The Parafac Model for Multi-way Data Analysis

Alwin Stegeman
University of Groningen
The Netherlands

a.w.stegeman@rug.nl
http://www.ppsw.rug.nl/~stegeman
Contents

• 2-way Data Analysis
  Singular Value Decomposition (SVD)
  Principal Components Analysis (PCA)

• 3-way Data Analysis
  3-way rank
  Candecomp / Parafac (CP)
  CP compared to 2-way SVD and PCA
  Tucker Model
Contents

• The Parafac Model

  Uniqueness Properties
  Alternating Least Squares algorithm
  Preprocessing of the Data
  Example of a Parafac Analysis
  Including restrictions in Parafac
  Extension to multi-way Parafac
  Degenerate Parafac solutions
2-way Data Analysis

2-way data $\Rightarrow$ 2-way array $= \text{matrix}$

notation: $x$ scalar $1 \times 1$

$x$ column vector $n \times 1$

$x^T$ row vector $1 \times n$

$X$ matrix $n \times m$

row-rank($X$) = max # linearly independent rows
$= \text{dim(row space)}$

column-rank($X$) = max # linearly independent columns
$= \text{dim(column space)}$

rank($X$) = row-rank($X$) = column-rank($X$) $\leq \min(n,m)$
Examples

\[
X = \begin{bmatrix}
1 & 4 & 5 \\
2 & 5 & 7 \\
3 & 6 & 9
\end{bmatrix}
\]

column 1 + column 2 = column 3
columns 1 and 2 are linearly independent
\[\Rightarrow\] column-rank(\(X\)) = 2 \[\Rightarrow\] rank(\(X\)) = 2

\[
X = a \ b^T = a \circ b \quad \Leftrightarrow \quad x_{ij} = a_i \ b_j
\]

all columns of \(X\) are scalar multiples \[\Rightarrow\] rank(\(X\)) = 1
if \(X\) has a nonzero column
Let $X$ be an $n \times m$ matrix with $n \geq m$ and $\text{rank}(X) = R$.

Then the SVD of $X$ is $X = U S V^T$.

where $U$ is $n \times m$ and $U^T U = I_m$

$V$ is $m \times m$ and $V^T V = VV^T = I_m$

$S = \text{diag}\{s_1, \ldots, s_R, 0, \ldots, 0\}$ is $m \times m$

columns of $U$ are mutually orthogonal and have length 1

columns of $V$ are mutually orthogonal and have length 1

singular values of $X$ are $s_1 \geq \ldots \geq s_R > 0 \geq \ldots \geq 0$
\[
\text{rank}(\mathbf{X}) = R = \# \text{ positive singular values of } \mathbf{X}
\]

SVD \(\Rightarrow\) \(
\mathbf{X} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + s_R \mathbf{u}_R \mathbf{v}_R^T
\)

with \(\mathbf{u}_j\) and \(\mathbf{v}_j\) the \(j\)-th columns of \(\mathbf{U}\) and \(\mathbf{V}\)

\[
\text{rank}(\mathbf{u}_j \mathbf{v}_j^T) = 1 \Rightarrow \text{SVD decomposes } \mathbf{X} \text{ into } R \text{ rank-1 matrices}
\]

"economy size SVD" \(\mathbf{U}_R\) is \(n \times R\) and \((\mathbf{U}_R)^T \mathbf{U}_R = \mathbf{I}_R\)
\(\mathbf{V}_R\) is \(m \times R\) and \((\mathbf{V}_R)^T \mathbf{V}_R = \mathbf{I}_R\)
\(\mathbf{S}_R = \text{diag}\{s_1, \ldots, s_R\}\) is \(R \times R\)

\(\Rightarrow\) columns of \(\mathbf{U}_R\) and \(\mathbf{V}_R\) are unique up to sign if the singular values are all distinct
**Theorem**  Let the rank-\(p\) matrix \(Y\) (with \(p \leq R\)) be given by the truncated SVD of \(X\), i.e.

\[
Y = s_1 u_1 v_1^T + \ldots + s_p u_p v_p^T = U_p S_p (V_p)^T.
\]

