Tensor decompositions in Engineering

Pierre Comon

July 21, 2008
Canonical decomposition

**Goal:** Given a tensor $\mathbf{T}$ of order $d$, defined in space $V_1 \otimes \ldots \otimes V_d$ on $\mathbb{R}$ or $\mathbb{C}$, with fixed bases
Canonical decomposition

**Goal:** Given a tensor $T$ of order $d$, defined in space $V_1 \otimes \ldots \otimes V_d$ on $\mathbb{R}$ or $\mathbb{C}$, with fixed bases

- Determine the largest $P$ for which the decomposition

\[
T = \sum_{p=1}^{P} a(p) \otimes b(p) \otimes c(p) \otimes \ldots d(p)
\]

is *unique* (depending on $P$, $d$, and dimensions)
Canonical decomposition

**Goal:** Given a tensor $\mathbf{T}$ of order $d$, defined in space $V_1 \otimes \ldots \otimes V_d$ on $\mathbb{R}$ or $\mathbb{C}$, with fixed bases

- Determine the largest $P$ for which the decomposition

$$
\mathbf{T} = \sum_{p=1}^{P} a(p) \otimes b(p) \otimes c(p) \otimes \ldots d(p)
$$

is *unique* (depending on $P$, $d$, and dimensions)

- Or at least when there are *finitely* many
Canonical decomposition

**Goal:** Given a tensor $T$ of order $d$, defined in space $V_1 \otimes \ldots \otimes V_d$ on $\mathbb{R}$ or $\mathbb{C}$, with fixed bases

- Determine the largest $P$ for which the decomposition

$$T = \sum_{p=1}^{P} a(p) \otimes b(p) \otimes c(p) \otimes \ldots d(p)$$

is *unique* (depending on $P$, $d$, and dimensions)

- Or at least when there are *finitely* many

- Provide an algorithm to *compute* the collections of vectors $a(p), b(p), c(p), \ldots d(p)$
Applications

Application areas

1. Antenna Array Processing: Telecommunications (Cellular, Satellite, Military), Radar, Sonar, Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...

   • Techniques using High Order Statistics (e.g. cumulants)
   • Deterministic Techniques exploiting receiver diversities

2. Data & Factor Analysis (e.g. Psychometrics, Chemometrics, Food Sciences, Environment...)

3. Arithmetic Complexity Theory

4. Many other fields e.g. Medical imaging, Differential Geometry, Speech, Machine Learning, Control...
Applications

Application areas

1. Antenna Array Processing: Telecommunications (Cellular, Satellite, Military), Radar, Sonar, Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...
   - Techniques using High Order Statistics (e.g. cumulants)
Application areas

1. Antenna Array Processing: Telecommunications (Cellular, Satellite, Military), Radar, Sonar, Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...
   - Techniques using High Order Statistics (e.g. cumulants)
   - Deterministic Techniques exploiting receiver diversities
Application areas

1. Antenna Array Processing: Telecommunications (Cellular, Satellite, Military), Radar, Sonar, Biomedical (EchoGraphy, ElectroEncephaLoGraphy, ElectroCardioGraphy)...
   - Techniques using High Order Statistics (e.g. *cumulants*)
   - Deterministic Techniques exploiting receiver *diversities*

2. Data & Factor Analysis (e.g. Psychometrics, Chemometrics, Food Sciences, Environment...)

Application areas

1. Antenna Array Processing: Telecommunications (Cellular, Satellite, Military), Radar, Sonar, Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...
   - Techniques using High Order Statistics (e.g. cumulants)
   - Deterministic Techniques exploiting receiver diversities

2. Data & Factor Analysis (e.g. Psychometrics, Chemometrics, Food Sciences, Environment...)

3. Arithmetic Complexity Theory

Pierre Comon: Tensor decompositions in Eng.,
Application areas

1. Antenna Array Processing: Telecommunications (Cellular, Satellite, Military), Radar, Sonar, Biomedical (EchoGraphy, ElectroEncephaloGraphy, ElectroCardioGraphy)...
   - Techniques using High Order Statistics (e.g. *cumulants*)
   - Deterministic Techniques exploiting receiver *diversities*

2. Data & Factor Analysis (e.g. Psychometrics, Chemometrics, Food Sciences, Environment...)

3. Arithmetic Complexity Theory

4. Many other fields
   - e.g. Medical imaging, Differential Geometry, Speech, Machine Learning, Control...
Example: Antenna Array Processing (1)
Example: Antenna Array Processing (1)
Example: Antenna Array Processing (1)
Example: Antenna Array Processing (1)
Example: Antenna Array Processing (1)
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:

\[ T_{ij} = \sum_q \sum_{\ell} a_{i q \ell} \sum_k h_{q \ell k} s_{k q j} \]

- \( i \): space
- \( k \): symbol time
- \( a \): receiver geometry
- \( j \): time
- \( q \): transmitter
- \( h \): global channel impulse response
- \( \ell \): path
- \( s \): transmitted signal
Example: Antenna Array Processing (2)

Modeling the signals received on an array of antennas generally leads to a \textit{matrix decomposition}:

\[
T_{ijp} = \sum_q \sum_{\ell} a_{i q \ell} \sum_k h_{q \ell k p} s_{k q j}
\]

