KATHOLIEKE UNIVERSITEIT LEUVEN

Decomposing a Third-Order Tensor in Rank-(L,L,1) Terms by Means of Simultaneous Matrix Diagonalization

Dimitri Nion & Lieven De Lathauwer

K.U. Leuven, Kortrijk campus, Belgium

E-mails: Dimitri.Nion@kuleuven-kortrijk.be

Lieven.DeLathauwer@kuleuven-kortrijk.be

2009 SIAM Conference on Applied Linear Algebra,

Session MS33 “Computational Methods for Tensors”

Monterey, USA, October 26-29, 2009

Roadmap

I. Introduction

→ Tensor decompositions: PARAFAC, Tucker, Block-Component Decompositions

II. Block-Component Decomposition in Rank-(L,L,1) Terms

→ Definition of the BCD-(L,L,1), Uniqueness bound, ALS Algorithm

III. Reformulation of BCD-(L,L,1) in terms of simultaneous matrix diagonalization

→ New algorithm, relaxed uniqueness bound

IV. An application of the BCD-(L,L,1): blind source separation in telecommunications

V. Conclusion and Future Research

Roadmap

I. Introduction

→ Tensor decompositions: PARAFAC, Tucker, Block-Component Decompositions

II. Block-Component Decomposition in Rank-(L,L,1) Terms

→ Definition of the BCD-(L,L,1), Uniqueness bound, ALS Algorithm

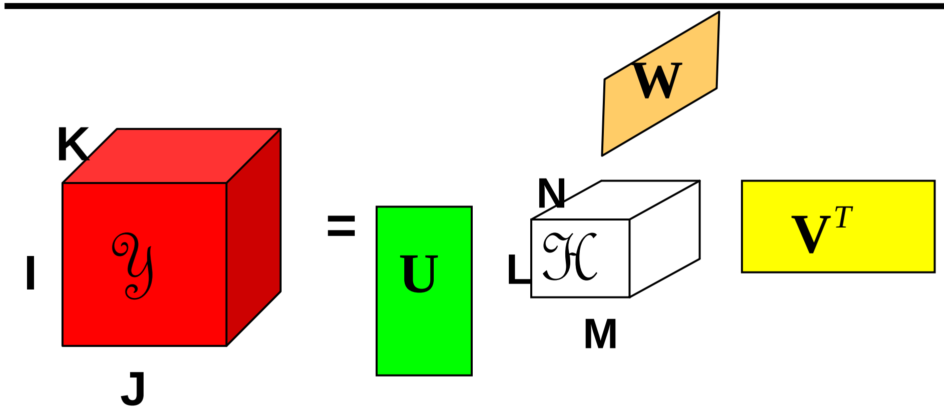
III. Reformulation of BCD-(L,L,1) in terms of simultaneous matrix diagonalization

→ New algorithm, relaxed uniqueness bound

IV. An application of the BCD-(L,L,1): blind source separation in telecommunications

V. Conclusion and Future Research

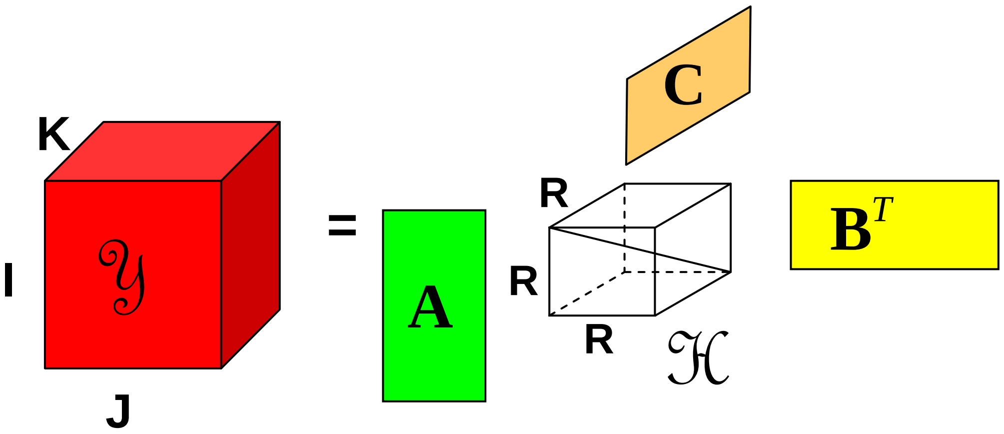
Tucker/ HOSVD and PARAFAC



[Tucker, 1966] / [De Lathauwer, 2000]

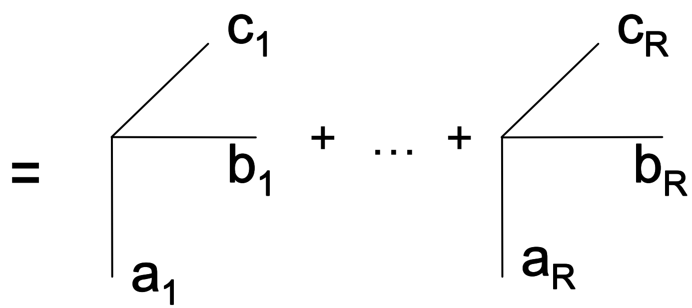
$$\mathcal{Y} = \mathcal{H} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$$

PARAFAC [Harshman, 1970]



$\mathcal{H}$ is diagonal

(if i=j=k, h_{ijk}=1, else, h_{ijk}=0)