Then \(Y\) is a best rank-\(p\) approximation of \(X\), i.e.

\[
\|X - Y\|^2 = \sum_{i,j} (x_{ij} - y_{ij})^2 \quad \text{is minimal.}
\]

\[\square\]

\(\Rightarrow\) **SVD** gives all best low-rank approximations of \(X\)

\(\Rightarrow\) \(\text{rank}(X) = \text{smallest } \# \text{ rank-1 matrices whose sum equals } X\)
Principal Components Analysis (PCA)

\(X (n \times m)\) contains scores of \(n\) subjects on \(m\) tests

PCA model \(X = AB^T + E\) \(\Leftrightarrow x_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr} + e_{ij}\)

- \(A = XD (n \times R)\) contains \(R\) factors/components as columns (e.g. extraversion, emotional stability, etc)

- \(B (m \times R)\) contains loadings of the tests on the factors
• columns of $\mathbf{X}$ have mean 0 and variance 1
• factors $\mathbf{A}$ have variance 1 and are uncorrelated

Objective: Minimize $\| \mathbf{X} - \mathbf{A} \mathbf{B}^T \|^2$

PCA solution $\mathbf{A} \mathbf{B}^T = \text{truncated SVD of } \mathbf{X}$

$\mathbf{A} \mathbf{B}^T = \mathbf{U}_R \mathbf{S}_R (\mathbf{V}_R)^T$ is a best rank-$R$ approximation of $\mathbf{X}$

$\mathbf{A} = n^{1/2} \mathbf{U}_R = (\mathbf{U} \mathbf{S} \mathbf{V}^T) \mathbf{V}_R (\mathbf{S}_R)^{-1} n^{1/2} = \mathbf{X} \mathbf{V}_R (\mathbf{S}_R)^{-1} n^{1/2}$

$\mathbf{B}^T = n^{-1/2} \mathbf{S}_R (\mathbf{V}_R)^T$
Factors $A$ are linear combinations of the data $X$.

Factors are uncorrelated:
$$A^TA/n = (U_R)^TU_R = I_R$$

SVD:

Principal components (columns of $A$) are ordered.

Explained variances:
$$(s_1)^2/n \geq \ldots \geq (s_R)^2/n$$

Are the $R$ largest eigenvalues of
$$\text{Cov}(X) = X^TX/n = V (S)^2 V^T/n$$
PCA solution is not unique !!

\[ AB^T = (A \ Q)(Q^T B^T) \quad \text{for } Q \quad \text{with} \quad Q^T Q = QQ^T = I_R \]

\[ \Rightarrow \quad \|X - AB^T\|^2 = \|X - AQ Q^T B^T\|^2 \]

\[ \Rightarrow \quad \text{factors are uncorrelated} \quad (A \ Q)^T(A \ Q)/n = Q^T Q = I_R \]

- orthogonal rotation \( Q \) gives a different basis for the factor space spanned by the columns of \( A \)

- if the rotation \( Q \) yields simple structure in the loadings matrix \( (Q^T B^T) \), then the factors are easier to interpret
3-way Data Analysis

3-way data $\Rightarrow$ 3-way array $\mathbf{X} \; n \times m \times p$

examples

- scores of $n$ subjects on $m$ tests at $p$ time points
- scores of $n$ air quality indicators on $m$ time points at $p$ locations
- scores of $n$ judges on $m$ quality indicators for $p$ food products
- fMRI data for $n$ voxels in $m$ scans of $p$ subjects
- chemometrics (spectroscopy, chromatography)
- signal processing (source identification from multi-channel signals)
Mode 1

Mode 2

Mode 3

Handling:

- Horizontal slices: Mode 1 fixed
- Lateral slices: Mode 2 fixed
- Frontal slices: Mode 3 fixed
Mode 1 fibers
Modes 2 and 3 fixed

