{K}\binom{K+d-1}{d} \rceil$, sometimes called *defective* cases.
Green: lower bound $\frac{1}{K}\binom{K+d-1}{d}$ is integer and nondefective, hence finite number of solutions with proba 1

Values of the Generic Rank (2)

Warning: for *unsymmetric* tensors of order d and dimension K, the generic rank is different

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: $\bar{R} = \lceil \frac{K^d}{Kd-d+1} \rceil$.
Green: lower bound $\frac{K^d}{Kd-d+1}$ is integer and nondefective

Numerical computation of the Generic Rank

Mapping (for unsymmetric tensors):

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

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

- The *smallest r* for wich $rank(\mathrm{Jacobian}(\varphi)) = \prod_i n_i$ is the generic rank, $\bar{R}$.
- Example of use of Terracini's lemma

Example of computation of Generic Rank

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

$\mathbf{T}$ 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:

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

Problem P5

- Similar theorem as AH for unsymmetric tensors?
- that is, decomposition of homogeneous polynomials of degree d but partial degree 1 into sum of products of linear forms

Lectures of Giorgio Ottaviani today...

Genericity in $\mathbb{R}$ vs. $\mathbb{C}$

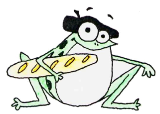
Define $\mathcal{Z}_r = \{\text{tensors of rank } r\}$

- A rank r is *typical* if $\mathcal{Z}_r$ is Zariski-dense
- In the complex field, there is only one typical rank, called the *generic rank*.
- In the real field, there can be *several* typical ranks (smallest equals generic rank in $\mathbb{C}$)

Problem P10

Case of the Real field

- Similar result as AH theorem in the real field?
- i.e. values of typical ranks for any order and dimensions



Low rank approximation

We need

1. to know the exact rank
   and
2. the rank to be sub-generic

Hence we make suboptimal rank reduction

- by a 2-stage rank reduction
  or
- by HOSVD truncation

Two-stage suboptimal rank reduction (1)

1. Associate one *linear* operator with tensor $\mathbf{T}$, defined by a matrix $\mathbf{M}$
2. Compute the best rank r approximate of $\mathbf{M}$, e.g. via truncated SVD *(not always possible)*
3. Unfold each of the r singular vectors into a matrix, and compute its rank-1 approximate
4. From this starting point, run a few iterations of a descent on

$$||\mathbf{T} - \sum_{p=1}^{r} \mathbf{u}_p \otimes \mathbf{v}_p \otimes \ldots \otimes \mathbf{w}_p||^2$$

Two-stage suboptimal rank reduction (2)

Example: If $\mathbf{T}$ is $K \times K \times K \times K$ symmetric

1. build the symmetric matrix $\mathbf{M}$ of size $K^2 \times K^2$.
2. Compute the r dominant eigenvectors $\mathbf{e}_p$ of $\mathbf{M}$.
3. We wish each $\mathbf{e}$ to be of the form $\mathbf{u} \otimes \mathbf{u}$.
   Hence minimize $||\mathbf{Unvec}_K(\mathbf{e}_p) - \mathbf{u}_p \, \mathbf{u}_p^{\mathsf{T}}||^2$ via rank-1 approximates
   Eventually:

$$\mathbf{T} \approx \sum_{p=1}^{r} \mathbf{u}_p \otimes \mathbf{u}_p \otimes \mathbf{u}_p \otimes \mathbf{u}_p$$

HOSVD s

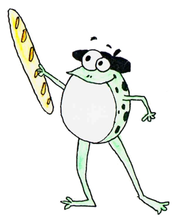
For a tensor of order d, compute first a rank-$(R_1, R_2, \ldots, R_d)$ approximate:

1. Associate d *linear* operators with tensor $\mathbf{T}$, defined by unfolding matrices $\mathbf{M}_i$, $1 \le i \le d$
2. Compute the rank-R_i approximation of each matrix $\mathbf{M}_i$ (reduction of the multilinear *mode-i ranks*)
3. Use truncated left singular marices $\mathbf{U}^{(i)}$ to compute the $R_1 \times R_2 \times \ldots R_d$ core tensor $\mathbf{T}_0$
4. Run an iterative descent on $||\mathbf{T}_0 - \sum_{p=1}^{r} \mathbf{u}_p \otimes \mathbf{v}_p \otimes \ldots \otimes \mathbf{w}_p||^2$

Problem P4

Approximation of a tensor by another of lower rank

1. ill-posed in general for free real/complex entries
2. other problem statement (e.g. border rank)?
3. case of positive entries
4. case of semi-definite positive quantic
5. other cases?