\(i\): space \hspace{0.5cm} \(k\): symbol time \hspace{0.5cm} \(a\): receiver geometry

\(j\): time \hspace{0.5cm} \(q\): transmitter \hspace{0.5cm} \(h\): global channel impulse response

\(\ell\): path \hspace{0.5cm} \(s\): transmitted signal

But in the presence of additional \textit{diversity}, a tensor can be constructed, thanks to new variable \(p\).
Applications

Example: Antenna Array Processing (3)

New variable \( p \) can represent:

- Oversampling (sample index),
- Spreading code (chip index),
- Frequency (multicarrier),
- Geometrical invariance (subarray index),
- Polarization
- Nonstationarity...
Example: Antenna Array Processing (3)

New variable $p$ can represent:

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

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

Data not in tensor format

If no additional diversity available, problem cannot be solved, unless other properties on sources exist:

▶ Statistical independence
▶ Sparsity
▶ Discrete...
Data not in tensor format

If no additional diversity available, problem cannot be solved, unless other properties on sources exist:

- Statistical independence $\Rightarrow$ *symmetric tensors*
- Sparsity
- Discrete...
Example: Speech

The Coktail Party problem

In free space: \[ T_{ij} = \sum_q a_{iq} \sum_k h_q(k) s_q(j - k) \]

**Remark:** Data not in tensor format ⇒ statistical tools
Applications

Example: Fluorescence Spectroscopy

An optical excitation produces several effects

At low concentrations, Beer-Lambert law can be linearized:

\[ I(\lambda_e, \lambda_f, k_n) = I_o \sum_n \gamma_n(\lambda_f) \epsilon_n(\lambda_e) c_k, n \]

But there are also non-linear effects:

• Screen effect (if some components are too concentrated)
• Reabsorption of fluorescent emissions by the medium
Example: Fluorescence Spectroscopy

An optical excitation produces several effects

- Rayleigh diffusion
- Raman diffusion
- Fluorescence

At low concentrations, Beer-Lambert law can be linearized:

$$I(\lambda_e, \lambda_f, k) = I_o \sum_n \gamma_n(\lambda_f) \epsilon_n(\lambda_e) c_k, n$$

But there are also non-linear effects:

- Screen effect (if some components are too concentrated)
- Reabsorption of fluorescent emissions by the medium
Example: Fluorescence Spectroscopy

An optical excitation produces several effects

- Rayleigh diffusion
- Raman diffusion
- Fluorescence

At low concentrations, Beer-Lambert law can be linearized

\[
I(\lambda_e, \lambda_f, k) = I_o \sum_n \gamma_n(\lambda_f) \epsilon_n(\lambda_e) c_{k,n}
\]
Applications

Example: Fluorescence Spectroscopy

An optical excitation produces several effects

► Rayleigh diffusion
► Raman diffusion
► Fluorescence

At low concentrations, Beer-Lambert law can be linearized

\[
I(\lambda_e, \lambda_f, k) = I_0 \sum_n \gamma_n(\lambda_f) \epsilon_n(\lambda_e) c_{k,n}
\]

► But there are also non linear effects:
  • Screen effect (if some components are too concentrated)
  • Reabsorption of fluorescent emissions by the medium
Example: Factor Analysis

Food Sciences:

one of the numerous application areas

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

$$T_{ijk} = \sum_p A_{ip} B_{jp} C_{kp}$$
Particular constraints

- Toeplitz matrix (convolution)
- Symmetries (partial or total, real or Hermitean)
- Positivity
- Blocks → not $r$–secant of Segre variety...
In practice, the tensor we are given is subject to errors (model, calculation) or noise, and is hence \textit{generic}.
The practical problem

- In practice, the tensor we are given is subject to errors (model, calculation) or noise, and is hence *generic*.
- If we expect a rank different from generic, we have an *approximation problem* ⇒ optimization.
Classification

Direct vs Inverse

Two formulations possible if rank bounded by dimension:

1. Direct: look for $A, B, \ldots D$:

$$
\min_{A,B,\ldots D} \left\| T_{ijk,\ldots,\ell} - \sum_{p=1}^{P} A_{ip} B_{jp} \ldots D_{\ell p} \right\|^2
$$

i.e. decompose $T$ into a sum of $P$ rank-one terms

2. Inverse: look for $A', B', \ldots D'$:

$$
\min_{A',B',\ldots D'} \left\| T_{ijk,\ldots,\ell} - \sum_{m,n,p,q} A'_{im} B'_{jn} \ldots D'_{\ell q} \right\|^2
$$

i.e. try to "diagonalize" $T$ by linear invertible change of coordinates in each space
Direct vs Inverse

Two formulations possible if rank bounded by dimension:

1. **Direct**: look for $A, B, \ldots D$:

   $$
   \min_{A,B,\ldots D} \left\| T_{ijk\ldots\ell} - \sum_{p=1}^{P} A_{ip} B_{jp} \ldots D_{\ell p} \right\|^2
   $$

   *i.e.* decompose $T$ into a sum of $P$ rank-one terms

2. **Inverse**: look for $A', B', \ldots D'$:

   $$
   \min_{A',B',\ldots D'} \left\| \sum_{mnp\ldots q \neq ppp\ldots p} T_{ijk\ldots\ell} A'_{im} B'_{jn} \ldots D'_{\ell q} \right\|^2
   $$