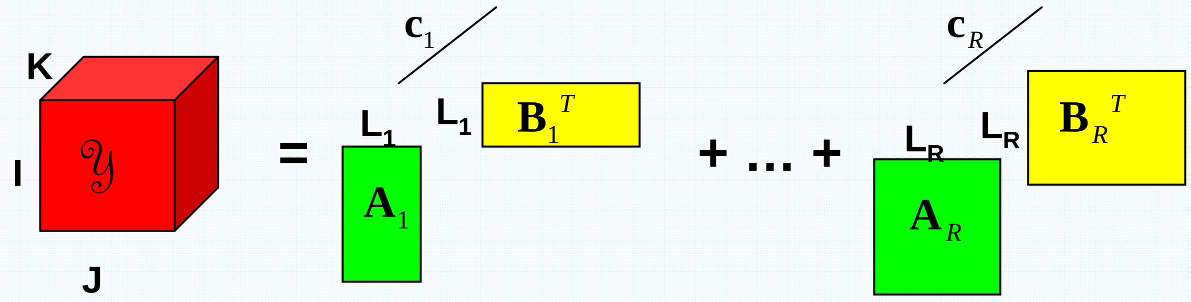


Sum of R rank-1 tensors:

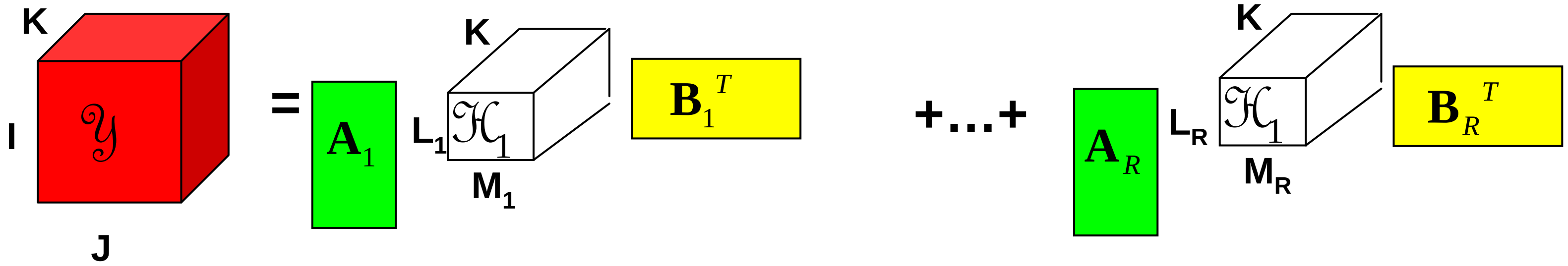
$\mathcal{Y}_1 + \ldots + \mathcal{Y}_R$

From PARAFAC/HOSVD to Block Components Decompositions (BCD) *[De Lathauwer and Nion, SIMAX 2008]*

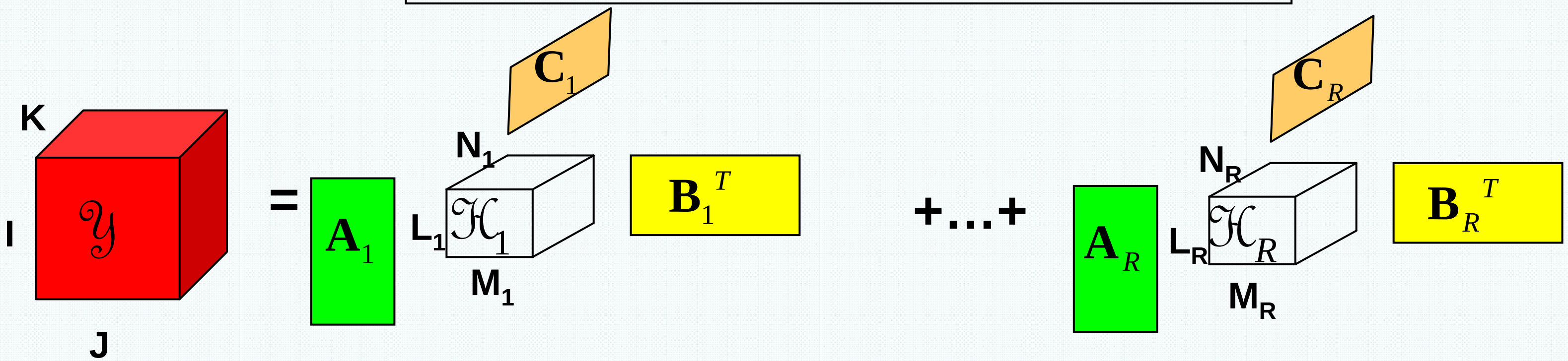
BCD in rank (L_r,L_r,1) terms



BCD in rank (L_r, M_r, .) terms



BCD in rank (L_r, M_r, N_r) terms



Roadmap

I. Introduction

→ Tensor decompositions: PARAFAC, Tucker, Block-Component Decompositions

II. Block-Component Decomposition in Rank-(L,L,1) Terms

→ Definition of the BCD-(L,L,1), Uniqueness bound, ALS Algorithm

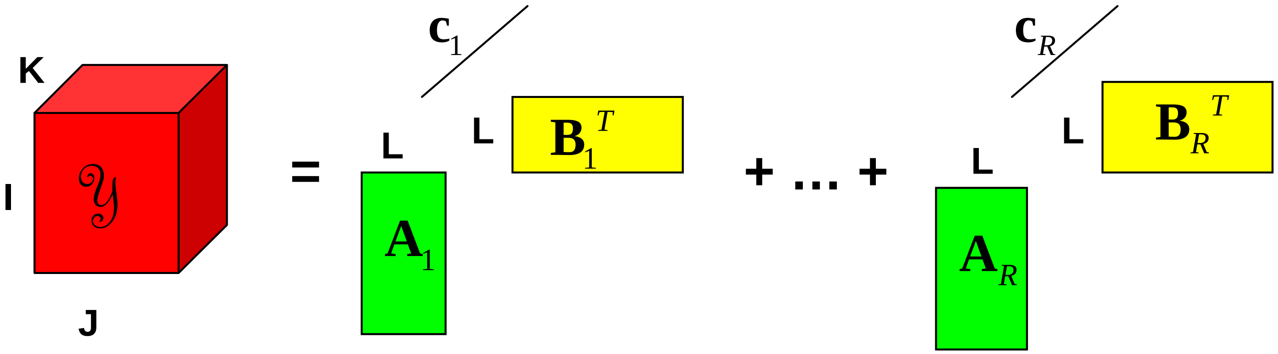
III. Reformulation of BCD-(L,L,1) in terms of simultaneous matrix diagonalization

