Tensor decompositions in statistical signal processing

Pierre Comon

I3S, CNRS, University of Nice, Sophia-Antipolis, France

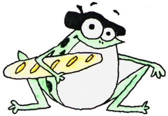
July 2, 2009

Linear statistical model

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

with

$$\begin{cases} \mathbf{y}: & K \times 1 \text{ random} \\ \mathbf{s}: & P \times 1 \text{ random } \textit{stat. independent} \\ \mathbf{A}: & K \times P \text{ deterministic} \\ \mathbf{b}: & \text{errors} \end{cases}$$

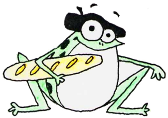


Linear statistical model

$$\mathbf{y} = \mathbf{A}\,\mathbf{s} + \mathbf{b} \quad (1)$$

with

$$\left\{ \begin{array}{lll} \mathbf{y}: & K \times 1 \text{ random} & \text{"sensors"} \\ \mathbf{s}: & P \times 1 \text{ random } \textit{stat. independent} & \text{"sources"} \\ \mathbf{A}: & K \times P \text{ deterministic} & \text{"mixing matrix"} \\ \mathbf{b}: & \text{errors} & \text{"noise"} \end{array} \right.$$

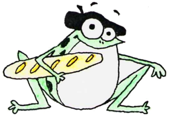


Linear statistical model

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

with

$$\begin{cases} \mathbf{y}: & K \times 1 \text{ random} & \text{“sensors”} \\ \mathbf{s}: & P \times 1 \text{ random } \textit{stat. independent} & \text{“sources”} \\ \mathbf{A}: & K \times P \text{ deterministic} & \text{“mixing matrix”} \\ \mathbf{b}: & \text{errors} & \text{“noise”} \\ & \text{(may be removed for } P \text{ large enough)} & \end{cases}$$



Other writing

$$\mathbf{y} = [\mathbf{A},\ \mathbf{I}] \begin{bmatrix} \mathbf{s} \\ \mathbf{b} \end{bmatrix}$$

Other writing

$$\mathbf{y} = [\mathbf{A},\ \mathbf{I}] \begin{bmatrix} \mathbf{s} \\ \mathbf{b} \end{bmatrix}$$

which is of the *noiseless* form

$$\mathbf{y} = \mathbf{A}\,\mathbf{s} \tag{2}$$

with a *larger* dimension P

Taxinomy

Problems addressed in the literature:

Taxinomy

Problems addressed in the literature:

- $K \geq P$: "over-determined"

  can be reduced to a square $P \times P$ regular mixture
  - **A** orthogonal or unitary
  - **A** square invertible

Taxinomy

Problems addressed in the literature:

- $K \geq P$: "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

Taxinomy

Problems addressed in the literature:

- $K \geq P$: "over-determined" $\rightarrow$ Approximation Pb
  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

Taxinomy

Problems addressed in the literature:

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

Goals

Solely from realizations of observation vector $\mathbf{y}$

Goals

Solely from realizations of observation vector $\mathbf{y}$

In the stochastic framework:

Goals

Solely from realizations of observation vector $\mathbf{y}$

In the stochastic framework:

- Estimate matrix $\mathbf{A}$: Blind identification

Goals

Solely from realizations of observation vector $\mathbf{y}$

In the stochastic framework:

- Estimate matrix $\mathbf{A}$: Blind identification
- Estimate realizations of the "source" vector $\mathbf{s}$: Blind separation/extraction/equalization

Goals

Solely from realizations of observation vector $\mathbf{y}$

In the stochastic framework:

- Estimate matrix $\mathbf{A}$: Blind identification
- Estimate realizations of the "source" vector $\mathbf{s}$: Blind separation/extraction/equalization

In the deterministic framework:

Goals

Solely from realizations of observation vector $\mathbf{y}$

In the stochastic framework:

- Estimate matrix $\mathbf{A}$: Blind identification
- Estimate realizations of the "source" vector $\mathbf{s}$: Blind separation/extraction/equalization