   *i.e.* try to “diagonalize” $T$ by linear invertible change of coordinates in each space
Orthogonal decomposition

If $A, B, C, D$ orthogonal, the two formulations are equivalent:

1. Direct:

$$\min |T_{ijk\ell} - \sum_{p=1}^A A_{ip} B_{jp} C_{kp} D_{\ell p}|^2$$

2. Inverse:

$$\min \left| \sum_{ijk\ell} A_{pi} B_{qj} C_{rk} D_{s\ell} T_{ijk\ell} - \sum_{pppp} \delta_{pqrs} \right|^2$$

or

$$\max \left| \sum_{ijk\ell} A_{pi} B_{jp} C_{pk} D_{\ell p} T_{ijk\ell} \right|^2$$

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

If $A$, $B$, $C$, $D$ orthogonal, the two formulations are equivalent:

1. *Direct:*

$$\min_{A,B,C,D,\Delta} \| T_{ijk\ell} - \sum_{p=1}^{P} A_{ip} B_{jp} C_{kp} D_{\ell p} \Delta_{ppp} \|^2$$
If $A, B, C, D$ orthogonal, the two formulations are equivalent:

1. **Direct:**

$$
\min_{A,B,C,D,\Delta} \left\| T_{ijkl} - \sum_{p=1}^{P} A_{ip} B_{jp} C_{kp} D_{\ell p} \Delta_{pppp} \right\|^2
$$

2. **Inverse:**

$$
\min \left\| \sum_{ijkl} A_{pi} B_{qj} C_{rk} D_{sl} T_{ijkl} - \Delta_{pppp} \delta_{pqrs} \right\|^2 \text{ or }
\max \left( \sum_{ijkl} A_{pi} B_{pj} C_{pk} D_{pl} T_{ijkl} \right)^2
$$
Orthogonal decomposition

If $A, B, C, D$ orthogonal, the two formulations are equivalent:

1. **Direct:**
   \[
   \min_{A,B,C,D,\Delta} \| T_{ijkl} - \sum_{p=1}^{P} A_{ip} B_{jp} C_{kp} D_{\ell p} \Delta_{pppp} \|^2
   \]

2. **Inverse:**
   \[
   \min \| \sum_{ijkl} A_{pi} B_{qj} C_{rk} D_{s\ell} T_{ijkl} - \Delta_{pppp} \delta_{pqrs} \|^2 \text{ or }
   \max \sum_{A,B,C,D} \left| \sum_{ijkl} A_{pi} B_{pj} C_{pk} D_{\ell p} T_{ijkl} \right|^2
   \]

**Proof.** The Frobenius norm is invariant under orthogonal change of coordinates. \qed
If rank exceeds dimensions (occurs generically), only the direct formulation is possible:

$$\min_{A,B,..D} \| T_{ijk...\ell} - \sum_{p=1}^{P} A_{ip} B_{jp} \ldots D_{\ell p} \|^2$$

Rank reduction is often necessary to restore uniqueness.
Orthogonal symmetric decomposition

Examples of objectives for orthogonal decomposition

\[ C_{ijkl} \overset{\text{def}}{=} \sum_{p} Q_{ip} Q_{jq} Q_{kr} Q_{ls} T_{pqrs} \]

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

Matrix slices diagonalization \( \neq \) Tensor diagonalization
Problem P2

1. At each step, a plane rotation is computed and yields the \textit{global maximum} of the objective $\Upsilon$ restricted to one variable

$$\max Q \sum_{p=1}^{n} \left| \sum_{ijk\ell} Q_{ip} Q_{jp} Q_{kp} Q_{lp} C_{ijk\ell} \right|^2$$
Orthogonal symmetric decomposition

Problem P2

1. At each step, a plane rotation is computed and yields the **global maximum** of the objective \( \Upsilon \) restricted to one variable

\[
\max_Q \sum_{p=1}^{n} \left| \sum_{ijk\ell} Q_{ip} Q_{jp} Q_{kp} Q_{\ell p} C_{ijk\ell} \right|^2
\]

2. There is no proof that the sequence of successive plane rotations yields the global maximum, in the general case (i.e. for symmetric tensors of dimension \( n \) and general form)
Problem P2

1. At each step, a plane rotation is computed and yields the \textit{global maximum} of the objective $\Upsilon$ restricted to one variable

$$\max_Q \sum_{p=1}^{n} \left| \sum_{ijk\ell} Q_{ip} Q_{jp} Q_{kp} Q_{lp} C_{ijk\ell} \right|^2$$

2. There is no proof that the sequence of successive plane rotations yields the global maximum, in the general case (i.e. for symmetric tensors of dimension $n$ and general form)

3. Yet, no counter-example has been found since 1991
Orthogonal symmetric decomposition

Computation: bibliographical comments

- Orthogonal diagonalization of symmetric tensors:
  - JAD in 2 modes: \[\text{DeLe78}\] (R), \[\text{CardS93}\] (C), with matrix exponential \[\text{TanaF07}\]
  - JAD 2 modes with positive definite matrices: \[\text{Flur86}\]
  - JAD in 3 modes: \[\text{DelaDV01}\]
  - Direct diagonal without slicing (pairs): \[\text{Como92}\] \[\text{Como94}\]