→ New algorithm, relaxed uniqueness bound

IV. An application of the BCD-(L,L,1): blind source separation in telecommunications

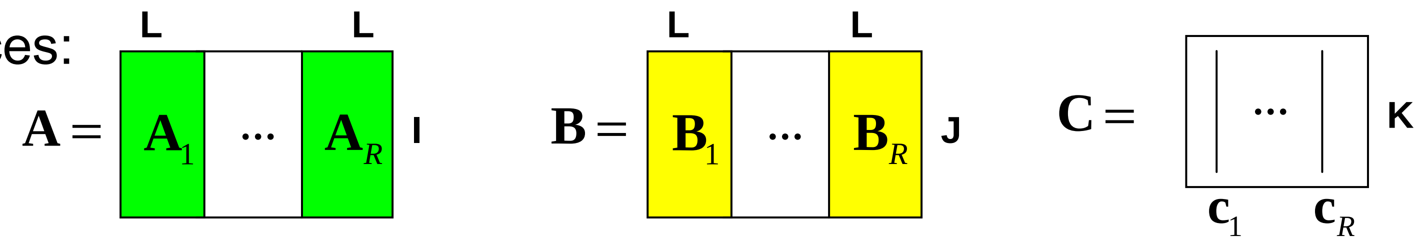
V. Conclusion and Future Research

The BCD(L,L,1) as a generalization of PARAFAC.



BCD-(L,L,1)

- Generalization of PARAFAC *[De Lathauwer, de Baynast, 2003]*
  BCD-(1,1,1)=PARAFAC
- Unknown matrices:



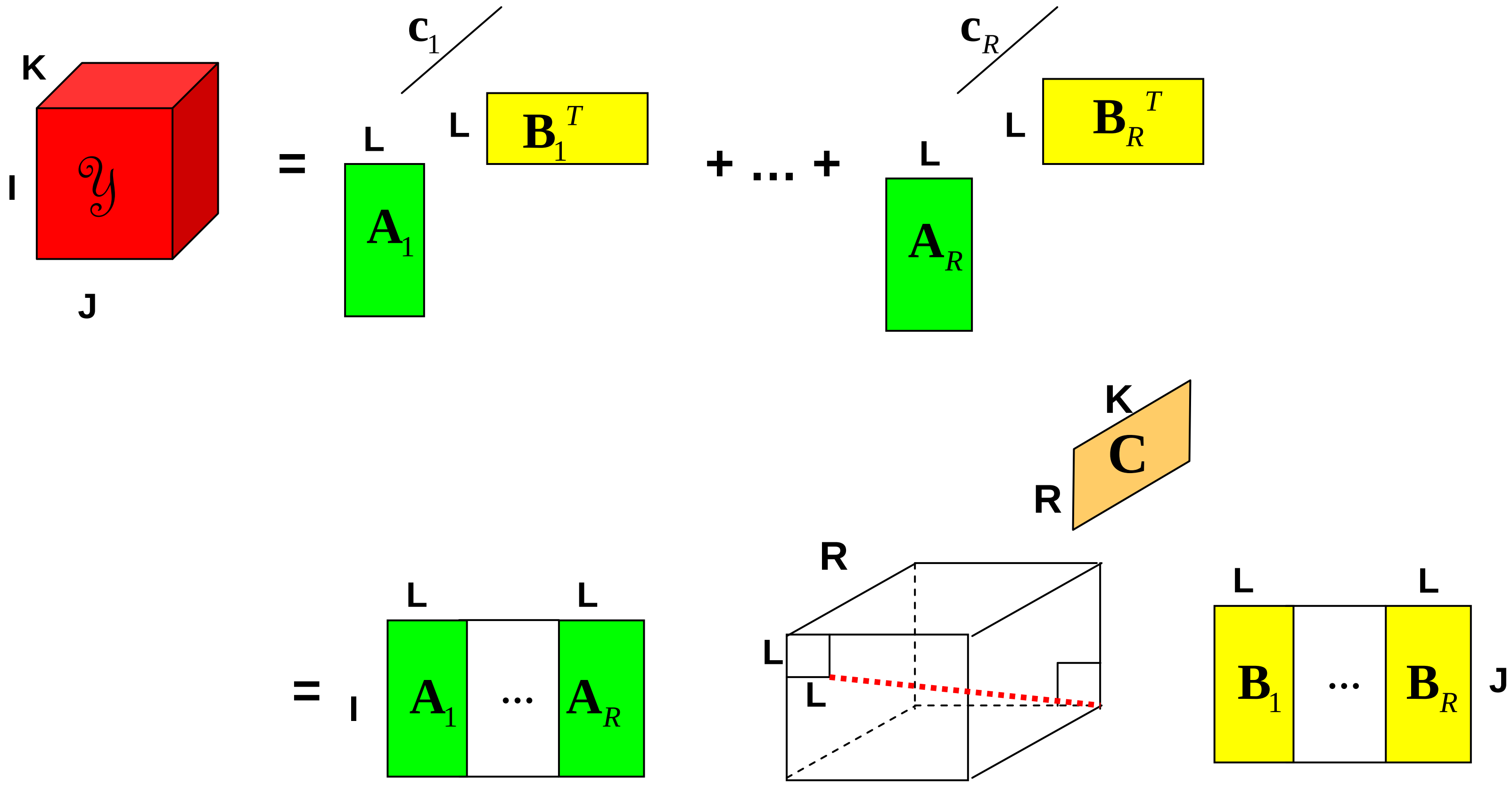
- BCD-(L,L,1) is said essentially unique if only remaining ambiguities are:

→ Arbitrary permutation of the blocks in A and B and of the columns of C

→ Rotational freedom of each block (block-wise subspace estimation) + scaling ambiguity on the columns of C

The BCD(L,L,1) as a constrained Tucker model.