Algorithms when $K > 2$

Now survey some numerical algorithms to compute the decomposition:

- when rank is strictly smaller than generic
- when dimension is larger than 2
- suboptimality (symmetry not fully imposed - link with P15)

BIOME algorithms

- These algorithms work with a cumulant tensor of even order $2r > 4$
- Related to *symmetric flattening* introduced in previous lectures
- We take the case $2r = 6$ for the presentation, and denote

$$\mathcal{C}_{ijk}^{\ell mn} \overset{\text{def}}{=} \mathrm{Cum}\{x_i, x_j, x_k, x_l^*, x_m^*, x_n^*\} \qquad (9)$$

- 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 $\Delta_p \overset{\text{def}}{=} \mathrm{Cum}\{s_p, s_p, s_p, s_p^*, s_p^*, s_p^*\}$ denote the diagonal entries of a $P \times P$ diagonal matrix, $\boldsymbol{\Delta}^{(6)}$

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

$$\mathbf{C}_x^{(6)} = \mathbf{H}^{\odot 3}\, \boldsymbol{\Delta}^{(6)}\, \mathbf{H}^{\odot 3\mathsf{H}} \qquad (10)$$

- Because $\mathbf{C}_x^{(6)}$ is Hermitian, $\exists \mathbf{V}$ unitary, such that

$$(\mathbf{C}_x^{(6)})^{1/2} = \mathbf{H}^{\odot 3}\, (\boldsymbol{\Delta}^{(6)})^{1/2}\, \mathbf{V} \qquad (11)$$

- **Idea:** Use redundancy existing between blocks of $(\mathbf{C}_x^{(6)})^{1/2}$.

Using the invariance to estimate V

1. 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, $\boldsymbol{\Gamma}[n]$, satisfies:

$$\boldsymbol{\Gamma}[n] = (\mathbf{H} \odot \mathbf{H}^{\mathsf{H}})\mathbf{D}[n]\, (\boldsymbol{\Delta}^{(6)})^{1/2}\, \mathbf{V}$$

   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 $\boldsymbol{\Gamma}[n]$ share the same common right singular space
2. Compute the joint EVD of the $K(K-1)$ matrices

$$\boldsymbol{\Theta}[m, n] \overset{\text{def}}{=} \boldsymbol{\Gamma}[m]^{-}\boldsymbol{\Gamma}[n]$$

   as: $\boldsymbol{\Theta}[m, n] = \mathbf{V}\, \boldsymbol{\Lambda}[m, n]\, \mathbf{V}^{\mathsf{H}}$.

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:

$$\widehat{\mathbf{H}^{\odot 3}} = (\mathbf{C}_x^{(6)})^{1/2}\,\mathbf{V} \tag{12}$$

One possibility to get $\mathbf{H}$ from $\mathbf{H}^{\odot 3}$ is as follows:

3. Build K^2 matrices $\mathbf{\Xi}[m]$ of size $K \times P$ form rows of $\widehat{\mathbf{H}^{\odot 3}}$
4. From $\mathbf{\Xi}[m]$ find $\widehat{\mathbf{H}}$ and diagonal matrices $\mathbf{D}[m]$, in the LS sense:
$$\mathbf{\Xi}[m]\,\mathbf{D}[m] \approx \widehat{\mathbf{H}}, \quad 1 \le m \le K^2$$

Estimation of $\mathbf{H}$ (details) s

Stationary values of criterion $\sum_{m=1}^{M} \|\mathbf{\Xi}_m \mathbf{D}_m - \mathbf{H}\|_F^2$, $M \stackrel{\text{def}}{=} K^2$, yield the solution below

