When tensor decomposition meets compressed sensing

Pierre Comon

I3S, CNRS, University of Nice - Sophia Antipolis, France
Collaborator: Lek-Heng Lim

Sept. 27-30, 2010
Blind Source Separation & Sparse Representation

\[ x = Hs \]

\( H \) is \( K \times P \), \textit{underdetermined}: \( K < P \)

- \textit{Sparse representation}: Columns \( h_n \in \mathbb{D} \), known dictionary
- \textit{BSS}: \( H \) unknown
  - \( s \) is sparse
  - \( s \) is not sparse
Blind Source Separation & Sparse Representation

\[ x = Hs \]

- \( H \) is \( K \times P \), **underdetermined**: \( K < P \)

- **Sparse representation**: Columns \( h_n \in \mathbb{D} \), known dictionary
- **BSS**: \( H \) unknown
  - \( s \) sparse
  - \( s \) not sparse
Blind Source Separation & Sparse Representation

\[ x = Hs \]

\( H \) is \( K \times P \), \textit{underdetermined}: \( K < P \)

- **Sparse representation**: Columns \( h_n \in \mathbb{D} \), known dictionary
- **BSS**: \( H \) unknown
  - \( s \) sparse
  - \( s \) not sparse
Blind Source Separation & Sparse Representation

\[ \mathbf{x} = \mathbf{H} \mathbf{s} \]

\( \mathbf{H} \) is \( K \times P \), \textit{underdetermined}: \( K < P \)

- \textit{Sparse representation}: Columns \( \mathbf{h}_n \in \mathbb{D} \), known dictionary
- \textit{BSS}: \( \mathbf{H} \) unknown
  - \( \mathbf{s} \) sparse
  - \( \mathbf{s} \) \textit{not} sparse
Blind Source Separation & Sparse Representation

\[ x = Hs \]

\( H \) is \( K \times P \), \textit{underdetermined}: \( K < P \)

- \textit{Sparse representation}: Columns \( h_n \in \mathbb{D} \), known dictionary
- \textit{BSS}: \( H \) unknown
  - \( s \) sparse
  - \( s \) \textit{not sparse}
Blind identification of linear mixtures

- **Linear mixtures:**

  \[ x = Hs \]

  If \( s_\ell \) statistically independent, we have the core equation:

  \[ \Psi(u) = \sum_{\ell=1}^{P} \varphi_\ell(u^TH) \]

- Take 3rd derivatives at point \( u \):

  \[ T_{ijk}(u) = \sum_{\ell=1}^{P} H_{i\ell} H_{j\ell} H_{k\ell} C_{\ell\ell\ell}(u) \] (1)

  At \( u = 0 \) \( \Rightarrow \) symmetric decomposition of \( T_{ijk} \) \[ \text{[HOS]} \]

  At \( u \neq 0 \) \( \Rightarrow \) \[ \text{[Taleb, Comon-Rajih, Yeredor]} \]
Blind identification of linear mixtures

- **Linear mixtures:**
  
  \[ x = Hs \]
  
  If \( s_\ell \) statistically independent, we have the core equation:
  
  \[
  \Psi(u) = \sum_{\ell=1}^{P} \varphi_{\ell}(u^T H)
  \]

- Take 3rd derivatives at point \( u \):
  
  \[
  T_{ijk}(u) = \sum_{\ell=1}^{P} H_{i\ell} H_{j\ell} H_{k\ell} C_{\ell\ell\ell}(u)
  \] (1)

At \( u = 0 \) \( \text{symmetric decomposition of } T_{ijk} \) [HOS]
At \( u \neq 0 \) \( \text{[Taleb, Comon-Rajih, Yeredor]} \)
Blind identification of linear mixtures

- **Linear mixtures:**

\[ \mathbf{x} = \mathbf{Hs} \]

If \( s_\ell \) statistically independent, we have the *core equation*:

\[ \Psi(\mathbf{u}) = \sum_{\ell=1}^{P} \varphi_\ell(\mathbf{u}^T \mathbf{H}) \]

- Take 3rd derivatives at point \( \mathbf{u} \):

\[ T_{ijk}(\mathbf{u}) = \sum_{\ell=1}^{P} H_{i\ell} H_{j\ell} H_{k\ell} \; C_{\ell\ell\ell}(\mathbf{u}) \]  \hspace{1cm} (1)