The BCD-(L ,L , 1) can be seen as a particular case of Tucker model, where the core tensor is « block-diagonal », with L by L blocks on its diagonal.



BCD(L,L,1): existing results on algorithms and uniqueness

- Several usual algorithms used to compute PARAFAC have been adapted to the BCD(L,L,1).

<u>Example 1:</u> ALS algorithm (alternate between Least Squares updates of unknowns **A**, **B** and **C**).

<u>Example 2:</u> ALS with Enhanced Line Search to speed up convergence.

<u>Example 3:</u> Gauss-Newton based algorithms (Levenberg-Marquardt).

- First result on essential uniqueness, in the generic sense *[De Lathauwer, 2006]*

$$LR \leq IJ \text{ and } \min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K,R) \geq 2(R+1) \qquad (1)$$

Starting point of this work

- In 2005, De Lathauwer has shown that, under certain assumptions on the dimensions, PARAFAC can be reformulated as a simultaneous diagonalization (SD) problem. This yields:
  - A very fast and accurate algorithm to compute PARAFAC
  - A new, relaxed, uniqueness bound

- Is it possible to generalize these results to the BCD-(L,L,1)?
- If so, does it also yield a fast algoritm and a new uniqueness bound (more relaxed than the one on previous slide)?
- The answer is YES

Roadmap

I. Introduction

→ Tensor decompositions: PARAFAC, Tucker, Block-Component Decompositions

II. Block-Component Decomposition in Rank-(L,L,1) Terms

→ Definition of the BCD-(L,L,1), Uniqueness bound, ALS Algorithm

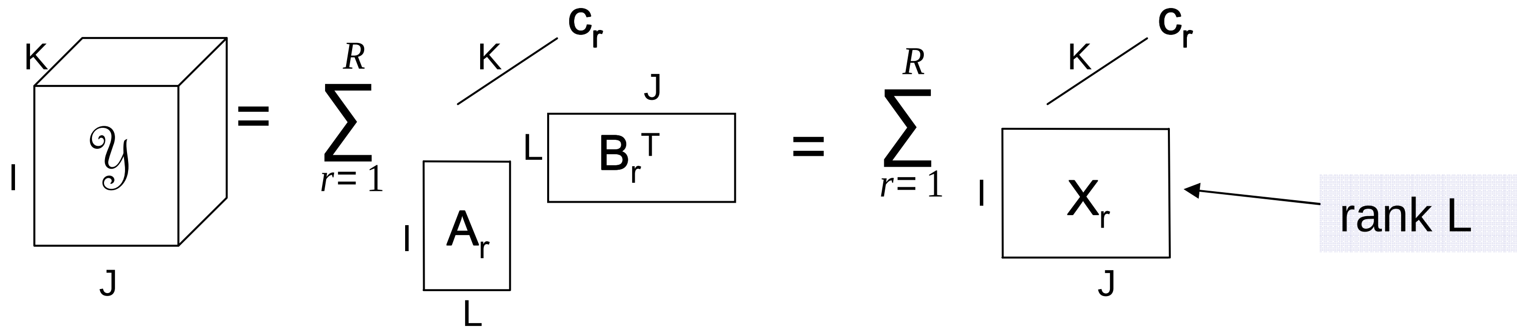
III. Reformulation of BCD-(L,L,1) in terms of simultaneous matrix diagonalization

→ New algorithm, relaxed uniqueness bound

IV. An application of the BCD-(L,L,1): blind source separation in telecommunications

V. Conclusion and Future Research

Reformulation of DCB-(L,L,1) in terms of SD: overview (1)



<u>Assumption:</u> $R \leq \min(IJ, K)$

i.e., K has to be a sufficiently long dimension

Build **Y**, the JI by K matrix unfolding of $\mathcal{Y}$

<u>BCD-(L,L,1) in matrix format :</u>

$$\mathbf{Y} = \left(vec(\mathbf{X}_1)\cdots vec(\mathbf{X}_R)\right)\cdot \mathbf{C}^T = \tilde{\mathbf{X}}\cdot \mathbf{C}^T \qquad (1)$$

<u>SVD of **Y** (generically rank-R):</u>

$$\mathbf{Y} = \mathbf{U}\cdot \mathbf{\Sigma}\cdot \mathbf{V}^H = \mathbf{E}\cdot \mathbf{V}^H \qquad (2)$$

⇒

$$\exists \mathbf{W} \in C^{R\times R}$$

$$\begin{cases} \tilde{\mathbf{X}} = \mathbf{E}\cdot \mathbf{W} \\ \mathbf{C}^T = \mathbf{W}^{-1}\cdot \mathbf{V}^H \end{cases}$$

<u>Goal:</u> Find **W**, i.e., find the linear combinations of the columns of **E** that yield vectorized rank-L matrices.

Reformulation of DCB-(L,L,1) in terms of SD: overview (2)

Note 1: Once **W** found, the unknown matrices **A**, **B**, **C** of the BCD-(L,L,1) follow

$$\begin{cases} \tilde{\mathbf{X}} = \mathbf{E} \cdot \mathbf{W} \\ \mathbf{C}^T = \mathbf{W}^{-1} \cdot \mathbf{V}^H \end{cases}$$

$$\tilde{\mathbf{X}} = \left(vec(\mathbf{X}_1) \cdots vec(\mathbf{X}_R) \right)$$
$$= \left(vec(\mathbf{A}_1 \mathbf{B}_1^T) \cdots vec(\mathbf{A}_R \mathbf{B}_R^T) \right)$$

Matricize and estimate $\mathbf{A}_1$ and $\mathbf{B}_1$ from best rank-L approximation.

