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

\[ Y = \mathcal{H} \times_1 U \times_2 V \times_3 W \]


PARAFAC [Harshman, 1970]

\( \mathcal{H} \) is diagonal

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

Sum of R rank-1 tensors:

\[ y_1 + \ldots + y_R \]
From PARAFAC/HOSVD to Block Components Decompositions (BCD) [De Lathauwer and Nion, SIMAX 2008]

**BCD in rank \((L_r, L_r, 1)\) terms**

\[
Y_{IJK} = \sum_{c_l} A_{l}^{T} B_{l} + \ldots + \sum_{c_R} A_{R}^{T} B_{R}^{T}
\]

**BCD in rank \((L_r, M_r, \ldots)\) terms**

\[
Y_{IJK} = \sum_{c_l} A_{l}^{T} B_{l} + \ldots + \sum_{c_R} A_{R}^{T} B_{R}^{T}
\]

**BCD in rank \((L_r, M_r, N_r)\) terms**

\[
Y_{IJK} = \sum_{c_l} A_{l}^{T} B_{l} + \ldots + \sum_{c_R} A_{R}^{T} B_{R}^{T}
\]
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.

Generalization of PARAFAC [De Lathauwer, de Baynast, 2003]

BCD-(1,1,1)=PARAFAC

Unknown matrices:

\[
\begin{align*}
A &= \begin{bmatrix}
A_1 & \ldots & A_R
\end{bmatrix}
\end{align*}
\]

\[
\begin{align*}
B &= \begin{bmatrix}
B_1 & \ldots & B_R
\end{bmatrix}
\end{align*}
\]

\[
\begin{align*}
C &= \begin{bmatrix}
\vdots \\
c_1 & \ldots & c_R
\end{bmatrix}
\end{align*}
\]

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).
  
  **Example 1:** ALS algorithm (alternate between Least Squares updates of unknowns \( A \), \( B \) and \( C \)).
  
  **Example 2:** ALS with Enhanced Line Search to speed up convergence.
  
  **Example 3:** Gauss-Newton based algorithms (Levenberg-Marquardt).

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

\[
LR \leq IJ \quad \text{and} \quad \min\left(\frac{I}{L}, R\right) + \min\left(\frac{J}{L}, R\right) + \min(K, R) \geq 2(R+1) \quad (1)
\]
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 algorithm 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)

\[
Y = \sum_{r=1}^{R} \left[ \begin{array}{c} A_r \\ B_r \end{array} \right] \cdot \left[ \begin{array}{c} C_r \\ 0 \end{array} \right] = \sum_{r=1}^{R} \left[ \begin{array}{c} X_r \\ 0 \end{array} \right] = X_r
\]

Assumption: \( R \leq \min(IJ, K) \)

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

Build \( Y \), the \( JI \) by \( K \) matrix unfolding of \( \tilde{y} \)

BCD-(L,L,1) in matrix format:

\[
Y = (\text{vec}(X_1) \cdots \text{vec}(X_R)) \cdot C^T = \tilde{X} \cdot C^T \tag{1}
\]

SVD of \( Y \) (generically rank-\( R \)):

\[
Y = U \cdot \Sigma \cdot V^H = E \cdot V^H \tag{2}
\]

Goal: 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

$$\begin{align*}
\hat{X} &= E \cdot W \\
C^T &= W^{-1} \cdot V^H
\end{align*}$$

$$\begin{align*}
\hat{X} &= \left( \text{vec}(X_1) \cdots \text{vec}(X_R) \right) \\
&= \left( \text{vec}(A_1B_1^T) \cdots \text{vec}(A_RB_R^T) \right)
\end{align*}$$

Matricize and estimate $A_1$ and $B_1$ from best rank-L approximation.

Matricize and estimate $A_R$ and $B_R$ from best rank-L approximation.

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

$$\begin{align*}
\hat{X} &= \left( \text{vec}(a_1b_1^T), \cdots, \text{vec}(a_Rb_R^T) \right) \\
&= (b_1 \otimes a_1, \cdots, b_R \otimes a_R) \\
&= B \circ A \quad \text{where} \quad \circ \quad \text{is the Khatri-Rao product}
\end{align*}$$

$\hat{X} = E \cdot W$ is a Khatri-Rao structure recovery problem, and can be solved by simultaneous diagonalization [De Lathauwer, 2005]
Remark: on typical matrix factorization problems in Signal Processing

Problem formulation: Given only an \((M\times N)\) rank-\(R\) observed matrix \(X\), find the \((M\times R)\) and \((R\times N)\) matrices \(H\) and \(S\) s.t. \(X = HS\)

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

Examples:

- **ICA** (Independent Component Analysis) \(\rightarrow\) find \(H\) that makes the \(R\) source signals in \(S\) as much statistically independent as possible. **Blind Source Separation**.