At \( \mathbf{u} = 0 \) ➔ symmetric decomposition of \( T_{ijk} \)  \hspace{1cm} [HOS]

At \( \mathbf{u} \neq 0 \) ➔  \hspace{1cm} [Taleb, Comon-Rajih, Yeredor]
Blind identification of linear mixtures

- **Linear mixtures:**

  \[ x = Hs \]

  If \( s_\ell \) statistically independent, we have the core equation:

  \[ \Psi(u) = \sum_{\ell=1}^{P} \varphi_\ell(u^T H) \]

- **Take 3rd derivatives at point \( u \):**

  \[ T_{ijk}(u) = \sum_{\ell=1}^{P} H_{i\ell} H_{j\ell} H_{k\ell} C_{\ell\ell\ell}(u) \]  
  \[ 1 \]

  At \( u = 0 \) \( \triangleright\) symmetric decomposition of \( T_{ijk} \) [HOS]

  At \( u \neq 0 \) \( \triangleright \) [Taleb, Comon-Rajih, Yeredor]
Statistical approach

Blind identification of linear mixtures

- **Linear mixtures:**

\[ x = Hs \]

If \( s_\ell \) **statistically independent**, we have the **core equation**:

\[
\Psi(u) = \sum_{\ell=1}^{P} \varphi_\ell(u^T H)
\]

- Take 3rd derivatives at point \( u \):

\[
T_{ijk}(u) = \sum_{\ell=1}^{P} H_{i\ell} H_{j\ell} H_{k\ell} C_{\ell\ell\ell}(u) \quad (1)
\]

At \( u = 0 \) ➤ symmetric decomposition of \( T_{ijk} \)  
[HOS]

At \( u \neq 0 \) ➤ [Taleb, Comon-Rajih, Yeredor]
Tensors: General items
Arrays and Multi-linearity

- A tensor of order $d$ is a multi-linear map:

$$S_1^* \otimes \ldots \otimes S_m^* \rightarrow S_{m+1} \otimes \ldots \otimes S_d$$

- Once bases of spaces $S_{\ell}$ are fixed, they can be represented by $d$-way arrays of coordinates

- bilinear form, or linear operator: represented by a matrix
- trilinear form, or bilinear operator: by a 3rd order tensor.
Arrays and Multi-linearity

- A tensor of \textit{order d} is a multi-linear map:

\[ S_1^* \otimes \ldots \otimes S_m^* \rightarrow S_{m+1} \otimes \ldots \otimes S_d \]

- Once bases of spaces \( S_\ell \) are fixed, they can be represented by \textit{d-way arrays} of coordinates

\begin{itemize}
  \item bilinear form, or linear operator: represented by a \textit{matrix}
  \item trilinear form, or bilinear operator: by a \textit{3rd order tensor.}
\end{itemize}
Multi-linearity

Compact notation

Linear change in contravariant spaces:

\[ T'_{ijk} = \sum_{npq} A_{in} B_{jp} C_{kq} T_{npq} \]

Denoted compactly

\[ T' = (A, B, C) \cdot T \]

Example: covariance matrix

\[ z = Ax \implies R_z = (A, A) \cdot R_x = A R_x A^T \]
Decomposable tensor

- A $d$th order "decomposable" tensor is the tensor product of $d$ vectors:
  \[ T = u \otimes v \otimes \ldots \otimes w \]
  and has coordinates \[ T_{ij\ldots k} = u_i v_j \ldots w_k. \]
- May be seen as a discretization of multivariate functions whose variables separate:
  \[ t(x, y, \ldots, z) = u(x) v(y) \ldots w(z) \]
- Nothing else but rank-1 tensors, with forthcoming definition
Decomposable tensor

- A $d$th order "decomposable" tensor is the tensor product of $d$ vectors:
  $$\mathcal{T} = u \otimes v \otimes \ldots \otimes w$$
  and has coordinates $T_{ij\ldots k} = u_i v_j \ldots w_k$.
- may be seen as a discretization of multivariate functions whose variables separate:
  $$t(x, y, \ldots, z) = u(x) v(y) \ldots w(z)$$
- Nothing else but rank-1 tensors, with forthcoming definition
Decomposable tensor

- A $d$th order "decomposable" tensor is the tensor product of $d$ vectors:

$$\mathcal{T} = u \otimes v \otimes \ldots \otimes w$$

and has coordinates $T_{ij\ldots k} = u_i v_j \ldots w_k$.