Matricize and estimate $\mathbf{A}_R$ and $\mathbf{B}_R$ from best rank-L approximation.

$$\mathbf{C} = \mathbf{V}^* \cdot \mathbf{W}^{-T}$$

Note 2: For PARAFAC (i.e. L=1), we have

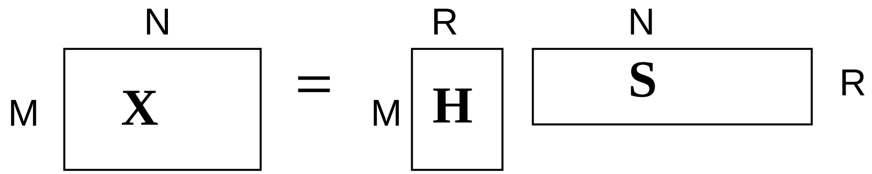
$$\tilde{\mathbf{X}} = \left(vec(\mathbf{a}_1 \mathbf{b}_1^T), \cdots, vec(\mathbf{a}_R \mathbf{b}_R^T) \right)$$
$$= \left(\mathbf{b}_1 \otimes \mathbf{a}_1, \cdots, \mathbf{b}_R \otimes \mathbf{a}_R \right)$$
$$= \mathbf{B} \circ \mathbf{A} \quad \text{where } \circ \text{ is the Khatri-Rao product}$$

⇒ $\tilde{\mathbf{X}} = \mathbf{E} \cdot \mathbf{W}$ is a Khatri-Rao structure recovery problem, and can be solved by simultaneous diagonalization *[De Lathauwer, 2005]*

Reformulation of DCB-(L,L,1) in terms of SD: overview (3)

<u>Remark:</u> on typical matrix factorization problems in Signal Processing

<u>Problem formulation:</u> Given only an (MxN) rank-R observed matrix **X**, find the (MxR) and (RxN) matrices **H** and **S** s.t. **X** = **H S**



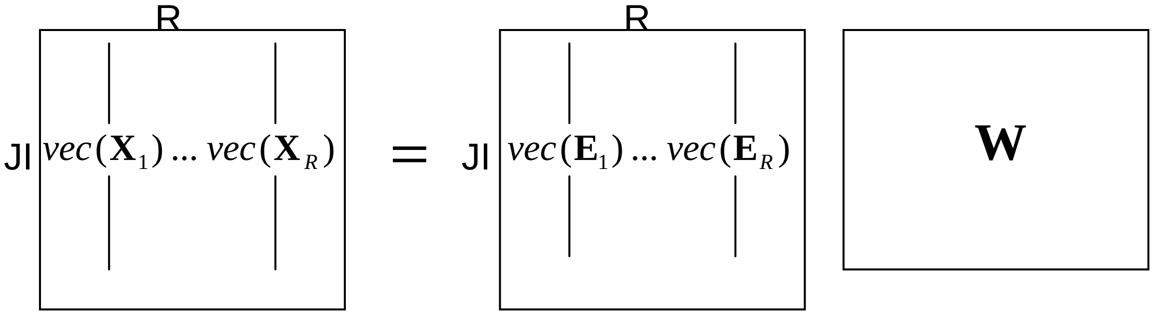
But infinite number of solutions **X** = (**HF**) ($\mathbf{F}^{-1}$**S**) so we need extra constraints.

<u>Examples:</u>

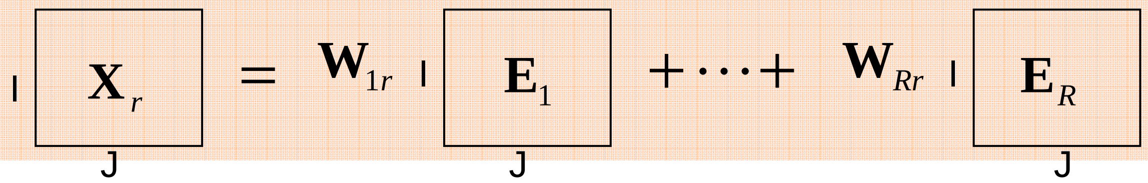
- **ICA** (Independent Component Analysis) → find **H** that makes the R source signals in **S** as much statistically independent as possible. Blind Source Separation.
- **FIR** filter estimation → **H** holds the impulse response of a FIR filter, and **S** is Toeplitz. Blind Channel Estimation in telecommunications.
- Source localization → **H** is Vandermonde and holds the individual response of the M antennas to the R source signals, each signal impinging with a Direction Of Arrival (DOA).
- Non-negative matrix factorization
- Finite Alphabet projection → **S** holds numerical symbols

Reformulation of DCB-(L,L,1) in terms of SD: overview (4)

$$\tilde{\mathbf{X}} = \mathbf{E} \cdot \mathbf{W}$$



For $r = 1 \ldots R$



How to find the coefficients of the linear combinations of the $\mathbf{E}_r$ that yield rank-L matrices?

<u>Tool:</u> mapping ϕ_L for rank-L detection. Let $\mathbf{X}_r \in C^{I \times J}$, then $\phi_L(\mathbf{X}_r, \mathbf{X}_r, \ldots, \mathbf{X}_r) = 0$
iif $\mathbf{X}_r$ is at most rank-L.