In the deterministic framework:

- Build a data tensor $\mathcal{Y}$ thanks to a diversity k, and decompose it:

$$\mathcal{Y}_{ijk} = \sum_p A_{ip} s_{jp} B_{kp}$$

Goals

Solely from realizations of observation vector $\mathbf{y}$

In the stochastic framework:

- Estimate matrix $\mathbf{A}$: Blind identification
- Estimate realizations of the "source" vector $\mathbf{s}$: Blind separation/extraction/equalization

In the deterministic framework:

- Build a data tensor $\mathcal{Y}$ thanks to a diversity k, and decompose it:

$$\mathcal{Y}_{ijk} = \sum_p A_{ip} s_{jp} B_{kp}$$

- $\mathbf{A}$ and $\mathbf{s}$ are jointly estimated

Application areas

1. Telecommunications (Cellular, Satellite, Military),
2. Radar, Sonar,
3. Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...
4. Speech, Audio,
5. Machine learning,
6. Data mining,
7. Control...

Identifiability & Uniqueness

- Uniqueness/Identifiability up to inherent ambiguities
- Finite number of solutions

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.

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.

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 & 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.

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.

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

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

Characteristic functions

First c.f.

Characteristic functions

First c.f.

- Real Scalar: $\Phi_x(t) \stackrel{\text{def}}{=} \mathbb{E}\{e^{\imath\, tx}\} = \int_u e^{\imath\, tu}\, dF_x(u)$

Characteristic functions

First c.f.

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

Characteristic functions

First c.f.

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

Second c.f.

Characteristic functions

First c.f.

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

Second c.f.

- $\Psi(\mathbf{t}) \stackrel{\text{def}}{=} \log \Phi(\mathbf{t})$

Characteristic functions

First c.f.

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

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):

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)*

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)*

Proof.

complicated... □

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)*

Proof.

complicated... □

  - if (x, y) statistically independent, then

$$\Psi_{x,y}(u, v) = \Psi_x(u) + \Psi_y(v) \tag{3}$$

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)*

Proof.

complicated... □

  - if (x, y) statistically independent, then

$$\Psi_{x,y}(u,v) = \Psi_x(u) + \Psi_y(v) \tag{3}$$

Proof.

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

□

Problem posed in terms of Characteristic Functions

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

$$\Psi_y(\mathbf{u}) = \sum_p \psi_p\left(\sum_q u_q\, A_{qp}\right) \tag{4}$$

where $\psi_p(w) \stackrel{\text{def}}{=} \log \mathbb{E}\{\exp \imath(s_p\, w)\}$.

Proof.

Problem posed in terms of Characteristic Functions

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

$$\Psi_y(\mathbf{u}) = \sum_p \psi_p \left(\sum_q u_q\, A_{qp} \right) \tag{4}$$

where $\psi_p(w) \stackrel{\text{def}}{=} \log \mathbb{E}\{\exp \imath(s_p\, w)\}$.

Proof.

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

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

Problem posed in terms of Characteristic Functions

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

$$\Psi_y(\mathbf{u}) = \sum_p \psi_p\left(\sum_q u_q\, A_{qp}\right) \tag{4}$$

where $\psi_p(w) \stackrel{\text{def}}{=} \log \mathbb{E}\{\exp \imath(s_p\, w)\}$.

Proof.

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

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

- Since s_p independent, $\phi_y(\mathbf{u}) = \prod_p \mathbb{E}\{\exp \imath(\sum_q u_q\, A_{qp}\, s_p)\}$

Function decomposition

Problem:

Equation (4) shows that we have to decompose a *mutlivariate* function into a sum of *univariate* ones.

Function decomposition

Problem:

Equation (4) shows that we have to decompose a *mutlivariate* function into a sum of *univariate* ones.

Weierstrass (1885), Stone (1948), Hilbert (1900), Kolmogorov (1957), Sprecher (1965), Hornik (1989), Cybenko (1989)...

- But here, Ψ_y and ψ_p's are characteristic functions.