- may be seen as a discretization of multivariate functions whose variables separate:

$$t(x, y, \ldots, z) = u(x) v(y) \ldots w(z)$$

- Nothing else but rank-1 tensors, with forthcoming definition
Example

- Take \( v = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \)

- Then \( v \otimes 3 \overset{\text{def}}{=} v \otimes v \otimes v = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \)

- This is a “decomposable” symmetric tensor \( \rightarrow \text{rank-1} \)

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

■ Take \( \mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \)

■ Then \( \mathbf{v} \otimes 3 \equiv \mathbf{v} \otimes \mathbf{v} \otimes \mathbf{v} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \)

■ This is a “decomposable” symmetric tensor \( \rightarrow \) rank-1

blue bullets = 1, red bullets = \(-1\).
Example

- Take \( v = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \)
- Then \( v \otimes 3 \overset{\text{def}}{=} v \otimes v \otimes v = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix} \)
- This is a “decomposable” symmetric tensor → rank-1

blue bullets = 1, red bullets = \(-1\).
Matrix SVD, $\mathbf{M} = (\mathbf{U}, \mathbf{V}) \cdot \Sigma$, may be extended in at least two ways to tensors

- **Keep orthogonality:** Orthogonal Tucker, HOSVD
  
  $$\mathcal{T} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \mathcal{C}$$

  $\mathcal{C}$ is $R_1 \times R_2 \times R_3$: *multilinear rank* $= (R_1, R_2, R_3)$

- **Keep diagonality:** Canonical Polyadic decomposition (CP)
  
  $$\mathcal{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{L}$$

  $\mathcal{L}$ is $R \times R \times R$ diagonal, $\lambda_i \neq 0$: *rank* $= R$. 

---

From SVD to tensor decompositions
From SVD to tensor decompositions

Matrix SVD, \( \mathbf{M} = (\mathbf{U}, \mathbf{V}) \cdot \mathbf{\Sigma} \), may be extended in at least two ways to tensors

- **Keep orthogonality:** Orthogonal Tucker, HOSVD
  \[ T = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \mathbf{C} \]
  \( \mathbf{C} \) is \( R_1 \times R_2 \times R_3 \): multilinear rank = \( (R_1, R_2, R_3) \)

- **Keep diagonality:** Canonical Polyadic decomposition (CP)
  \[ T = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathbf{L} \]
  \( \mathbf{L} \) is \( R \times R \times R \) diagonal, \( \lambda_i \neq 0 \): rank = \( R \).
From SVD to tensor decompositions

Matrix SVD, \( \mathbf{M} = (\mathbf{U}, \mathbf{V}) \cdot \mathbf{\Sigma} \), may be extended in at least two ways to tensors

- **Keep orthogonality:** Orthogonal Tucker, HOSVD
  \[
  \mathbf{T} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \mathbf{C}
  \]
  \( \mathbf{C} \) is \( R_1 \times R_2 \times R_3 \): multilinear rank = \((R_1, R_2, R_3)\)

- **Keep diagonality:** Canonical Polyadic decomposition (CP)
  \[
  \mathbf{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathbf{L}
  \]
  \( \mathbf{L} \) is \( R \times R \times R \) diagonal, \( \lambda_i \neq 0 \): rank = \( R \).
From SVD to tensor decompositions

Matrix SVD, $M = (U, V) \cdot \Sigma$, may be extended in at least two ways to tensors

- **Keep orthogonality:** Orthogonal Tucker, HOSVD
  \[ T = (U, V, W) \cdot C \]
  $C$ is $R_1 \times R_2 \times R_3$: **multilinear rank** $= (R_1, R_2, R_3)$

- **Keep diagonality:** Canonical Polyadic decomposition (CP)
  \[ T = (A, B, C) \cdot L \]
  $L$ is $R \times R \times R$ diagonal, $\lambda_i \neq 0$: **rank** $= R$. 