After several algebraic manipulations, one can show that **W** is solution of a SD problem

$$\begin{cases} \mathbf{Q}_1 = \mathbf{W} \cdot \mathbf{D}_1 \cdot \mathbf{W}^T \\ \mathbf{Q}_2 = \mathbf{W} \cdot \mathbf{D}_2 \cdot \mathbf{W}^T \\ \vdots \qquad\qquad \vdots \\ \mathbf{Q}_R = \mathbf{W} \cdot \mathbf{D}_R \cdot \mathbf{W}^T \end{cases}$$

Reformulation of DCB-(2,2,1) in terms of SD

Technical details

Trilinear mapping ϕ_2 for rank-2 detection:

$$\phi_2 : (\mathbf{X},\mathbf{Y},\mathbf{Z}) \in (C^{I\times J}, C^{I\times J}, C^{I\times J}) \to \phi_2(\mathbf{X},\mathbf{Y},\mathbf{Z}) \in C^{I\times I\times I\times J\times J\times J}$$

$$[\phi_2(\mathbf{X},\mathbf{Y},\mathbf{Z})]_{i_1 i_2 i_3 j_1 j_2 j_3} = \begin{vmatrix} x_{i_1 j_1} & x_{i_1 j_2} & x_{i_1 j_3} \\ y_{i_2 j_1} & y_{i_2 j_2} & y_{i_2 j_3} \\ z_{i_3 j_1} & z_{i_3 j_2} & z_{i_3 j_3} \end{vmatrix} + \begin{vmatrix} x_{i_1 j_1} & x_{i_1 j_2} & x_{i_1 j_3} \\ z_{i_2 j_1} & z_{i_2 j_2} & z_{i_2 j_3} \\ y_{i_3 j_1} & y_{i_3 j_2} & y_{i_3 j_3} \end{vmatrix} + \begin{vmatrix} y_{i_1 j_1} & y_{i_1 j_2} & y_{i_1 j_3} \\ x_{i_2 j_1} & x_{i_2 j_2} & x_{i_2 j_3} \\ z_{i_3 j_1} & z_{i_3 j_2} & z_{i_3 j_3} \end{vmatrix}$$

$$+ \begin{vmatrix} y_{i_1 j_1} & y_{i_1 j_2} & y_{i_1 j_3} \\ z_{i_2 j_1} & z_{i_2 j_2} & z_{i_2 j_3} \\ x_{i_3 j_1} & x_{i_3 j_2} & x_{i_3 j_3} \end{vmatrix} + \begin{vmatrix} z_{i_1 j_1} & z_{i_1 j_2} & z_{i_1 j_3} \\ x_{i_2 j_1} & x_{i_2 j_2} & x_{i_2 j_3} \\ y_{i_3 j_1} & y_{i_3 j_2} & y_{i_3 j_3} \end{vmatrix} + \begin{vmatrix} z_{i_1 j_1} & z_{i_1 j_2} & z_{i_1 j_3} \\ y_{i_2 j_1} & y_{i_2 j_2} & y_{i_2 j_3} \\ x_{i_3 j_1} & x_{i_3 j_2} & x_{i_3 j_3} \end{vmatrix}$$

Then we have $\phi_2(\mathbf{X},\mathbf{X},\mathbf{X}) = 0$ iif $\mathbf{X}$ is at most rank - 2.

Reformulation of DCB-(2,2,1) in terms of SD

Technical details

For $r = 1 \ldots R$

$$\underset{I \times J}{\mathbf{E}_r} = \mathbf{W}^{-1}_{1r}\,\underset{I \times J}{\mathbf{X}_1} + \cdots + \mathbf{W}^{-1}_{Rr}\,\underset{I \times J}{\mathbf{X}_R}$$

Build the set of R[3] tensors $\mathcal{P}_{rst} = \phi_2(\mathbf{E}_r, \mathbf{E}_s, \mathbf{E}_t)$ r=1,…,R, s=1,…,R, t=1,…,R

Since ϕ_2 is trilinear, we have: $\mathcal{P}_{rst} = \sum_{u,v,w=1}^{R} (\mathbf{W}^{-1})_{ur} (\mathbf{W}^{-1})_{vs} (\mathbf{W}^{-1})_{wt} \phi_2(\mathbf{X}_u, \mathbf{X}_v, \mathbf{X}_w)$

One can show that, if the $(C^3_{R+2} - R)$ tensors of the set $\mathbf{\Omega}$ are linearly independent,

$$\mathbf{\Omega} = \{\phi_2(\mathbf{X}_u, \mathbf{X}_v, \mathbf{X}_w), 1 \le u \le v \le w \le R\} - \{\phi_2(\mathbf{X}_u, \mathbf{X}_u, \mathbf{X}_u), 1 \le u \le R\},$$

then **W** is solution of

$$\mathcal{Q} = \mathcal{D} \times_1 \mathbf{W} \times_2 \mathbf{W} \times_3 \mathbf{W}$$

where $\mathcal{D}$ is an arbitrary diagonal tensor and $\mathcal{Q}$ is a symmetric tensor satisfying

$$\sum_{r,s,t}^{R} q_{rst} \mathcal{P}_{rst} = 0$$

Reformulation of DCB-(2,2,1) in terms of SD:

A new uniqueness bound

- Crucial assumption in the reformulation:

« The $(C^3_{R+2} - R)$ tensors of the set Ω are linearly independent »

