Block Component Analysis

Lieven De Lathauwer

K.U.Leuven
Belgium
Lieven.DeLathauwer@kuleuven-kortrijk.be
Lieven.DeLathauwer@esat.kuleuven.be
Overview

- Algebraic preliminaries
  - Multilinear rank and associated decomposition
  - Rank and associated decomposition
- Factor analysis and blind source separation
- Block Term Decompositions
- Block Component Analysis
- Variants
- Conclusion
**Multilinear rank and associated decomposition**

**Multilinear rank:** (dimension column space, dimension row space, ...)  
[Hitchcock, 1927]

**Decomposition:** \( \mathcal{A} = S \cdot _1 U^{(1)} \cdot _2 U^{(2)} \cdot _3 \ldots \cdot _N U^{(N)} \)

In multilinear SVD: \( S \) is all-orthogonal and ordered \( U^{(1)}, U^{(2)}, \ldots, U^{(N)} \) are orthogonal

\[ [Tucker '64], [De Lathauwer '00] \]

Cf. subspace variety
**Rank and associated decomposition**

**Rank**: minimal number of rank-1 terms \([Hitchcock, 1927]\)

Canonical Polyadic Decomposition (CPD) of a tensor \(\mathcal{A}\) is its decomposition in a minimal sum of rank-1 tensors

\[
\mathcal{A} = \lambda_1 u_1^{(1)} + \lambda_2 u_2^{(1)} + \ldots + \lambda_R u_R^{(1)}
\]

\([Harshman ’70], [Carroll and Chang ’70]\)

Unique under mild conditions

Cf. \(R\)-th secant variety of Segre variety (unsymmetric), Veronese variety (symmetric)
Overview

• Algebraic preliminaries
• Factor analysis and blind source separation
  – Principal Component Analysis (PCA)
  – Canonical Polyadic Analysis (CPA)
  – Independent Component Analysis (ICA)
• Block Term Decompositions
• Block Component Analysis
• Variants
• Conclusion
Factor analysis and blind source separation

- Decompose a data matrix in rank-1 terms
  E.g. independent component analysis, telecommunication, biomedical applications, chemometrics, exploratory data analysis, ... 

\[
A = F \cdot G^T 
\]

\[
\begin{bmatrix}
A \\
\end{bmatrix} = \begin{bmatrix}
g_1 \\
g_2 + \ldots + \\
g_R \\
\end{bmatrix} \begin{bmatrix}
f_1 \\
f_2 \\
f_R \\
\end{bmatrix}
\]

- Decomposition in rank-1 terms is not unique

\[
A = (FM) \cdot (M^{-1}G^T) 
\]

\[
= \tilde{F} \cdot \tilde{G}^T 
\]
Exploitation of prior knowledge

PCA, SVD: uniqueness obtained by adding orthogonality constraints

\[ A = U^{(1)} \cdot \Sigma \cdot U^{(2)^T} \]

\( U^{(1)}, U^{(2)} \) orthogonal, \( \Sigma \) diagonal
Example: excitation-emission fluorescence in chemometrics

Matrix approach

row vector $\sim$ emission spectrum
column vector $\sim$ excitation spectrum
coefficients $\sim$ concentrations

\[
\begin{bmatrix}
A
\end{bmatrix} = \begin{bmatrix}
\lambda_1 & \lambda_2 & \ldots & \lambda_R
\end{bmatrix}
\begin{bmatrix}
\mathbf{u}_1^{(1)} & \mathbf{u}_2^{(1)} & \ldots & \mathbf{u}_R^{(1)}
\end{bmatrix}
\]
Tensor solution: CP Analysis

Tensorization: one matrix $\rightarrow$ several matrices, stacked in tensor

row vector $\sim$ emission spectrum
column vector $\sim$ excitation spectrum
coefficients $\sim$ concentrations

\[ A = \left[ \begin{array}{c} \lambda_1 u_1^{(3)} \\ \lambda_2 u_2^{(3)} \\ \vdots \end{array} \right] = \lambda_1 u_1^{(1)} + \lambda_2 u_2^{(1)} + \ldots + \lambda_R u_R^{(1)} \]

[Smilde, Geladi, Bro '04]
Independent Component Analysis (ICA)

Model:

\[ Y = MX \]

- statistical independence implies:
  - the variables are uncorrelated
  - additional conditions on the HOS
ICA: basic equations

Model:

\[ Y = MX \]