Definitions
From SVD to tensor decompositions

Matrix SVD, \( \mathbf{M} = (\mathbf{U}, \mathbf{V}) \cdot \Sigma \), may be extended in at least two ways to tensors

- **Keep orthogonality:** Orthogonal Tucker, HOSVD

  \[ \mathcal{T} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \mathcal{C} \]

  \( \mathcal{C} \) is \( R_1 \times R_2 \times R_3 \): *multilinear rank* = \( (R_1, R_2, R_3) \)

- **Keep diagonality:** Canonical Polyadic decomposition (CP)

  \[ \mathcal{T} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{L} \]

  \( \mathcal{L} \) is \( R \times R \times R \) diagonal, \( \lambda_i \neq 0 \): *rank* = \( R \).
Exact Canonical Polyadic (CP) decomposition
**Canonical Polyadic (CP) decomposition**

- Any $I \times J \times \cdots \times K$ tensor $\mathcal{T}$ can be decomposed as

\[
\mathcal{T} = \sum_q \lambda_q \mathbf{u}^{(q)} \otimes \mathbf{v}^{(q)} \otimes \cdots \otimes \mathbf{w}^{(q)}
\]

- “Polyadic form” [Hitchcock’27]

- The tensor rank of $\mathcal{T}$ is the minimal number $R(\mathcal{T})$ of “decomposable” terms such that equality holds.

- May impose unit norm vectors $\mathbf{u}^{(q)}, \mathbf{v}^{(q)}, \ldots \mathbf{w}^{(q)}$
Canonical Polyadic (CP) decomposition

- Any $I \times J \times \cdots \times K$ tensor $\mathcal{T}$ can be decomposed as

\[
\mathcal{T} = \sum_{q}^{R(T)} \lambda_q \mathbf{u}^{(q)} \otimes \mathbf{v}^{(q)} \otimes \ldots \otimes \mathbf{w}^{(q)}
\]

- “Polyadic form” [Hitchcock’27]
- The tensor rank of $\mathcal{T}$ is the minimal number $R(\mathcal{T})$ of “decomposable” terms such that equality holds.
- May impose unit norm vectors $\mathbf{u}^{(q)}, \mathbf{v}^{(q)}, \ldots \mathbf{w}^{(q)}$
Any $I \times J \times \cdots \times K$ tensor $T$ can be decomposed as

$$T = \sum_{q=1}^{R(T)} \lambda_q u^{(q)} \otimes v^{(q)} \otimes \cdots \otimes w^{(q)}$$

“Polyadic form” [Hitchcock’27]

The tensor rank of $T$ is the minimal number $R(T)$ of “decomposable” terms such that equality holds.

May impose unit norm vectors $u^{(q)}$, $v^{(q)}$, $\ldots$, $w^{(q)}$
Hitchcock

Frank Lauren Hitchcock
(1875-1957)

[Courtesy of L-H.Lim]
Hitchcock

Frank Lauren Hitchcock
(1875-1957)

Claude Elwood Shannon
(1916-2001)

[Courtesy of L-H.Lim]
Towards a unique terminology

- Minimal Polyadic Form [Hitchcock’27]
- Canonical decomposition [Weinstein’84, Carroll’70, Chiantini-Ciliberto’06, Comon’00, Khoromskij, Tyrtyshnikov]
- Parafac [Harshman’70, Sidiropoulos’00]
- Optimal computation [Strassen’83]
- Minimum-length additive decomposition (AD) [Iarrobino’96]

Suggestion:
- Canonical Polyadic decomposition (CP) [Comon’08, Grasedyk, Espig…]
- CP does also already stand for Candecomp/Parafac [Bro’97, Kiers’98, tenBerge’04…]
Psychometrics

Richard A. Harshman
(1970)

J. Douglas Carroll
(1970)
### Uniqueness: Kruskal 1/2

The **Kruskal rank** of a matrix $\mathbf{A}$ is the maximum number $k_A$, such that any subset of $k_A$ columns are linearly independent.
Uniqueness: Kruskal 2/2

Sufficient condition for uniqueness of CP

[Kruskal’77, Sidiropoulos-Bro’00, Landsberg’09]:

Essential uniqueness is ensured if tensor rank $R$ is below the so-called Kruskal’s bound:

$$2R + 2 \leq k_A + k_B + k_C$$  \hspace{1cm} (3)

or generically, for a tensor of order $d$ and dimensions $N_\ell$:

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

Bound much smaller than expected rank:

$\exists$ a much better bound, in almost sure sense
Uniqueness: Kruskal 2/2

**Sufficient condition** for uniqueness of CP

[Kruskal’77, Sidiropoulos-Bro’00, Landsberg’09]:

Essential uniqueness is ensured if tensor rank $R$ is below the so-called *Kruskal’s bound*:

$$2R + 2 \leq k_A + k_B + k_C$$ (3)

or *generically*, for a tensor of order $d$ and dimensions $N_\ell$:

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

*Bound much smaller than expected rank:*

$\exists$ *a much better bound*, in almost sure sense
Rank-3 example 1/2

Example

\[ T = 2 = + + 2 \]
Rank-3 example 1/2

Example

\[ T = 2 \quad = \quad + \quad + 2 \]

blue bullets = 1, red bullets = −1.
Rank-3 example 2/2

Conclusion: the $2 \times 2 \times 2$ tensor

$$T = 2 \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

admits the CP

$$T = \left( \begin{array}{c} 1 \\ 1 \end{array} \right) \otimes^3 + \left( \begin{array}{c} -1 \\ 1 \end{array} \right) \otimes^3 + 2 \left( \begin{array}{c} 0 \\ -1 \end{array} \right) \otimes^3$$

and has rank 3, hence *larger* than dimension
Approximate Canonical Polyadic (CP) decomposition
Why need for approximation?

- Additive noise in measurements
- Noise has a continuous probability distribution
- Then tensor rank is *generic*
- Hence there are often *infinitely many* CP decompositions

Approximations aim at getting rid of noise, and at restoring uniqueness:

$$\text{Arg} \inf_{a(p), b(p), c(p)} \| T - \sum_{p=1}^{r} a(p) \otimes b(p) \ldots \otimes c(p) \|^2$$  \hspace{1cm} (4)

But infimum may be reached for tensors of rank $> r$!
Why need for approximation?

- Additive noise in measurements
- Noise has a continuous probability distribution
- Then tensor rank is *generic*
- Hence there are often *infinitely many* CP decompositions

Approximations aim at getting rid of noise, *and* at restoring uniqueness:

\[
\text{Arg} \inf_{a(p), b(p), c(p)} \|T - \sum_{p=1}^{r} a(p) \otimes b(p) \ldots \otimes c(p)\|^2
\]  

But infimum may be reached for tensors of rank \( > r \)!
$\mathcal{T}$ has \textit{border rank} $R$ iff it is the limit of tensors of rank $R$, and not the limit of tensors of lower rank.

[Bini'79, Schönhage'81, Strassen'83, Likteig'85]
Example

Let \( u \) and \( v \) be non collinear vectors. Define \( T_0 \) [Comon et al.'08]:

\[
T_0 = u \otimes u \otimes u \otimes v + u \otimes u \otimes v \otimes u + u \otimes v \otimes u \otimes u + v \otimes u \otimes u \otimes u
\]

And define sequence \( T_\varepsilon = \frac{1}{\varepsilon} \left[ (u + \varepsilon v)^\otimes 4 - u^\otimes 4 \right] \).

Then \( T_\varepsilon \rightarrow T_0 \) as \( \varepsilon \rightarrow 0 \)

\( \Rightarrow \) Hence \( \text{rank}\{T_0\} = 4 \), but \( \text{rank}\{T_0\} = 2 \)
Let \( u \) and \( v \) be non collinear vectors. Define \( \mathcal{T}_0 \) [Comon et al.’08]:

\[
\mathcal{T}_0 = u \otimes u \otimes u \otimes v + u \otimes u \otimes v \otimes u + u \otimes v \otimes u \otimes u + v \otimes u \otimes u \otimes u
\]

And define sequence \( \mathcal{T}_\varepsilon = \frac{1}{\varepsilon} \left[ (u + \varepsilon v)^\otimes 4 - u^\otimes 4 \right] \).

Then \( \mathcal{T}_\varepsilon \rightarrow \mathcal{T}_0 \) as \( \varepsilon \rightarrow 0 \)

\( \blacksquare \) Hence \( \text{rank}\{\mathcal{T}_0\} = 4 \), but \( \text{rank}\{\mathcal{T}_0\} = 2 \)
Example

Let $\mathbf{u}$ and $\mathbf{v}$ be non collinear vectors. Define $\mathcal{T}_0$ [Comon et al.'08]:

$$\mathcal{T}_0 = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}$$

And define sequence $\mathcal{T}_\varepsilon = \frac{1}{\varepsilon} \left[ (\mathbf{u} + \varepsilon \mathbf{v}) \otimes^4 - \mathbf{u} \otimes^4 \right]$. Then $\mathcal{T}_\varepsilon \rightarrow \mathcal{T}_0$ as $\varepsilon \rightarrow 0$

Hence $\text{rank}\{\mathcal{T}_0\} = 4$, but $\text{rank}\{\mathcal{T}_0\} = 2$
Example

Let $\mathbf{u}$ and $\mathbf{v}$ be non collinear vectors. Define $\mathcal{I}_0$ [Comon et al.’08]:

$$\mathcal{I}_0 = \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u} + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}$$

And define sequence $\mathcal{I}_\varepsilon = \frac{1}{\varepsilon} \left[ (\mathbf{u} + \varepsilon \mathbf{v}) \otimes^4 - \mathbf{u} \otimes^4 \right]$.