- Orthogonal diagonalization of non-symmetric tensors:
  - ALS type \[\text{Kroo83}\] \[\text{Kier92}\]
  - ALS on pairs: \[\text{MartV08}\] \[\text{SoreC08}\]
  - JAD in 2 modes (R): \[\text{Pesq01}\]
  - Matrix exponential \[\text{SoreICD08}\]
Orthogonal symmetric decomposition

Computation: bibliographical comments

- Orthogonal diagonalization of symmetric tensors:
  - JAD in 2 modes: [DeLe78] ($\mathbb{R}$), [CardS93] ($\mathbb{C}$), with matrix exponential [TanaF07]
  - ALS type [Kroo83] [Kier92]
  - ALS on pairs: [MartV08] [SoreC08]
  - JAD in 2 modes ($\mathbb{R}$): [Pesq01]
  - Matrix exponential [SoreICD08]
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]
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]
Orthogonal symmetric decomposition

Computation: bibliographical comments

- 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]
Computation: bibliographical comments

► 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:
Orthogonal symmetric decomposition

Computation: bibliographical comments

▶ Orthogonal diagonalization of symmetric tensors:
  • JAD in 2 modes: [DeLe78] (R), [CardS93] (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]
Orthogonal symmetric decomposition

Computation: bibliographical comments

- 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]
Orthogonal symmetric decomposition

Computation: bibliographical comments

- 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]
Orthogonal symmetric decomposition

Computation: bibliographical comments

- **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]
Invertible symmetric decomposition

Problem P3

Now allow general invertible transforms
  ▶ Ill-posedness of optimization over set of invertible matrices

Possible solutions:
Invertible symmetric decomposition

Problem P3

Now allow general invertible transforms
  ▶ Ill-posedness of optimization over set of invertible matrices

Possible solutions:
  ▶ Impose a constraint like \( \det A \geq \eta > 0 \)?
Invertible symmetric decomposition

**Problem P3**

Now allow general invertible transforms

- Ill-posedness of optimization over set of invertible matrices

Possible solutions:
- Impose a constraint like $\det A \geq \eta > 0$?
- Restrict $A$ to diagonally dominant, i.e. $|A_{ii}| \geq \eta + \sum_{j \neq i} |A_{ij}|$, and parameterize it as a product of matrices of the form $(I + W)$?
Invertible symmetric decomposition

Problem P3

Now allow general invertible transforms

- Ill-posedness of optimization over set of invertible matrices

Possible solutions:

- Impose a constraint like \( \det A \geq \eta > 0 \)?
- Restrict \( A \) to diagonally dominant, i.e. \( |A_{ii}| \geq \eta + \sum_{j \neq i} |A_{ij}| \), and parameterize it as a product of matrices of the form \((I + W)\)?
- Decompose as \( A = QR \) where \( R \) triangular with \( R_{ii} \geq \eta > 0 \)?
Invertible symmetric decomposition

Problem P3

Now allow general invertible transforms

- Ill-posedness of optimization over set of invertible matrices

Possible solutions:

- Impose a constraint like $\det A \geq \eta > 0$?
- Restrict $A$ to diagonally dominant, i.e. $|A_{ii}| \geq \eta + \sum_{j \neq i} |A_{ij}|$, and parameterize it as a product of matrices of the form $(I + W)$?
- Decompose as $A = QR$ where $R$ triangular with $R_{ii} \geq \eta > 0$?
- Just add a penalty term of the form $\log \det A$ to the objective?
Invertible symmetric decomposition

Computation: bibliographical comments

- Invertible diagonalization of a collection of symmetric matrices:

  - maximizes iteratively a lower bound to the decrease on a probabilistic objective [Pham01]
  - alternately minimize
    \[ \sum \| T(q) - A \Lambda(q) A^T \|_2 \]
  - with respect to \( A \) and \( \Lambda(q) \) [Yere02]. See also: [Li07] [VollO06]

  - minimizes iteratively
    \[ \sum \| T(q) - A \Lambda(q) A^T \|_2 \]
  - under the assumption that \( A \) is diagonally dominant [Zieh04].

- Invertible diagonalization of a collection of non-symmetric matrices:

  - factor \( A \) into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].

  - others: algorithms applicable to underdetermined case work here [AlbeFCC05].
Invertible symmetric decomposition

Computation: bibliographical comments

- Invertible diagonalization of a collection of symmetric matrices:
  - maximizes iteratively a lower bound to the decrease on a probabilistic objective [Pham01]

- Invertible diagonalization of a collection of non-symmetric matrices:
  - factor $A$ into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].
  - others: algorithms applicable to underdetermined case work here [AlbeFCC05].