Problem seen as non symmetric tensor decomposition

Back to core equation (4):

$$\Psi_y(\mathbf{u}) = \sum_p \psi_p \left(\sum_q u_q \, A_{qp} \right), \; \mathbf{u} \in \mathcal{G}$$

- Assumption: functions ψ_p, $1 \leq p \leq P$ admit finite derivatives up to order d in a neighborhood of the origin, containing $\mathcal{G}$.

Problem seen as non symmetric tensor decomposition

Back to core equation (4):

$$\Psi_y(\mathbf{u}) = \sum_p \psi_p \left(\sum_q u_q \, A_{qp} \right), \; \mathbf{u} \in \mathcal{G}$$

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

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

Problem seen as non symmetric tensor decomposition

Back to core equation (4):

$$\Psi_y(\mathbf{u}) = \sum_p \psi_p \left(\sum_q u_q \, A_{qp} \right), \ \mathbf{u} \in \mathcal{G}$$

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

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

- of the form $T_{ijk\ell} = \sum_p A_{ip} \, A_{jp} \, A_{kp} \, F_{\ell p}$, if L points $\mathbf{u}_\ell \in \mathcal{G}$.

Problem seen as symmetric tensor decomposition (1/2)

If only the origin in $\mathcal{G}$, *i.e.* $\mathbf{u}_\ell = \mathbf{0}$, then

$$C_{ijk} = \sum_p A_{ip}\, A_{jp}\, A_{kp}\, f_p$$

Multi-linear relation relating *cumulants*.

Problem seen as symmetric tensor decomposition (2/2)

- **Equivalent writing** (still with $d = 3$ as working example)

$$\mathbf{C} = \sum_{p=1}^{P} f_p \, \mathbf{h}(p) \otimes \mathbf{h}(p) \otimes \mathbf{h}(p)$$

where $\mathbf{h}(p) \overset{\text{def}}{=}$ pth column of $\mathbf{A}$.

Problem seen as symmetric tensor decomposition (2/2)

- **Equivalent writing** (still with $d = 3$ as working example)

$$\mathbf{C} = \sum_{p=1}^{P} f_p \, \mathbf{h}(p) \otimes \mathbf{h}(p) \otimes \mathbf{h}(p)$$

where $\mathbf{h}(p) \stackrel{\text{def}}{=} p$th column of $\mathbf{A}$.

- If vectors $\mathbf{h}(p)$ are not pairwise collinear, P is hopefully minimal, *i.e.* equal to the *rank* of tensor $\mathbf{C}$

Problem seen as symmetric tensor decomposition (2/2)

- **Equivalent writing** (still with $d = 3$ as working example)

$$\mathbf{C} = \sum_{p=1}^{P} f_p \, \mathbf{h}(p) \otimes \mathbf{h}(p) \otimes \mathbf{h}(p)$$

  where $\mathbf{h}(p) \stackrel{\text{def}}{=} p$th column of $\mathbf{A}$.

- If vectors $\mathbf{h}(p)$ are not pairwise collinear, P is hopefully minimal, *i.e.* equal to the *rank* of tensor $\mathbf{C}$ → condition

Expected rank

For an order-d symmetric tensor of dimension K

- Number of equations: $\binom{K+d-1}{d}$
  Number of unknowns: PK
  For what P can one have an exact fit?

Expected rank

For an order-d symmetric tensor of dimension K

- Number of equations: $\binom{K+d-1}{d}$
  Number of unknowns: PK
  For what P can one have an exact fit?
- "Expected rank":

$$R(K,d) \stackrel{\text{def}}{=} \left\lceil \frac{\binom{K+d-1}{d}}{K} \right\rceil \tag{5}$$

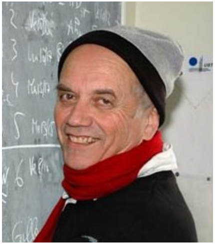
Clebsch's statement

Alfred Clebsch (1833-1872)

For $(d, K) = (4, 3)$, a generic tensor cannot be written as the sum of 5 rank-1 terms, even if $\#unknowns = 15 = \#equations$

Hirschowitz theorem

From Alexander-Hirschowitz thm (1995), one deduces [CGLM08]:



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

$$\bar{R}_s = R(K, d) \tag{6}$$

except for $(d, K) \in \{(3, 5), (4, 3), (4, 4), (4, 5)\}$

➡ Only a *finite number* of exceptions !