Mode 2 fibers
Modes 1 and 3 fixed

Mode 3 fibers
Modes 1 and 2 fixed
Rank of 3-way arrays (3-way rank)

mode \( t \) rank of \( \mathbf{X} \) = rank\{ mode \( t \) fibers \} \quad t = 1, 2, 3

analogous to row- and column-rank of matrices

\[ \text{rank}(\mathbf{X}) = \text{smallest } \# \text{ rank-1 arrays whose sum equals } \mathbf{X} \]

\[ \text{rank}(\mathbf{Y}) = 1 \quad \leftrightarrow \quad \mathbf{Y} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \quad \text{for non-zero vectors } \mathbf{a}, \mathbf{b}, \mathbf{c} \]

\[ y_{ijk} = a_i b_j c_k \]
example

\( \mathbf{X} \) is 2×2×2 with frontal slices

\[
\begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
\quad \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
1 & 1 \\
0 & 0
\end{bmatrix}
\]

mode 1 rank = mode 2 rank = mode 3 rank = \( \text{rank}(\mathbf{X}) = 2 \)

mode 2 rank = 1

example

\( \mathbf{X} \) is 2×2×2 with frontal slices

\[
\begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
\quad \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}
\]

mode 1 rank = mode 2 rank = mode 3 rank = 2

\( \text{rank}(\mathbf{X}) = 3 \)
Differences between 2-way and 3-way rank

- mode $t$ ranks of $\mathbf{X}$ may be different
- the rank of $\mathbf{X}$ can be larger than $n$, $m$ and $p$
- the maximal rank of $n \times m \times p$ arrays is usually different from the rank of random $n \times m \times p$ arrays (typical rank)
- for random $n \times m \times p$ arrays there may be more than one rank value which occurs with positive probability
- the rank of $\mathbf{X}$ over the complex field may be different from the rank of $\mathbf{X}$ over the real field
- maximal and typical ranks over the real field are generally not known
Using 2-way PCA on a matrix unfolding of $X$

$p = 5$ frontal slices $\Rightarrow n \times 5m$ matrix

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

problem: PCA yields a $5m \times R$ loadings matrix, which is difficult to interpret

$\Rightarrow$ we need 3-way models to analyze 3-way data
Candecomp / Parafac (CP)

\[ X = a_1 \circ b_1 \circ c_1 + \ldots + a_R \circ b_R \circ c_R + E \]

CP decomposes \( X \) into \( R \) rank-1 arrays and a residual array \( E \) by minimizing \( \|E\|^2 \).
Parafac solution \((A,B,C)\) with

\[
A = [a_1 \ldots a_R] \quad n \times R
\]
\[
B = [b_1 \ldots b_R] \quad m \times R
\]
\[
C = [c_1 \ldots c_R] \quad p \times R
\]

\(X\) is \(n \times m \times p\)

\(A,B,C\) are called component matrices

**CP in matrix form:**

\[
X_k = A \ C_k \ B^T + E_k \quad k = 1, \ldots, p
\]

\(X_k\) and \(E_k\) are the \(k\)-th frontal slices of \(X\) and \(E\)

\(C_k\) is diagonal \(R \times R\) with row \(k\) of \(C\) as diagonal

**CP in element form:**

\[
x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk}
\]
\[ \text{rank}(\mathbf{X}) = \text{smallest } R \text{ for which } \mathbf{X} \text{ has a full CP decomposition} \]

\[ \Rightarrow \text{ CP finds a best rank-} R \text{ approximation of } \mathbf{X} \]

**Differences with matrix SVD**

- CP solution cannot be found analytically – iterative algorithm is needed
- a best rank-\( R \) approximation of \( \mathbf{X} \) may not exist !!
- the \( R \) components are not ordered
- any set of \( R-1 \) components of the CP solution is usually not the best rank-(\( R-1 \)) approximation of \( \mathbf{X} \)
Note: other ways exist to generalize the SVD to three-way arrays, but these have no clear relation to the rank of the array

Comparison with 2-way PCA

- vectors $a_r$ may be seen as factors and $b_r$ and $c_r$ as loadings, but essentially the CP model is symmetric in the 3 modes
- as in 2-way PCA the components are determined by maximizing the explained variance
- usually the CP solution is unique and no rotation is possible without changing the model part
Complex- versus real-valued Parafac