Pierre Comon: Tensor decompositions in Eng.,
Invertible symmetric decomposition

Computation: bibliographical comments

- Invertible diagonalization of a collection of symmetric matrices:
  - maximizes iteratively a lower bound to the decrease on a probabilistic objective [Pham01]
  - alternately minimize \( \sum_q \| T(q) - A \Lambda(q) A^T \|_2 \) wrt \( A \) and \( \Lambda(q) \) [Yere02]. See also: [Li07] [VollO06]

- Invertible diagonalization of a collection of non-symmetric matrices:
  - factor \( A \) into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].
  - others: algorithms applicable to underdetermined case work here [AlbeFCC05].
Invertible symmetric decomposition

Computation: bibliographical comments

- Invertible diagonalization of a collection of symmetric matrices:
  - maximizes iteratively a lower bound to the decrease on a probabilistic objective [Pham01]
  - alternately minimize $\sum_q \| T(q) - A \Lambda(q) A^T \|^2$ wrt $A$ and $\Lambda(q)$ [Yere02]. See also: [Li07] [VollO06]
  - minimizes iteratively $\sum_q \| T(q) - A \Lambda(q) A^T \|^2$ under the assumption that $A$ is diagonally dominant [Zieh04].
Invertible symmetric decomposition

Computation: bibliographical comments

- Invertible diagonalization of a collection of symmetric matrices:
  - maximizes iteratively a lower bound to the decrease on a probabilistic objective [Pham01]
  - alternately minimize \( \sum_q \| T(q) - A \Lambda(q) A^T \|^2 \) wrt \( A \) and \( \Lambda(q) \) [Yere02]. See also: [Li07] [VollO06]
  - minimizes iteratively \( \sum_q \| T(q) - A \Lambda(q) A^T \|^2 \) under the assumption that \( A \) is diagonally dominant [Zieh04].

- Invertible diagonalization of a collection of non symmetric matrices:
  - factor \( A \) into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].
  - others: algorithms applicable to underdetermined case work here [AlbeFCC05].
Invertible symmetric decomposition

Computation: bibliographical comments

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

- Invertible diagonalization of a collection of non symmetric matrices:
  - factor $A$ into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].
Invertible symmetric decomposition

Computation: bibliographical comments

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

- Invertible diagonalization of a collection of non symmetric matrices:
  - factor $A$ into orthogonal and triangular parts, and perform a joint Schur decomposition [DelaDV04].
  - others: algorithms applicable to underdetermined case work here [AlbeFCC05]
Now rank larger than dimensions

Again we don’t know well:

- Uniqueness
- Computation

What we know rather well

- Uniqueness in symmetric case (AH theorem)
- Computation in dimension 2 (Sylvester, matrix pencils...)
Why symmetric tensors are important

- In examples of Antenna array processing (Telecommunications, Speech, Sonar...), source signals are statistically independent.
- This yields equations

\[ E\{f(s_i)g(s_j)\} = E\{f(s_i)\}E\{g(s_j)\} \]

for all \( i \neq j \) and any functions \( f \) and \( g \).
Characteristic functions

First c.f.

Real Scalar: $\Phi_x(t) \overset{\text{def}}{=} E\{e^{tx}\} = \int u e^{tx} dF_x(u)$

Real Multivariate: $\Phi_x(t) \overset{\text{def}}{=} E\{e^{t^T x}\} = \int u e^{t^T u} dF_x(u)$

Second c.f.: $\Psi(t) \overset{\text{def}}{=} \log \Phi(t)$

Properties:
• Always exists in the neighborhood of 0
• Uniquely defined as long as $\Phi(t) \neq 0$
Characteristic functions

First c.f.

- Real Scalar: \( \Phi_x(t) \overset{\text{def}}{=} \mathbb{E}\{e^{jt^x}\} = \int u e^{jt^u} dF_x(u) \)
- Real Multivariate: \( \Phi_x(t) \overset{\text{def}}{=} \mathbb{E}\{e^{jt^tx}\} = \int u e^{jt^Tu} dF_x(u) \)
Characteristic functions

First c.f.
- Real Scalar: $\Phi_x(t) \overset{\text{def}}{=} E\{e^{jtx}\} = \int_u e^{jtu} dF_x(u)$
- Real Multivariate: $\Phi_x(t) \overset{\text{def}}{=} E\{e^{jT_tx}\} = \int_u e^{jT_tu} dF_x(u)$

Second c.f.
Characteristic functions

First c.f.

- Real Scalar: \( \Phi_x(t) \overset{\text{def}}{=} E\{e^{jtx}\} = \int_u e^{jtu} dF_x(u) \)
- Real Multivariate: \( \Phi_x(t) \overset{\text{def}}{=} E\{e^{jT_t x}\} = \int_u e^{jT_t u} dF_x(u) \)

Second c.f.

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

$$\Psi_{x,y}(u, v) = \Psi_x(u) + \Psi_y(v) \quad \text{(1)}$$
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

$$\Psi_{x,y}(u, v) = \Psi_x(u) + \Psi_y(v) \quad (1)$$

**Proof.**

$$\Psi_{x,y}(u, v) = \log[E\{\exp(ux + vy)\}]$$

$$= \log[E\{\exp(ux)\} E\{\exp(vy)\}] .$$
General problem: Blind identification

Linear statistical model

\[ x = Hs + b \]  \hspace{1cm} (2)

with

\[
\begin{align*}
\mathbf{x} & : K \times 1 \text{ random} \\
\mathbf{s} & : P \times 1 \text{ random with stat. independent entries} \\
\mathbf{H} & : K \times P \text{ deterministic} \\
\mathbf{b} & : \text{ errors (may be removed for } P \text{ large enough)}
\end{align*}
\]
Problem posed in terms of Characteristic Functions

If $s_p$ independent and $x = Hs$, we have the core equation:

$$\Psi_x(u) = \sum_p \psi_{s_p} \left( \sum_q u_q H_{qp} \right)$$

(3)

Proof.
Problem posed in terms of Characteristic Functions

- If \( s_p \) independent and \( \mathbf{x} = \mathbf{H} \mathbf{s} \), we have the core equation:

\[
\Psi_x(\mathbf{u}) = \sum_p \psi_{s_p} \left( \sum_q u_q H_{qp} \right)
\]  

(3)

Proof.