Identifiability

Identifiability in wider sense: *finite number* of solutions

Identifiability

Identifiability in wider sense: *finite number* of solutions

- *Necessary* condition for identifiability: $P \leq R(K, d)$.
  Not sufficient, cf. Clebsch (1861), Hirschowitz (1995)

Identifiability

Identifiability in wider sense: *finite number* of solutions

- *Necessary* condition for identifiability: $P \leq R(K, d)$.
  Not sufficient, cf. Clebsch (1861), Hirschowitz (1995)
- *Sufficient* condition for *almost sure* identifiability:
  $P < R(K, d)$.

Identifiability

Identifiability in wider sense: *finite number* of solutions

- *Necessary* condition for identifiability: $P \leq R(K, d)$.
  Not sufficient, cf. Clebsch (1861), Hirschowitz (1995)
- *Sufficient* condition for *almost sure* identifiability:
  $P < R(K, d)$.
- *NS* condition for *almost sure* identifiability:

$$P = \frac{\binom{K+d-1}{d}}{K} = R(K, d) = \bar{R}_s$$

Generic rank of symmetric tensors

Symmetric tensors of order d and dimension K:

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

Deterministic approaches

- Model:

$$Y_{ijk} = \sum_p A_{ip}\, s_{jp}\, B_{kp}$$

- $\mathbf{Y} = \sum_{p=1}^{P} \mathbf{a}(p) \otimes \mathbf{s}(p) \otimes \mathbf{b}(p)$

Expected rank

For an order-d symmetric tensor of dimensions K_ℓ, $1 \leq \ell \leq d$

- Number of equations: $\prod_\ell K_\ell$
  Number of unknowns: $\sum_\ell K_\ell P - (d-1)P$

Expected rank

For an order-d symmetric tensor of dimensions K_ℓ, $1 \leq \ell \leq d$

- Number of equations: $\prod_\ell K_\ell$
  Number of unknowns: $\sum_\ell K_\ell P - (d-1)P$
- "Expected rank" again given by the ceil rule:

$$R(K_1, .., K_d) \stackrel{\text{def}}{=} \left\lceil \frac{\prod_\ell K_\ell}{\sum_\ell K_\ell - d + 1} \right\rceil \tag{7}$$

Identifiability

- *Necessary* condition for identifiability:

$$P \leq R(K_1, .., K_d)$$

Identifiability

- *Necessary* condition for identifiability:

$$P \leq R(K_1, .., K_d)$$

- *Sufficient* condition for *almost sure* identifiability:

$$P < R(K_1, .., K_d)$$

Identifiability

- *Necessary* condition for identifiability:

$$P \leq R(K_1, .., K_d)$$

- *Sufficient* condition for *almost sure* identifiability:

$$P < R(K_1, .., K_d)$$

- For general tensors, generic rank $\bar{R}$ not yet known everywhere.

Generic rank of order 3 free tensors (1)

	K	2				3			4	
	J	2	3	4	5	3	4	5	4	5
I	2	2	3	4	4	3	4	5	4	5
	3	3	3	4	5	5	5	5	6	6
	4	4	4	4	5	5	6	6	7	8
	5	4	5	5	5	5	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	10	11
	9	4	6	8	9	9	9	9	10	12
	10	4	6	8	10	9	10	10	10	12
	11	4	6	8	10	9	11	11	11	13
	12	4	6	8	10	9	12	12	12	13

Perspectives

- Efficient algorithms to compute finite set of solutions

Perspectives

- Efficient algorithms to compute finite set of solutions
- Detection of weak defectivity: rare cases when ∞ many solutions

Perspectives

- Efficient algorithms to compute finite set of solutions
- Detection of weak defectivity: rare cases when ∞ many solutions
- Construction of additional diversity: increases order

Perspectives

- Efficient algorithms to compute finite set of solutions
- Detection of weak defectivity: rare cases when ∞ many solutions
- Construction of additional diversity: increases order
- **Thanks for your attention**

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 $\mathbb{C}$

Proof of Darmois-Skitovich thm

Assume $[a_k, \ b_k]$ not collinear and ψ_p differentiable.