Second order:

\[
\begin{align*}
C_2^Y &= E\{YY^T\} \\
      &= M \cdot C_2^X \cdot M^T \\
      &= C_2^X \cdot_1 M \cdot_2 M \\
\end{align*}
\]

uncorrelated sources: \( C_2^X \) is diagonal

“diagonalization by congruence”

\[
\begin{bmatrix}
\sigma_1^2 \\
\sigma_2^2 \\
\vdots \\
\sigma_R^2
\end{bmatrix}_1 \begin{bmatrix}
m_1 \\
m_2 \\
\vdots \\
m_R
\end{bmatrix} + \ldots = \begin{bmatrix}
m_1 \\
m_2 \\
\vdots \\
m_R
\end{bmatrix}
\]
Higher order:

\[ C^Y_4 = C^X_4 \cdot_1 M \cdot_2 M \cdot_3 M \cdot_4 M \]

independent sources: \( C^X_4 \) is diagonal

Tensorization: decomposition data matrix \( \rightarrow \) CPD cumulant tensor

[Comon '94]
Overview

- Algebraic preliminaries
- Factor analysis and blind source separation
- Block Term Decompositions
- Block Component Analysis
- Variants
- Conclusion
**Rank and associated decomposition**

**Rank:** minimal number of rank-1 terms  

[**Hitchcock, 1927**]

**Canonical Polyadic Decomposition (CPD)** of a tensor $\mathcal{A}$ is its decomposition in a minimal sum of rank-1 tensors

$$\mathcal{A} = \lambda_1 u_1^{(1)} + \lambda_2 u_2^{(1)} + \ldots + \lambda_R u_R^{(1)}$$

[**Harshman ’70**, **Carroll and Chang ’70**]

Unique under mild conditions

Cf. $R$-th secant variety of Segre variety (unsymmetric), Veronese variety (symmetric)
Decomposition in rank-$(L, L, 1)$ terms

\[ \mathcal{A} = I_1 \quad U_1^{(1)} + \ldots + L S_I \quad U_1^{(2)} + \ldots + L S_I \quad U_1^{(3)} \]

Unique under mild conditions

[De Lathauwer '06]
Decomposition in rank-\((R_1, R_2, R_3)\) terms

Unifies: multilinear rank and rank Tucker and CPD

[De Lathauwer '06]
Overview

- Algebraic preliminaries
- Factor analysis and blind source separation
- Block Term Decompositions
- Block Component Analysis
- Variants
- Conclusion
Block Component Analysis

Demo
Wireless communication

e.g. Direct Sequence Code Division Multiple Access

DS-CDMA signal $\sim$ rank-1 matrix

Signal transmitted by user $r$:

$$
\begin{align*}
S_{1r} [c_{1r} c_{2r} \cdots c_{jr}] \\
S_{2r} [c_{1r} c_{2r} \cdots c_{jr}] \\
S_{3r} [c_{1r} c_{2r} \cdots c_{jr}] \\
\vdots \\
\sim S_r c_r^T = s_r \odot c_r^T
\end{align*}
$$
DS-CDMA signal received by antenna array $\sim$ rank-1 tensor

Dimension 1 $\sim$ temporal diversity
Dimension 2 $\sim$ spectral diversity
Dimension 3 $\sim$ spatial diversity

\[
T = \sum_{r=1}^{R} s_r a_r c_r
\]

Deterministic blind signal separation based on CPD

[Sidiropoulos et al. ’00]
Generalization propagation model:

- CPD: line-of-sight propagation (no ISI)
- decomposition in rank-\((L_r, L_r, 1)\) terms: reflexions in far field of antenna array
  
  \[\textit{De Lathauwer and de Baynast '08}\]

- decomposition in rank-\((L_r, M_r, \cdot)\) terms: reflexions not only in far field of antenna array
  
  \[\textit{Nion and De Lathauwer '08}\]

General comment: rank-1 terms are easiest to compute
rank-1 terms may be too simple
**Exponentials, sinusoids, polynomials, exponential polynomials**

**Principle:** Map every row of $A = F \cdot G^T$ to Hankel matrix

- Hankel matrices are often very ill-conditioned
- Hankel matrices generated by exponential polynomials are exactly low-rank
theoretical values: \((L_1, L_2) = (2, 3)\)
perfect separation: \((L_1, L_2) = (2, 3), (3, 3), (2, 4), (3, 4), (4, 4)\)
good separation: \((L_1, L_2) = (2, 2), (1, 2)\)
501 samples, SNR = 5 dB

good separation: \((L_1, L_2) = (1, 2), (2, 2), (2, 3)\)
theoretical values: $L_1 = 2$, $L_2 = 251$
theoretical values: \( L_1 = 2, L_2 = 251 \)
results: \( L_1 = 2, L_2 = 2, 3, \ldots, 7 \)
Audio signals