- Obtain vectors $\mathbf{d}_p \stackrel{\text{def}}{=} [\mathbf{D}_1(p,p)\,,\, \mathbf{D}_2(p,p)\,,\, \cdots\, \mathbf{D}_M(p,p)]^{\mathsf{T}}$, by solving the linear systems:
$$\mathbf{F}_p\,\mathbf{d}_p = 0$$
where matrices $\mathbf{F}_p$ are defined as
$$\mathbf{F}_p(m_1,m_2) = \begin{cases} (M-1)\left\{\mathbf{\Xi}_{m_1}^{\mathsf{H}}\mathbf{\Xi}_{m_1}\right\}(p,p) & \text{if } m_1 = m_2 \\ -\left\{\mathbf{\Xi}_{m_1}^{\mathsf{H}}\mathbf{\Xi}_{m_2}\right\}(p,p) & \text{otherwise} \end{cases}$$
- Deduce the estimate $\widehat{\mathbf{H}} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{\Xi}_m \mathbf{D}_m$

Conditions of identifiability of BIOME($2r$) s

- Source cumulants of order $2r > 4$ are nonzero and have the same sign
- Columns vectors of mixing matrix $\mathbf{H}$ are not collinear
- Matrix $\mathbf{H}^{\odot(r-1)}$ is full column rank.
- This last condition implies that tensor rank must be at most K^{r-1} (*e.g.* $P \le K^2$ for order $2r = 6$).

FOOBI algorithms

Again same problem: Given a $K^2 \times P$ matrix $\mathbf{H}^{\odot 2}$, find a real orthogonal matrix $\mathbf{Q}$ such that the P columns of $\mathbf{H}^{\odot 2}\,\mathbf{Q}$ are of the form $\mathbf{h}[p] \otimes \mathbf{h}[p]^*$

- **FOOBI:** use the K^4 determinantal equations characterizing rank-1 matrices $\mathbf{h}[p]\,\mathbf{h}[p]^{\mathsf{H}}$ of the form: $\phi(\mathbf{X},\mathbf{Y})_{ijk\ell} = x_{ij}y_{\ell k} - x_{ik}y_{\ell j} + y_{ij}x_{\ell k} - y_{ik}x_{\ell j}$
- **FOOBI2:** use the K^2 equations of the form: $\Phi(\mathbf{X},\mathbf{Y}) = \mathbf{X}\,\mathbf{Y} + \mathbf{Y}\,\mathbf{X} + \text{trace}\{\mathbf{X}\}\mathbf{Y} + \text{trace}\{\mathbf{Y}\}\mathbf{X}$ where matrices $\mathbf{X}$ and $\mathbf{Y}$ are $K \times K$ Hermitean.

FOOBI

1. Normalize the columns of the $K^2 \times P$ matrix $\mathbf{H}^{\odot 2}$ such that matrices $\mathbf{H}[r] \stackrel{\text{def}}{=} \mathbf{Unvec}_K\left(\mathbf{h}^{\odot 2}[r]\right)$ are Hermitean
2. Compute the $K^2(K-1)^2 \times P(P-1)/2$ matrix $\mathbf{P}$ defined by $\phi(\mathbf{H}[r], \mathbf{H}[s])$, $1 \leq r \leq s \leq P$.
3. Compute the P weakest right singular vectors of $\mathbf{P}$, Unvec them and store them in P matrices $\mathbf{W}[r]$
4. Jointly diagonalize $\mathbf{W}[r]$ by a real orthogonal matrix $\mathbf{Q}$
5. Then compute $\mathbf{F} \stackrel{\text{def}}{=} (\mathbf{C}_x^{(4)})^{1/2}\mathbf{\Delta Q}$ and deduce $\hat{\mathbf{h}}[r]$ as the dominant left singular vectors of $\mathbf{Unvec}\,(\mathbf{f}[r])$.

FOOBI2

1. Normalize the columns of the $K^2 \times P$ matrix $\mathbf{H}^{\odot 2}$ such that matrices $\mathbf{H}[r] \stackrel{\text{def}}{=} \mathbf{Unvec}_K\left(\mathbf{h}^{\odot 2}[r]\right)$ are Hermitean
2. Compute the $K(K+1)/2$ Hermitean matrix $\mathbf{B}[r,s]$ of size $P \times P$ defined by:

$$\Phi(\mathbf{H}[r], \mathbf{H}[s])|_{ij} \stackrel{\text{def}}{=} \mathbf{B}[i,j]|_{rs}$$

3. Jointly cancel diagonal entries of matrices $\mathbf{B}[i,j]$ by a real congruent orthogonal transform $\mathbf{Q}$
4. Then compute $\mathbf{F} \stackrel{\text{def}}{=} (\mathbf{C}_x^{(4)})^{1/2}\mathbf{\Delta Q}$ and deduce $\hat{\mathbf{h}}[r]$ as the dominant left singular vectors of $\mathbf{Unvec}\,(\mathbf{f}[r])$.