- Plug \( \mathbf{x} = \mathbf{H} \mathbf{s} \), in definition of \( \Psi_x \) and get

\[
\Phi_x(\mathbf{u}) \overset{\text{def}}{=} \mathbb{E}\{\exp(\mathbf{u}^T \mathbf{H} \mathbf{s})\} = \mathbb{E}\{\exp(\sum_{p,q} u_q H_{qp} s_p)\}
\]
Problem posed in terms of Characteristic Functions

- If $s_p$ independent and $x = Hs$, we have the core equation:

$$\Psi_x(u) = \sum_p \Psi_{s_p} \left( \sum_q u_q H_{qp} \right)$$ (3)

Proof.
- Plug $x = Hs$, in definition of $\Psi_x$ and get

$$\Phi_x(u) \overset{\text{def}}{=} E\{\exp(u^T Hs)\} = E\{\exp(\sum_{p,q} u_q H_{qp} s_p)\}$$

- Since $s_p$ independent, $\Phi_x(u) = \prod_p E\{\exp(\sum_q u_q H_{qp} s_p)\}$
Problem posed in terms of Characteristic Functions

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

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

Proof.

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

$$\Phi_x(\mathbf{u}) \overset{\text{def}}{=} E\{\exp(\mathbf{u}^T \mathbf{H} \mathbf{s})\} = E\{\exp\left(\sum_{p,q} u_q H_{qp} s_p \right)\}$$

- Since $s_p$ independent, $\Phi_x(\mathbf{u}) = \prod_p E\{\exp(\sum_q u_q H_{qp} s_p)\}$

- Taking the log concludes.
Problem posed in terms of Characteristic Functions

- If $s_p$ independent and $x = Hs$, we have the core equation:

$$\Psi_x(u) = \sum_p \Psi_{s_p} \left( \sum_q u_q H_{qp} \right)$$  \hspace{1cm} (3)

**Proof.**

- Plug $x = Hs$, in definition of $\Psi_x$ and get

$$\Phi_x(u) \overset{\text{def}}{=} E\{\exp(u^T Hs)\} = E\{\exp(\sum_{p,q} u_q H_{qp} s_p)\}$$

- Since $s_p$ independent, $\Phi_x(u) = \prod_p E\{\exp(\sum_q u_q H_{qp} s_p)\}$

- Taking the log concludes.

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

Pierre Comon: Tensor decompositions in Eng., 26
Equations derived from the C.F.

Assumption: functions $\psi_p$, $1 \leq p \leq P$ admit finite derivatives up to order $r$ in a neighborhood of the origin, containing $G$. 

If $L > 1$ point in grid $G$, then yields another mode in tensor.
Equations derived from the C.F.

▶ Assumption: functions $\psi_p$, $1 \leq p \leq P$ admit finite derivatives up to order $r$ in a neighborhood of the origin, containing $G$.

▶ Then, Taking $r = 3$ as a working example:

$$\frac{\partial^3 \psi_x}{\partial u_i \partial u_j \partial u_k}(u) = \sum_{p=1}^{P} H_{ip} H_{jp} H_{kp} \psi_p^{(3)} \left( \sum_{q=1}^{K} u_q H_{qp} \right)$$
Assumption: functions $\psi_p$, $1 \leq p \leq P$ admit finite derivatives up to order $r$ in a neighborhood of the origin, containing $G$.

Then, Taking $r = 3$ as a working example:

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

If $L > 1$ point in grid $G$, then yields another mode in tensor
Putting the problem in tensor form

- Takes at \( L > 1 \) points on a grid:

\[
T_{ijkl} = \sum_p H_{ip} H_{jp} H_{kp} B_{\ell p}
\]

or

\[
T = \sum_p h(p) \otimes h(p) \otimes h(p) \otimes b(p)
\]

where tensor \( T \) is \( K \times K \times K \times L \), and partially symmetric
Putting the problem in tensor form

- Takes at \( L > 1 \) points on a grid:

\[
T_{ijkl} = \sum_p H_{ip} H_{jp} H_{kp} B_{\ell p}
\]

or
\[
T = \sum_p h(p) \otimes h(p) \otimes h(p) \otimes b(p)
\]

where tensor \( T \) is \( K \times K \times K \times L \), and partially symmetric.