Chirp (top) and train (bottom) signal, 31 samples
singular values of Hankel matrices generated by chirp (left) and train (right)
top: 31 samples; bottom: 1000 samples
L. De Lathauwer

<table>
<thead>
<tr>
<th>$L_1 / L_2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>48</td>
<td>49</td>
<td>37</td>
<td>20</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>47</td>
<td>49</td>
<td>48</td>
<td>44</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>47</td>
<td>23</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>48</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>44</td>
<td>23</td>
<td>47</td>
<td>45</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>17</td>
<td>20</td>
<td>20</td>
<td>29</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>16</td>
<td>19</td>
<td>18</td>
<td>16</td>
<td>33</td>
<td>24</td>
</tr>
</tbody>
</table>

**mean SIR [dB] (Hankel, noiseless)**

(ICA: COM2: 15 dB, JADE: 14 dB)

<table>
<thead>
<tr>
<th>$L_1 / L_2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>47</td>
<td>49</td>
<td>51</td>
<td>51</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>47</td>
<td>50</td>
<td>49</td>
<td>51</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>50</td>
<td>49</td>
<td>48</td>
<td>49</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>49</td>
<td>48</td>
<td>47</td>
<td>48</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>51</td>
<td>51</td>
<td>49</td>
<td>48</td>
<td>48</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>38</td>
<td>47</td>
<td>46</td>
<td>46</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>22</td>
<td>45</td>
<td>44</td>
<td>44</td>
<td>47</td>
<td>44</td>
</tr>
</tbody>
</table>

**median SIR [dB] (Hankel, noiseless)**

(ICA: COM2: 15 dB, JADE: 14 dB)
singular values of wavelet matrices generated by chirp (left) and train (right)
top: 31 samples; bottom: 1000 samples
### mean SIR [dB] (wavelet, noiseless)

<table>
<thead>
<tr>
<th>$L_1 / L_2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>41</td>
<td>19</td>
<td>19</td>
<td>16</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>33</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>47</td>
<td>46</td>
<td>27</td>
<td>25</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>48</td>
<td>27</td>
<td>52</td>
<td>16</td>
<td>47</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>33</td>
<td>25</td>
<td>16</td>
<td>13</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>17</td>
<td>18</td>
<td>47</td>
<td>28</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

### median SIR [dB] (wavelet, noiseless)

<table>
<thead>
<tr>
<th>$L_1 / L_2$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>16</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>46</td>
<td>49</td>
<td>52</td>
<td>33</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>49</td>
<td>46</td>
<td>41</td>
<td>32</td>
<td>33</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>52</td>
<td>41</td>
<td>52</td>
<td>21</td>
<td>48</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>33</td>
<td>32</td>
<td>21</td>
<td>13</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>18</td>
<td>33</td>
<td>48</td>
<td>35</td>
<td>47</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>17</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>25</td>
<td>20</td>
</tr>
</tbody>
</table>
Results for noisy data:

<table>
<thead>
<tr>
<th>SNR [dB]</th>
<th>SIR [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>60</td>
<td>65</td>
</tr>
</tbody>
</table>

- BCA Hankel L=1
- BCA Hankel L=2
- BCA Hankel L=3
- BCA Hankel L=4
- BCA wavelet L=1
- BCA wavelet L=2
- BCA wavelet L=3
- BCA wavelet L=4
- ICA COM2
Conclusion

- In BCA signals are separated on the basis of low intrinsic dimensionality
- Foundation related to Pareto principle, compressed sensing, etc.
  Q: How many samples are needed for separation?
- Data can be tensorized in many ways (HOS, Hankel, . . .)
- In some cases, the rank-1 hypothesis is questionable
- PCA, ICA, CPA, NMF, . . .: easier to use but assumptions should be verified
- Constrained BCA: nonnegativity, sparsity, orthogonality, statistical independence, . . .

Related work: CPA with independence constraint

[De Vos, Van Huffel, De Lathauwer ’10]
CPA with orthogonality constraint

[Sørensen, De Lathauwer, Deneire ’10]