- One can show that this is generically true if

$$R \leq \min(IJ, K) \text{ and } C^3_I . C^3_J \geq C^3_{R+2} - R$$

$$C^k_n = \frac{n!}{k!(n-k)!}$$

- The generalization to any value of L yields that the DCB-(L,L,1) is unique if

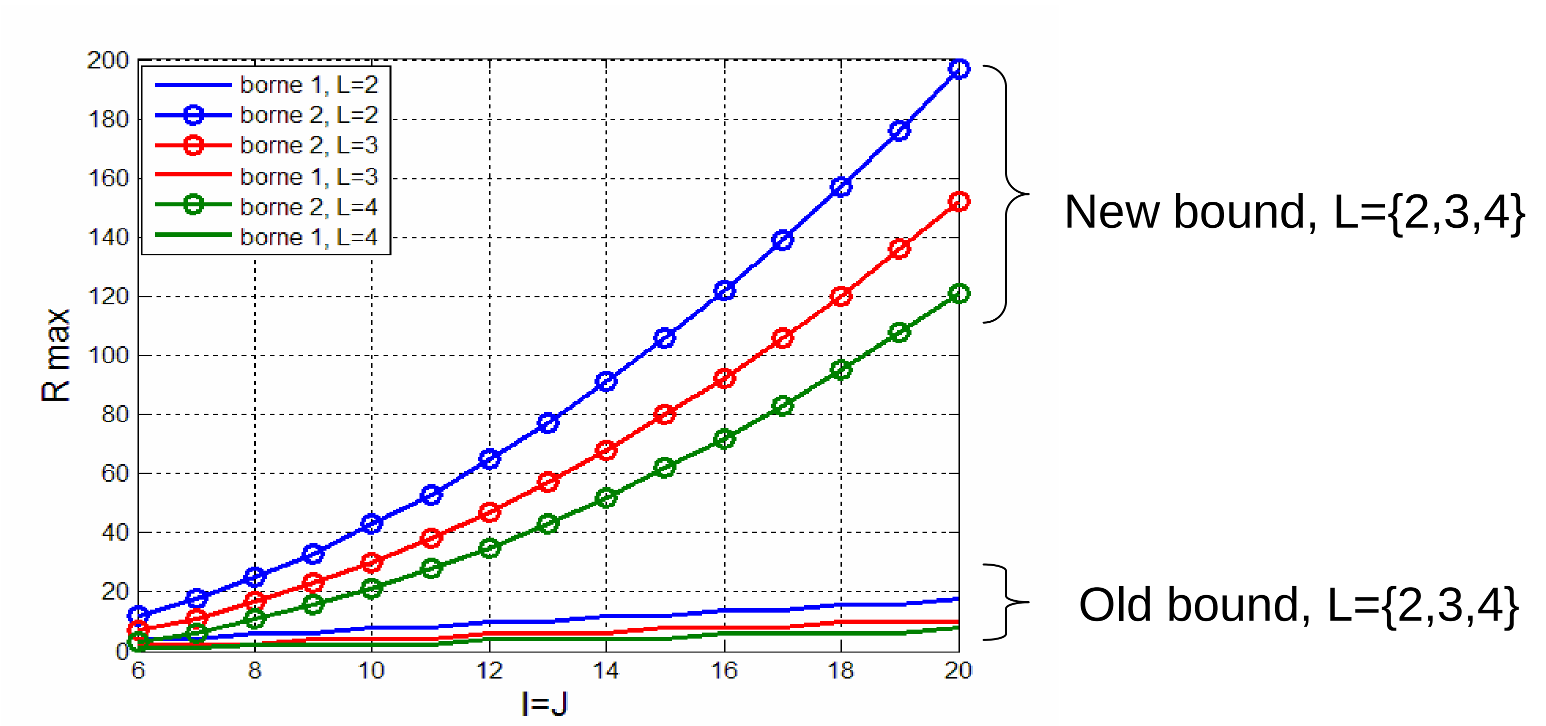
$$R \leq \min(IJ, K) \text{ and } C^{L+1}_I . C^{L+1}_J \geq C^{L+1}_{R+L} - R$$

- To be compared to the old uniqueness bound

$$LR \leq IJ \text{ and } \min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K, R) \geq 2(R+1)$$

Reformulation of DCB-(2,2,1) in terms of SD

Uniqueness



Roadmap

I. Introduction

→ Tensor decompositions: PARAFAC, Tucker, Block-Component Decompositions

II. Block-Component Decomposition in Rank-(L,L,1) Terms

→ Definition of the BCD-(L,L,1), Uniqueness bound, ALS Algorithm

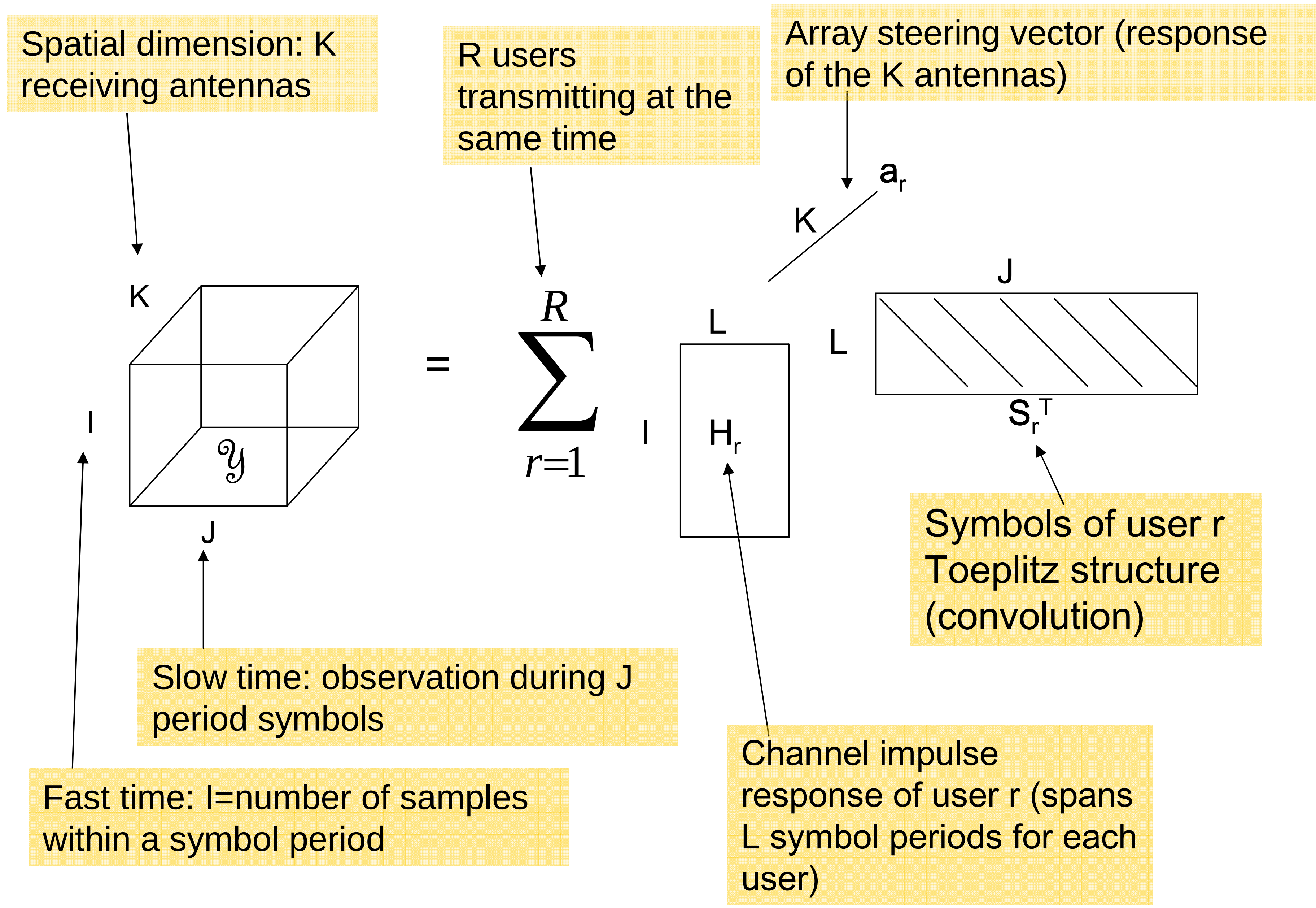
III. Reformulation of BCD-(L,L,1) in terms of simultaneous matrix diagonalization