- **FIR** filter estimation \(\rightarrow\) \(H\) holds the impulse response of a FIR filter, and \(S\) is Toeplitz. **Blind Channel Estimation in telecommunications**.

- **Source localization** \(\rightarrow\) \(H\) 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** \(\rightarrow\) \(S\) holds numerical symbols
Reformulation of DCB-\((L,L,1)\) in terms of SD: overview (4)

\[
\tilde{X} = E \cdot W
\]

For \( r = 1 \ldots R \)

\[
\begin{align*}
X_r & = W_{1r} L_1 + \cdots + W_{Rr} L_R
\end{align*}
\]

How to find the coefficients of the linear combinations of the \( E_r \) that yield rank-\( L \) matrices?

Tool: mapping \( \phi_L \) for rank-\( L \) detection. Let \( X_r \in C^{I \times J} \), then \( \phi_L(X_r, X_r, \ldots, X_r) = 0 \) iif \( X_r \) is at most rank-\( L \).

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

\[
\begin{align*}
Q_1 & = W \cdot D_1 \cdot W^T \\
Q_2 & = W \cdot D_2 \cdot W^T \\
& \vdots \\
Q_R & = W \cdot D_R \cdot W^T
\end{align*}
\]
Reformulation of DCB-(2,2,1) in terms of SD

Technical details

Trilinear mapping $\phi_2$ for rank-2 detection:

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

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

Then we have $\phi_2 (X, X, X) = 0$ iif $X$ is at most rank - 2.
Reformulation of DCB-(2,2,1) in terms of SD

Technical details

For $r = 1\ldots R$

\[
\begin{bmatrix}
E_r
\end{bmatrix}_J = W^{-1}_{1r} \begin{bmatrix}
X_1
\end{bmatrix}_J + \cdots + W^{-1}_{Rr} \begin{bmatrix}
X_R
\end{bmatrix}_J
\]

Build the set of $R^3$ tensors $\mathcal{P}_{rst} = \phi_2(E_r, E_s, E_t)$ for $r=1,\ldots,R$, $s=1,\ldots,R$, $t=1,\ldots,R$

Since $\phi_2$ is trilinear, we have:

\[
\mathcal{P}_{rst} = \sum_{u,v,w=1}^{R} (W^{-1})_{ur} (W^{-1})_{vs} (W^{-1})_{wt} \phi_2(X_u, X_v, X_w)
\]

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

\[
\Omega = \left\{ \phi_2(X_u, X_v, X_w), 1 \leq u \leq v \leq w \leq R \right\} - \left\{ \phi_2(X_u, X_u, X_u), 1 \leq u \leq R \right\}
\]

then $W$ is solution of

\[
\mathcal{Q} = D \times_1 W \times_2 W \times_3 W
\]

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

\[
\sum_{r,s,t} q_{rst} \mathcal{P}_{rst} = 0
\]
Reformulation of DCB-(2,2,1) in terms of SD:

A new uniqueness bound

- Crucial assumption in the reformulation:
  
  « The \((C_{R+2}^3 - R)\) tensors of the set \(\Omega\) are linearly independent »

- One can show that this is generically true if
  
  \[
  R \leq \min(IJ, K) \quad \text{and} \quad C_i^R.C_j^R \geq C_{R+2}^3 - R
  \]

  \[
  C_n^k = \frac{n!}{k!(n-k)!}
  \]

- The generalization to any value of \(L\) yields that the DCB-(L,L,1) is unique if
  
  \[
  R \leq \min(IJ, K) \quad \text{and} \quad C_i^{L+1}.C_j^{L+1} \geq C_{R+L}^{L+1} - R
  \]

- To be compared to the old uniqueness bound
  
  \[
  LR \leq IJ \quad \text{and} \quad \min \left( \frac{I}{L}, R \right) + \min \left( \frac{J}{L}, R \right) + \min(K, R) \geq 2(R+1)
  \]
Reformulation of DCB-(2,2,1) in terms of SD

Uniqueness

New bound, $L=\{2,3,4\}$

Old bound, $L=\{2,3,4\}$
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

Spatial dimension: K receiving antennas

R users transmitting at the same time

Array steering vector (response of the K antennas)

Channel impulse response of user r (spans L symbol periods for each user)

Symbols of user r Toeplitz structure (convolution)

Fast time: \( l \) = number of samples within a symbol period

Slow time: observation during \( J \) period symbols
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.