Proof of Darmois-Skitovich thm

Assume $[a_k,\ b_k]$ not collinear and ψ_p differentiable.

1 Hence $\sum_{k=1}^{P} \psi_p(u\, a_k + v\, b_k) = \sum_{k=1}^{P} \psi_k(u\, a_k) + \psi_k(v\, b_k)$
Trivial for terms for which $a_k b_k = 0$.
From now on, restrict the sum to terms $a_k b_k \neq 0$

Proof of Darmois-Skitovich thm

Assume $[a_k,\ b_k]$ not collinear and ψ_p differentiable.

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

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

Proof of Darmois-Skitovich thm

Assume $[a_k,\ b_k]$ not collinear and ψ_p differentiable.

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

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

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

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

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

Proof of Darmois-Skitovich thm

Assume $[a_k,\ b_k]$ not collinear and ψ_p differentiable.

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

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

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

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

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

Conclusion: We have one term less

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

$$\prod_{j=2}^{P}(\frac{a_1}{a_j}-\frac{b_1}{b_j})\,\psi_1^{(P-1)}(u\,a_1+v\,b_1)=f^{(P-1)}(u)+g^{(P-1)}(v)$$

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

$$\prod_{j=2}^{P}(\frac{a_1}{a_j}-\frac{b_1}{b_j})\,\psi_1^{(P-1)}(u\,a_1+v\,b_1)=f^{(P-1)}(u)+g^{(P-1)}(v)$$

5. Hence $\psi_1^{(P-1)}(u\,a_1+v\,b_1)$ is linear, as a sum of two univariate functions ($\psi_1^{(P)}$ is a constant because $a_1 b_1 \neq 0$).

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

$$\prod_{j=2}^{P}\left(\frac{a_1}{a_j}-\frac{b_1}{b_j}\right)\psi_1^{(P-1)}(u\,a_1+v\,b_1)=f^{(P-1)}(u)+g^{(P-1)}(v)$$

5 Hence $\psi_1^{(P-1)}(u\,a_1+v\,b_1)$ is linear, as a sum of two univariate functions ($\psi_1^{(P)}$ is a constant because $a_1 b_1 \neq 0$).

6 Eventually ψ_1 is a polynomial.

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

$$\prod_{j=2}^{P}\left(\frac{a_1}{a_j}-\frac{b_1}{b_j}\right)\psi_1^{(P-1)}(u\,a_1+v\,b_1)=f^{(P-1)}(u)+g^{(P-1)}(v)$$

5. Hence $\psi_1^{(P-1)}(u\,a_1+v\,b_1)$ is linear, as a sum of two univariate functions ($\psi_1^{(P)}$ is a constant because $a_1 b_1 \neq 0$).
6. Eventually ψ_1 is a polynomial.
7. Lastly invoke Marcinkiewicz theorem to conclude that s_1 is Gaussian.

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

$$\prod_{j=2}^{P}\left(\frac{a_1}{a_j}-\frac{b_1}{b_j}\right)\psi_1^{(P-1)}(u\,a_1+v\,b_1)=f^{(P-1)}(u)+g^{(P-1)}(v)$$

5. Hence $\psi_1^{(P-1)}(u\,a_1+v\,b_1)$ is linear, as a sum of two univariate functions ($\psi_1^{(P)}$ is a constant because $a_1 b_1 \neq 0$).
6. Eventually ψ_1 is a polynomial.
7. Lastly invoke Marcinkiewicz theorem to conclude that s_1 is Gaussian.
8. Same is true for any ψ_p such that $a_p b_p \neq 0$: s_p is Gaussian.