Then $\mathcal{I}_\varepsilon \rightarrow \mathcal{I}_0$ as $\varepsilon \rightarrow 0$

$\Rightarrow$ Hence rank$\{\mathcal{I}_0\} = 4$, but rank$\{\mathcal{I}_0\} = 2$
Ill-posedness

Tensors for which $\text{rank}\{\mathcal{T}\} < \text{rank}\{\mathcal{T}\}$ are such that the approximating sequence contains several decomposable tensors which

- tend to infinity \textit{and}
- cancel each other, viz, some columns become close to collinear

Ideas towards a well-posed problem:
- Prevent collinearity \textit{or} bound columns.
Ill-posedness

Tensors for which \( \text{rank}\{\mathcal{T}\} < \text{rank}\{\mathcal{T}\} \) are such that the approximating sequence contains several decomposable tensors which

- tend to infinity \textit{and}
- cancel each other, viz, some columns become close to collinear

Ideas towards a well-posed problem:

- Prevent collinearity \textit{or} bound columns.
Ill-posedness

Tensors for which $\text{rank}\{\mathcal{T}\} < \text{rank}\{\mathcal{T}\}$ are such that the approximating sequence contains several decomposable tensors which

- tend to infinity \textit{and}
- cancel each other, viz, some columns become close to collinear

Ideas towards a well-posed problem:

- Prevent collinearity \textit{or} bound columns.
Ill-posedness

Tensors for which $\text{rank}\{\mathcal{T}\} < \text{rank}\{\mathcal{T}\}$ are such that the approximating sequence contains several decomposable tensors which

- tend to infinity \textit{and}

- cancel each other, viz, some columns become close to collinear

\textbf{Ideas towards a well-posed problem:}

- Prevent collinearity \textit{or} bound columns.
Remedies

1. Impose orthogonality of columns within factor matrices \[\text{[Comon’92]}\]

2. Impose orthogonality between decomposable tensors \[\text{[Kolda’01]}\]

3. Prevent tendency to infinity by norm constraint on factor matrices \[\text{[Paatero’00]}\]

4. Nonnegative tensors: impose decomposable tensors to be nonnegative \[\text{[Lim-Comon’09]} \rightarrow \text{“nonnegative rank”}\]

5. Impose minimal angle between columns of factor matrices \[\text{[Lim-Comon’10]}\]
Remedies

1. Impose *orthogonality* of columns within factor matrices [Comon’92]

2. Impose *orthogonality* between decomposable tensors [Kolda’01]

3. Prevent tendency to infinity by *norm constraint* on factor matrices [Paatero’00]

4. Nonnegative tensors: impose decomposable tensors to be *nonnegative* [Lim-Comon’09] → “nonnegative rank”

5. Impose *minimal angle* between columns of factor matrices [Lim-Comon’10]
Nonnegativity constraint: example
Fluorescence Spectroscopy 1/3

An optical excitation produces several effects

- Rayleigh diffusion
- Raman diffusion
- Fluorescence

At low concentrations, Beer-Lambert law can be linearized [Luciani'09]

\[ I(\lambda_f, \lambda_e, k) = I_o \sum \gamma_k(\lambda_f) \epsilon_k(\lambda_e) c_{k,e} \]

Hence 3rd array decomposition with real nonnegative factors [Bro'97].
Fluorescence Spectroscopy  1/3

An optical excitation produces several effects

- Rayleigh diffusion
- Raman diffusion
- Fluorescence

At low concentrations, Beer-Lambert law can be linearized

\[ I(\lambda_f, \lambda_e, k) = I_o \sum_{\ell} \gamma_{\ell}(\lambda_f) \epsilon_{\ell}(\lambda_e) c_{k,\ell} \]

Hence 3rd array decomposition with real nonnegative factors \([Bro'97]\).
Fluorescence Spectroscopy  1/3