NB: Better bound than FOOBI and BIOME(4), but iterative algorithm sensitive to initialization

Algorithms based on characteristic functions

Fit with a model of exact rank

1. Back to the core equation (3):

$$\Psi_x(\mathbf{u}) = \sum_p \Psi_{s_p}\left(\sum_q u_q\, A_{qp}\right)$$

2. **Goal:** Find a matrix $\mathbf{H}$ such that the $K-$variate function $\Psi_x(\mathbf{u})$ decomposes into a sum of P univariate functions $\psi_p \stackrel{\text{def}}{=} \Psi_{s_p}$.
3. **Idea:** Fit both sides on a grid of values $\mathbf{u}[\ell] \in \mathcal{G}$

Equations derived from the CAF

- Assumption: functions ψ_p, $1 \leq p \leq P$ admit finite derivatives up to order r in a neighborhood of the origin, containing $\mathcal{G}$.
- Then, Taking $r = 3$ as a working example:

$$\frac{\partial^3 \Psi_x}{\partial u_i \partial u_j \partial u_k}(\mathbf{u}) = \sum_{p=1}^{P} H_{ip}\, H_{jp}\, H_{kp}\;\; \psi_p^{(3)}\Big(\sum_{q=1}^{K} u_q\, H_{qp}\Big)$$

- If $L > 1$ point in grid $\mathcal{G}$, then yields another mode in tensor

Putting the problem in tensor form

- A decomposition into a sum of rank-1 terms:

$$T_{ijk\ell} = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, B_{\ell p}$$

or equivalently

$$\mathbf{T} = \sum_p \mathbf{h}(p) \otimes \mathbf{h}(p) \otimes \mathbf{h}(p) \otimes \mathbf{b}(p)$$

- Tensor $\mathbf{T}$ is $K \times K \times K \times L$, symmetric in all modes except the last.

Joint use of different derivative orders

Example

- Derivatives of order 3:

$$T^{(3)}_{ijk\ell} = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, B_{\ell p}$$

- Derivatives of order 4:

$$T^{(4)}_{ijkm\ell} = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, H_{mp}\, C_{\ell p}$$

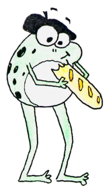
- Derivatives of orders 3 and 4:

$$T_{ijk\ell}[m] = \sum_p H_{ip}\, H_{jp}\, H_{kp}\, D_{\ell p}[m]$$

with $D_{\ell p}[m] = H_{mp}\, C_{\ell p}$ and $D_{\ell p}[0] = B_{\ell p}$.

Problem P13

- Results for tensors enjoying partial symmetries
- e.g., symmetric in the 3 first modes and not not in others
- e.g. tensors with symmetries in some modes and Hermitean symmetries in others
- Generic/typical ranks?
- Existence/well-poseness?
- Uniqueness?



Problem P18

What is best to do when *several* symmetric tensors, possibly of different orders, are supposed to be decomposed with *same* matrix $\mathbf{H}$?

Iterative algorithms

Many practitioners execute more or less brute force minimizations of $||\mathbf{T} - \sum_{p=1}^{r} \mathbf{u}_p \otimes \mathbf{v}_p \otimes \ldots \otimes \mathbf{w}_p||^2$

- Gradient with fixed or variable (ad-hoc) stepsize
- Alternate Least Squares (ALS)
- Levenberg-Marquardt
- Newton
- Conjugate gradient
- ...

Remarks

- Hessian is generally huge, but sparse
- Problem of *local minima*: ELS variants for all of the above

What we have seen so far

- In dimension 2, we are able to handle decompositions of symmetric tensors algebraically
- Suitable to extend this to higher dimensions (now partly done for sub-generic cases)
- Work remains for generic case or above
- A lot to do for unsymmetric tensors
- Problem of tensors enjoying some symmetries
- Collection of tensors sharing common terms is their decompositions

Summary of the lectures

- We want to know when the decomposition of a tensor is unique
- In the latter cases, we want to be able to compute the decomposition
- We have suboptimal ways of doing it in some cases