- Take only equations at the origin: Cumulant tensor:

\[
T = \sum_p h(p) \otimes h(p) \otimes h(p)
\]
Put the problem in tensor form

- Takes at $L > 1$ points on a grid:

\[ T_{ijk\ell} = \sum_p H_{ip} H_{jp} H_{kp} B_{\ell p} \]

or

\[ T = \sum_p h(p) \otimes h(p) \otimes h(p) \otimes b(p) \]

where tensor $T$ is $K \times K \times K \times L$, and partially symmetric

- Take only equations at the origin: **Cumulant tensor**

\[ T = \sum_p h(p) \otimes h(p) \otimes h(p) \]

- One can get equations at arbitrary orders.
Joint use of different derivative orders

Example

- Derivatives of order 3:

\[ T^{(3)}_{ijkl} = \sum_{p} H_{ip} H_{jp} H_{kp} B_{lp} \]
Joint use of different derivative orders

Example

- Derivatives of order 3:
  \[ T^{(3)}_{ijkl} = \sum_p H_{ip} H_{jp} H_{kp} B_{lp} \]

- Derivatives of order 4:
  \[ T^{(4)}_{ijklm} = \sum_p H_{ip} H_{jp} H_{kp} H_{mp} C_{lp} \]
Joint use of different derivative orders

Example

- Derivatives of order 3:

\[ T_{ijk\ell}^{(3)} = \sum_p H_{ip} H_{jp} H_{kp} B_{\ell p} \]

- Derivatives of order 4:

\[ T_{ijkm\ell}^{(4)} = \sum_p H_{ip} H_{jp} H_{kp} H_{mp} C_{\ell p} \]

- Derivatives of orders 3 and 4:

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

with \( D_{\ell p}[m] = H_{mp} C_{\ell p} \) and \( D_{\ell p}[0] = B_{\ell p} \).
Problem P18

Let two tensors $T_1 \in A = S^3 V_1 \otimes V_2$ and $T_2 \in B = S^2 V_1 \otimes V_3$ be defined by decompositions

$$T_1 = \sum_{p=1}^{r} a(p) \otimes a(p) \otimes a(p) \otimes b(p) \quad \text{and} \quad T_2 = \sum_{p=1}^{r} a(p) \otimes a(p) \otimes c(p)$$

where $r$ is the generic rank in $B$.

We are given $\tilde{T}_1 = T_1 + E_1$ and $\tilde{T}_2 = T_2 + E_2$ where $E_i$ are small.

How can we compute the $a(p)'s$, $1 \leq p \leq r$ from $\tilde{T}_1$ and $\tilde{T}_2$?
Alexander-Hirschowitz theorem

Theorem (1995) For $d > 2$, the generic rank of a $d$th order symmetric tensor of dimension $K$ is always equal to the lower bound

$$\bar{R}_s = \left\lceil \frac{(K+d-1)}{d} \right\rceil K$$

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 \( K \)**

<table>
<thead>
<tr>
<th>( d )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>21</td>
<td>30</td>
<td>42</td>
</tr>
</tbody>
</table>

\[
\tilde{R}_s \geq \frac{1}{K} \binom{K + d - 1}{d}
\]

**Bold:** exceptions to the ceil rule: \( \tilde{R}_s = \lceil \frac{1}{K} \binom{K + d - 1}{d} \rceil \), 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
Uniqueness

- Uniqueness, or at least finite number of solutions
- *Terracini’s lemma* allows to compute the dimension of any secant variety via the tangent space
- and in particular Segre, Veronese, or any other enjoying special symmetry properties.
Values of the Generic Rank (2)

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

$$
\begin{array}{|c|ccccccc|}
\hline
 d & K & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
 3 & 2 & 5 & 7 & 10 & 14 & 19 \\
 4 & 4 & 9 & 20 & 37 & 62 & 97 \\
\hline
\end{array}
$$

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

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

**Green:** lower bound $\frac{K^d}{Kd - d + 1}$ is integer and nondefective
Numerical computation of the Generic Rank

**Mapping** (for unsymmetric tensors):

\[ \{ u(\ell), v(\ell), \ldots, w(\ell), \ 1 \leq \ell \leq r \} \xrightarrow{\varphi} \sum_{\ell=1}^{r} u(\ell) \otimes v(\ell) \otimes \ldots \otimes w(\ell) \]

\[ \{ \mathbb{C}^{n_1} \otimes \ldots \otimes \mathbb{C}^{n_d} \}^r \xrightarrow{\varphi} \mathcal{A} \]

- The **smallest** \( r \) for which \( \text{rank}(\text{Jacobian}(\varphi)) = \prod_i n_i \) is the generic rank, \( \bar{R} \).
- Example of use of **Terracini’s lemma**
First example of computation of Generic Rank

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

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

\[
J = \begin{bmatrix}
I_{n_1} \otimes b^T(1) \otimes c^T(1) \\
: & \otimes & : & \otimes & : \\
I_{n_1} \otimes b^T(r) \otimes c^T(r) \\
a(1)^T \otimes I_{n_2} \otimes c^T(1) \\
: & \otimes & : & \otimes & : \\
a(r)^T \otimes I_{n_2} \otimes c^T(r) \\
a(1)^T \otimes b(1)^T \otimes I_{n_3} \\
: & \otimes & : & \otimes & : \\
a(r)^T \otimes b(r)^T \otimes I_{n_3}
\end{bmatrix}
\]

and \( \text{rank}\{J\} = \text{dim}(\text{Im}(\varphi)) \) 

\( \bar{R} = \text{Min}\{r : \text{Im}\{\varphi\} = \mathcal{A}\} \)
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
- Uniqueness?
Second Example: third order real tensors with symmetric slices

Typical ranks for $N_1 \times N_2 \times N_2$ arrays, with $N_2 \times N_2$ real symmetric slices.