An optical excitation produces several effects

- Rayleigh diffusion
- Raman diffusion
- Fluorescence
  At low concentrations, Beer-Lambert law can be linearized
  \[ I(\lambda_f, \lambda_e, k) = I_o \sum_\ell \gamma_\ell(\lambda_f) \epsilon_\ell(\lambda_e) c_{k,\ell} \]

\[ \Rightarrow \text{Hence 3rd array decomposition with real nonnegative factors [Bro’97].} \]
Mixture of 4 solutes (one concentration shown)
Fluorescence Spectroscopy  3/3

Obtained results with $R = 4$
Approximate CP decomposition: Coherence constraint
Coherence

Definition: [Donoho’03, Gribonval’03, Candès’07] let $A$ a matrix with unit-norm columns, $a_p$.

$$\mu_A = \max_{p \neq q} |\langle a_p, a_q \rangle|$$

this “coherence of $A$” is used in Sparse Representation theory
Existence

Best rank-$R$ approximate under angular constraint

- **Proposition:** \([\text{Lim-Comon}'2010]\)
  Let $\mathcal{L}$ diagonal, and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ have $R$ unit norm columns. If $R < \left[\mu_A \mu_B \mu_C\right]^{-1}$, then:

$$\inf_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathcal{T} - (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \mathcal{L}\|$$

*is attained.*
**Lemma:** (spark) \[ Gribonval’03 \]

\[ k_A \geq \frac{1}{\mu_A} \]
Uniqueness

**Proposition:** *[Lim-Comon’10]*

If $\mathcal{T} = (A, B, C) \cdot \mathcal{L}$, with $\lambda_p \neq 0$ for $1 \leq p \leq R$, $A$, $B$, $C$ matrices with unit norm columns, and:

$$2R + 2 \leq \frac{1}{\mu_A} + \frac{1}{\mu_B} + \frac{1}{\mu_C}$$  \hspace{1cm} (6)

then $\mathcal{T}$ has a unique CP decomposition of rank $R$, up to unit modulus scalar factors $(\rho_A, \rho_B, \rho_C)$, $\rho_A \rho_B \rho_C = 1$.

Hence it suffices that one $\mu$ is small, not every.
Uniqueness

Proposition: [Lim-Comon’10]
If $\mathcal{T} = (A, B, C) \cdot \mathcal{L}$, with $\lambda_p \neq 0$ for $1 \leq p \leq R$, $A, B, C$ matrices with unit norm columns, and:

$$2R + 2 \leq \frac{1}{\mu_A} + \frac{1}{\mu_B} + \frac{1}{\mu_C} \quad (6)$$

then $\mathcal{T}$ has a unique CP decomposition of rank $R$, up to unit modulus scalar factors $(\rho_A, \rho_B, \rho_C)$, $\rho_A \rho_B \rho_C = 1$.

Hence it suffices that one $\mu$ is small, not every
Angular constraint: example
Antenna Array Processing
Antenna Array Processing
Antenna Array Processing

Transmitter

Receiver
Antenna Array Processing
Antenna Array Processing
Antenna Array Processing

[Diagram of an antenna array processing system]
Narrow band model in the far field

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

\[ T_{ij} = \sum_q a_{iq} s_{jq} \]

\( i: \) space \hspace{1cm} \( q: \) path, source \hspace{1cm} \( A: \) antenna gains

\( j: \) time \hspace{1cm} \( S: \) transmitted signals

But in the presence of additional diversity, a tensor can be constructed, thanks to new index \( p \)
Narrow band model in the far field

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

\[ T_{ijp} = \sum_q a_{iq} s_{jq} h_{pq} \]

- \( i \): space
- \( q \): path, source
- \( A \): antenna gains
- \( j \): time
- \( S \): transmitted signals

But in the presence of additional *diversity*, a tensor can be constructed, thanks to new index \( p \)
Possible diversities in Signal Processing

- space
- time
- space translation (array geometry)
- time translation (chip)
- frequency/wavenumber (nonstationarity)
- excess bandwidth (oversampling)
- cyclostationarity
- polarization
- finite alphabet
- ...