→ New algorithm, relaxed uniqueness bound

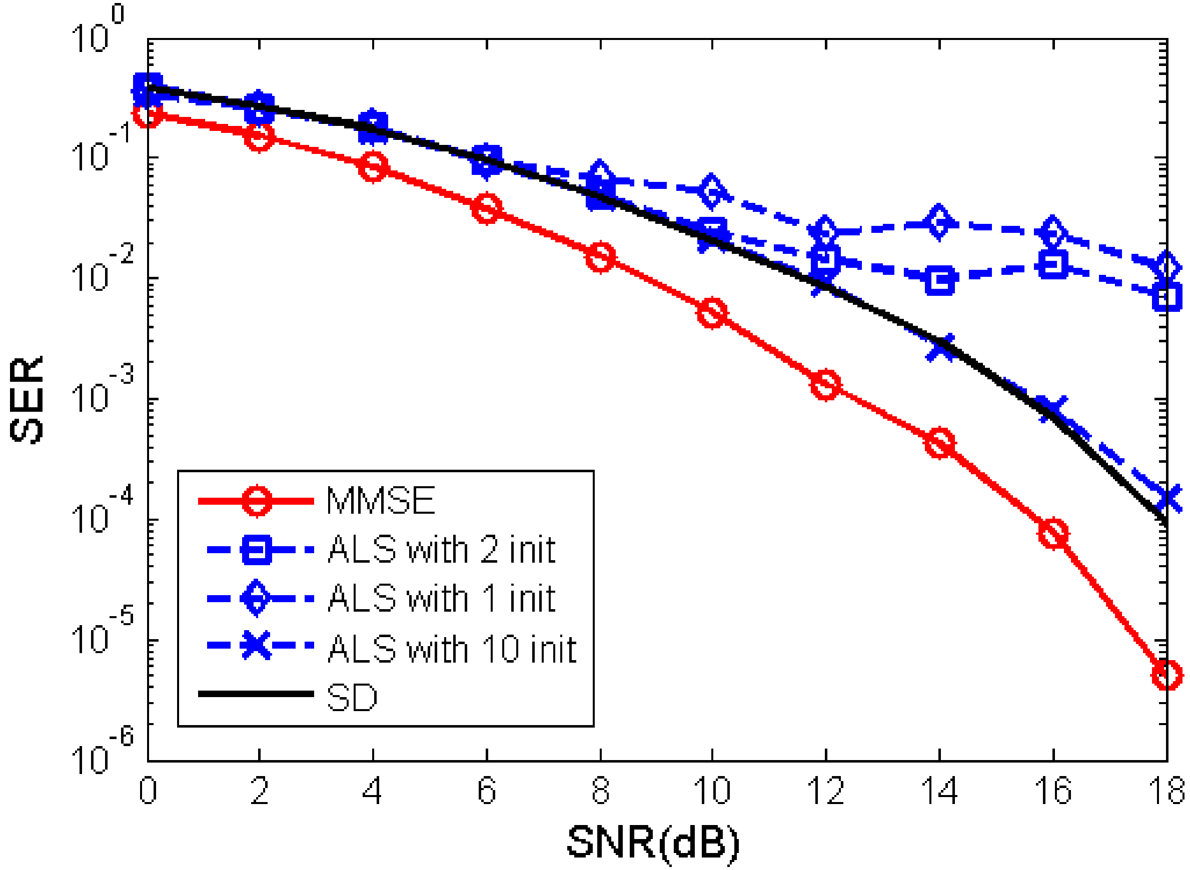
IV. An application of the BCD-(L,L,1): blind source separation in telecommunications

V. Conclusion and Future Research

Data model: DS-CDMA system



Performance: comparison between ALS and SD algorithms



Conclusion

- Reformulation of PARAFAC in terms of Simultaneous Diagonalization (SD) yields a fast and accurate algorithm, with improved identifiability results [De Lathauwer, 2005].

The starting point for this reformulation is that one dimension is long enough: $R \leq \min(IJ, K)$, where I,J and K can be interchanged.

- The BCD-(L,L,1), which is a generalization of PARAFAC, can also be formulated in terms of SD, which also yield a fast and accurate algorithm and improved identifiability result.

The starting point for this reformulation is that the third dimension (K) is long enough $R \leq \min(IJ, K)$. I,J and K *can not* be interchanged

- When the long dimension is I or J, i.e., $R \leq \min(JK, I)$ or $R \leq \min(IK, J)$

we have recently shown (CAMSAP 2009), that the BCD-(L,L,1) can be reformulated as Joint-Block-Diagonalization problem. This yields a new set of identifiability results.