<table>
<thead>
<tr>
<th>$N_1$</th>
<th>$N_2$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>2,3</td>
<td>3,4</td>
<td>4,5</td>
<td>5,6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td>4,5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>3</td>
<td>5,6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>3</td>
<td>6</td>
<td>9,10</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

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

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

- A rank $r$ is \textit{typical} if $\mathcal{Z}_r$ is Zariski-dense
- In the complex field, there is only one typical rank, called the \textit{generic rank}.
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
Problem P13

- Uniqueness 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-posedness?
- Computation?
Problem P7

**Computation of a decomposition** via flattening matrices when unique (e.g. general tensors with subgeneric ranks [ChiaC06])

Let $V_i$ be spaces of dimension $n_i$, and for some chosen $\ell$, let

$$p \overset{\text{def}}{=} \sum_{i=1}^{\ell} n_i, \quad q \overset{\text{def}}{=} \sum_{i=\ell+1}^{d} n_i, \quad B \overset{\text{def}}{=} V_1 \otimes \ldots \otimes V_\ell \quad \text{and} \quad C \overset{\text{def}}{=} V_{\ell + 1} \otimes \ldots \otimes V_d$$

Let $T \in B \otimes C$ be a tensor of rank at most $\min(p, q)$.

- Associate $T$ with a linear operator $\varphi$ from $B^* \to C$, defined by a matrix $M$ of size $p \times q$.
- Compute a basis $\{a(k)\}_{1 \leq k \leq r}$ of $\text{Im}\{\varphi\}$
- Find all linear combinations $b = \sum_k \lambda_k a(k)$ such that $b$ represents a tensor of rank 1 in $C$.

**How** can we solve this quadratic system of equations in $\lambda_j$?
Problem P4

Lower rank approximation

- One can reduce the rank by truncating the basis \( \{ a(k) \} \) of \( \text{Im}\{ \varphi \} \)
- This may restore uniqueness
- Sometimes hint on expected rank from practical problem
- But ill posed problem because of lack of closeness
FOOBI algorithms

Given a $K^2 \times P$ matrix $H \odot^2$, find a real orthogonal matrix $Q$ such that the $P$ columns of $H \odot^2 Q$ are of the form $h[p] \otimes h[p]^*$
FOOBi algorithms

Given a $K^2 \times P$ matrix $H \otimes^2$, find a real orthogonal matrix $Q$ such that the $P$ columns of $H \otimes^2 Q$ are of the form $h[p] \otimes h[p]^*$

- **FOOBI**: use the $K^4$ determinantal equations characterizing rank-1 matrices $h[p] h[p]$ of the form:

  $\phi(X, Y)_{ijk\ell} = x_{ij} y_{\ell k} - x_{ik} y_{\ell j} + y_{ij} x_{\ell k} - y_{ik} x_{\ell j}$
FOOBI algorithms

Given a $K^2 \times P$ matrix $H \otimes^2$, find a real orthogonal matrix $Q$ such that the $P$ columns of $H \otimes^2 Q$ are of the form $h[p] \otimes h[p]^*$

- **FOOBI**: use the $K^4$ determinantal equations characterizing rank-1 matrices $h[p] h[p]^H$ of the form:
  \[
  \phi(X, Y)_{ijk\ell} = x_{ij} y_{\ell k} - x_{ik} y_{\ell j} + y_{ij} x_{\ell k} - y_{ik} x_{\ell j}
  \]

- **FOOBI2**: use the $K^2$ equations of the form:
  \[
  \Phi(X, Y) = XY + YX - \text{trace}\{X\}Y - \text{trace}\{Y\}X
  \]
  where matrices $X$ and $Y$ are $K \times K$ Hermitean.
Decomposition of symmetric tensors with rank larger than dimension:

- **FOOBI**: Quite interesting recent algorithms based on matrix slices and rank-1 detecting criteria [DelaCC07]

  Lieven DeLathauwer

Decomposition of general tensors with rank larger than dimensions:

- Basically all iterative
Computation via iterative algorithms

Many practitioners execute more or less brute force minimizations of \( \| \mathbf{T} - \sum_{p=1}^{r} u_p \otimes v_p \otimes \ldots \otimes w_p \| ^2 \)

- 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
  - *initial point* could be provided by flattening matrices
Conclusion

Lack of special purpose algorithmic tools. Suboptimal because:

▶ Some “open” problems listed
▶ either minimize 2 successive criteria instead of a single one
▶ or treat a tensor as a collection of matrices
▶ or ignore some symmetries
▶ or are iterative without global convergence proof
▶ or need rank reduction
▶ or all together

Ignorance is the necessary condition for human being happiness.

Anatole France (1844-1924)
Conclusion

Lack of special purpose algorithmic tools. Suboptimal because:

- Some “open” problems listed
- either minimize 2 successive criteria instead of a single one
- or treat a tensor as a collection of matrices
- or ignore some symmetries
- or are iterative without global convergence proof
- or need rank reduction
- or all together

One starts to know the scale of our ignorance...

Ignorance is the necessary condition for human being happiness.
Anatole France (1844-1924)