Pierre Comon
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (1/2)
Space translation diversity (2/2)

\( A_{iq} \): gain between sensor \( i \) and source \( q \)

\( H_{pq} \): transfer between reference and subarray \( p \)

\( S_{jq} \): sample \( j \) of source \( q \)

\( \beta_q \): path loss, \( d_q \): DOA, \( b_i \): sensor location

Tensor model (NB far field) \cite{Sidiroopoulos2000}

- Reference subarray: \( A_{iq} = \beta_q \exp(j \frac{\omega}{c} b_i^T d_q) \)

- Space translation (from reference subarray):
  \( \beta_q \exp(j \frac{\omega}{c} [b_i + \Delta_p]^T d_q) \stackrel{\text{def}}{=} A_{iq} H_{pq} \)

- Trilinear model:
  \[
  T_{ijp} = \sum_q A_{iq} S_{jq} H_{pq}
  \]

\( p \): index of subarray
Space translation diversity (2/2)

$A_{iq}$: gain between sensor $i$ and source $q$

$H_{pq}$: transfer between reference and subarray $p$

$S_{jq}$: sample $j$ of source $q$

$\beta_q$: path loss, $d_q$: DOA, $b_i$: sensor location

**Tensor model** (NB far field) [Sidiropoulos'00]

- Reference subarray: $A_{iq} = \beta_q \exp(j\frac{\omega}{C} b_i^T d_q)$

- Space translation (from reference subarray):
  
  $\beta_q \exp(\frac{\omega}{C} [b_i + \Delta_p]^T d_q) \overset{\text{def}}{=} A_{iq} H_{pq}$

- Trilinear model:

  $$T_{ijp} = \sum_q A_{iq} S_{jq} H_{pq}$$

  $p$: index of subarray
Space translation diversity (2/2)

\( A_{iq} \): gain between sensor \( i \) and source \( q \)
\( H_{pq} \): transfer between reference and subarray \( p \)
\( S_{jq} \): sample \( j \) of source \( q \)
\( \beta_q \): path loss, \( d_q \): DOA, \( b_i \): sensor location

**Tensor model** (NB far field) \([\text{Sidiropoulos'00}]\)

- Reference subarray: \( A_{iq} = \beta_q \exp(j \frac{\omega}{C} b_i^T d_q) \)
- Space translation (from reference subarray):
  \( \beta_q \exp(\frac{\omega}{C} [b_i + \Delta_p]^T d_q) \overset{\text{def}}{=} A_{iq} H_{pq} \)
- Trilinear model:
  \[
  T_{ijp} = \sum_q A_{iq} S_{jq} H_{pq}
  \]

\( p \): index of subarray
Meaning of coherence

1. Unconstrained joint source estimation & localization
   \[ \text{[Sidiropoulos'2000]} \] (ill-posed if approximate)

2. Coherence-constrained joint source estimation & localization
   \[ \text{[Lim-Comon'2010]} \]
   - time diversity: \( \mu_A \rightarrow \text{cross correlation} \)
   - space diversity: \( \mu_B, \mu_C \rightarrow \text{angular separation} \)
Unaddressed topics

- spread spectrum communications
- brain inverse problems
- medical imaging (MRI)
- NL filtering
- noise reduction
- compression (Tensor trains...)
- probability
- hyperspectral imaging
- structured tensors
- convolutive mixtures
- nonnegative factors
- ...

Algorithms
QUIZ: Why use tensors?

Main reason: essential uniqueness
- Identifiability recovery, *up to scale-permutation*

Sometimes: powerful *deterministic* approaches

Secondary reason: more sources with fewer sensors
- Matrices $A$, $B$, $C$ may have more columns than rows
QUIZ: Why use tensors?

Main reason: essential uniqueness
   ➤ Identifiability recovery, *up to scale-permutation*

Sometimes: powerful *deterministic* approaches

Secondary reason: more sources with fewer sensors
   ➤ Matrices $A$, $B$, $C$ may have more columns than rows
QUIZ: Why use tensors?

**Main reason:** essential uniqueness
- Identifiability recovery, *up to scale-permutation*

*Sometimes:* powerful *deterministic* approaches

**Secondary reason:** more sources with fewer sensors
- Matrices $A, B, C$ may have more columns than rows
QUIZ: Why use tensors?

Main reason: essential uniqueness
   ➤ Identifiability recovery, *up to scale-permutation*

Sometimes: powerful *deterministic* approaches

Secondary reason: more sources with fewer sensors
   ➤ Matrices $A$, $B$, $C$ may have more columns than rows
Other perspectives

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

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

Only when the last tree has died, the last river has been poisoned and the last fish has been caught will we realize that we cannot eat money.
Cree proverb