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

$$\prod_{j=2}^{P}\left(\frac{a_1}{a_j}-\frac{b_1}{b_j}\right)\psi_1^{(P-1)}(u\,a_1+v\,b_1)=f^{(P-1)}(u)+g^{(P-1)}(v)$$

5. Hence $\psi_1^{(P-1)}(u\,a_1+v\,b_1)$ is linear, as a sum of two univariate functions ($\psi_1^{(P)}$ is a constant because $a_1 b_1 \neq 0$).
6. Eventually ψ_1 is a polynomial.
7. Lastly invoke Marcinkiewicz theorem to conclude that s_1 is Gaussian.
8. Same is true for any ψ_p such that $a_p b_p \neq 0$: s_p is Gaussian.

NB: also holds if ψ_p not differentiable

Definition of Cumulants

- Moments:

$$\mu_r \stackrel{\text{def}}{=} \mathbb{E}\{x^r\} = (-\imath)^r \left. \frac{\partial^r \Phi(t)}{\partial t^r} \right|_{t=0}$$

Definition of Cumulants

- Moments:

$$\mu_r \stackrel{\text{def}}{=} \mathbb{E}\{x^r\} = (-\imath)^r \left. \frac{\partial^r \Phi(t)}{\partial t^r} \right|_{t=0}$$

- Cumulants:

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

Definition of Cumulants

- Moments:

$$\mu_r \stackrel{\text{def}}{=} \mathbb{E}\{x^r\} = (-\imath)^r \left. \frac{\partial^r \Phi(t)}{\partial t^r} \right|_{t=0}$$

- Cumulants:

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

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

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

Polynomial interpolation

Alexander-Hirschowitz Theorem (1995) 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.

References

F. L. HITCHCOCK. “Multiple invariants and generalized rank of a p-way matrix or tensor.” *J. Math. and Phys.*, 7(1):39–79, 1927.

L. R. TUCKER. “Some mathematical notes for three-mode factor analysis.” *Psychometrika*, 31:279–311, 1966.

J. B. KRUSKAL. “Three-way arrays: Rank and uniqueness of trilinear decompositions.” *Linear Algebra and Applications*, 18:95–138, 1977.

V. STRASSEN. “Rank and optimal computation of generic tensors.” *Linear Algebra Appl.*, 52:645–685, July 1983.

P. COMON and B. MOURRAIN. “Decomposition of quantics in sums of powers of linear forms.” *Signal Processing, Elsevier*, 53(2):93–107, September 1996.

N. D. SIDIROPOULOS and R. BRO. “On the uniqueness of multilinear decomposition of N-way arrays.” *J. Chemometrics*, 14:229–239, 2000.

P. COMON. “Tensor decompositions.” In J. G. McWhirter and I. K. Proudler eds, *Mathematics in Signal Processing V*, pages 1–24. Clarendon Press, Oxford, UK, 2002.

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

P. COMON, G. GOLUB, L-H. LIM, and B. MOURRAIN. “Symmetric tensors and symmetric tensor rank.” *SIAM Journal on Matrix Analysis Appl.*, 30(3):1254–1279, 2008.

P. COMON, J. M. F. ten BERGE, L. De LATHAUWER and J. CASTAING. “Generic and Typical Ranks of Multi-Way Arrays.” *Linear Algebra Appl.*, 2009. See also http://arxiv.org/abs/0802.2371

P. COMON, X. LUCIANI and A. de ALMEIDA. “Tensor decompositions, alternating least squares and other

References

Participants in ANR-06-BLAN-0074 “DECOTES” project:

- Signal team, Lab. I3S UMR6070, Univ. Nice and CNRS
- Galaad team, INRIA, UR Sophia-Antipolis - Méditerranée.
- Lab. LTSI U642, INSERM and Univ. Rennes 1
- Thalès Communications