- Real-valued Parafac is mainly used in psychology, chemistry, neuro-imaging
- Complex-valued Parafac is mainly used in signal processing
The Tucker Model

\[ \mathbf{X} = \sum_{r=1}^{R} \sum_{p=1}^{P} \sum_{q=1}^{Q} g_{rpq} \mathbf{a}_r \circ \mathbf{b}_p \circ \mathbf{c}_q + \mathbf{E} \]

Parafac \( \Rightarrow \) \( R = P = Q \) and \( g_{rrr} = 1 \)

\( g_{rpq} = 0 \) if \( (r,p,q) \neq (r,r,r) \)

- factors/loadings \( \mathbf{a}_r, \mathbf{b}_p \) and \( \mathbf{c}_q \) and coefficients \( g_{rpq} \)
- a Tucker solution is not unique
- hybrid models in between Parafac and Tucker with restrictions \( g_{rpq} = 0 \) are used in chemistry
- uniqueness of a hybrid model depends on the pattern of restrictions on \( g_{rpq} \)
$X$ is $n \times m \times p$

Tucker solution $(A,B,C,G)$ with

- $A = [a_1 \ldots a_R]$ $n \times R$
- $B = [b_1 \ldots b_P]$ $m \times P$
- $C = [c_1 \ldots c_Q]$ $p \times Q$

$G$ is $R \times P \times Q$ with elements $g_{rpq} \Rightarrow $ core array $G$
Uniqueness of Parafac solutions

\[
\mathbf{X} = a_1 \circ b_1 \circ c_1 + \ldots + a_R \circ b_R \circ c_R + \mathbf{E}
\]

The fitted model part and residuals do not change if we

- change the order of the summation
- multiply \( a_r \) by \( \lambda_a \), \( b_r \) by \( \lambda_b \) and \( c_r \) by \( \lambda_c \), with \( \lambda_a \lambda_b \lambda_c = 1 \)

If \((\mathbf{A}, \mathbf{B}, \mathbf{C})\) is a Parafac solution, then \((\mathbf{A} \ \Pi \ \Lambda_a, \ \mathbf{B} \ \Pi \ \Lambda_b, \ \mathbf{C} \ \Pi \ \Lambda_c)\) has the same fitted model part, where \(\Pi\) is a permutation matrix and \(\Lambda_a, \Lambda_b, \Lambda_c\) are diagonal matrices such that \(\Lambda_a \Lambda_b \Lambda_c = \mathbf{I}_R\)
If a Parafac solution is unique up to these indeterminacies, then it is called **essentially unique**

To avoid the scaling indeterminacy, the columns of two component matrices can be set to length 1

**Note:** this type of uniqueness does not refer to an essentially unique global minimum of the Parafac objective function
Kruskal’s condition for essential uniqueness

The k-rank of $A$, denoted as $k$, is defined as the maximum number of columns in $A$ that are linearly independent. Specifically:

- $k_A = 0 \implies A$ has an all-zero column
- $k_A = 1 \implies A$ has no all-zero columns, but it has two proportional columns
- $k_A = 2 \implies A$ has no all-zero or proportional columns, but there are 3 linearly dependent columns

Kruskal’s condition for essential uniqueness:

$$2R + 2 \leq k_A + k_B + k_C$$
A Parafac algorithm: Alternating Least Squares

Minimize \[ \sum_{i,j,k} (x_{ijk} - \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr})^2 \]

0. (random) starting values for \((A,B,C)\)
1. find the best \(A\) for fixed \(B\) and \(C\)
2. find the best \(B\) for fixed \(A\) and \(C\)
3. find the best \(C\) for fixed \(A\) and \(B\)
4. if \(\|E_{\text{old}}\|^2 - \|E_{\text{new}}\|^2 > \varepsilon\), then step 1, else STOP

\(\Rightarrow\) each iteration decreases the objective function \(\|E\|^2\)
\(\Rightarrow\) to avoid local minima: try different starting values
Preprocessing for 2-way PCA

$X \ (n \times m \ )$ contains scores of $n$ subjects on $m$ variables

→ center and normalize columns of $X$
  i.e. columns have mean 0 and variance 1

Why center? Scores on variables are usually relative
Centering removes unknown constants

Why normalize? Assures equal influence of each variable
Preprocessing for Parafac and Tucker

\[ X \ (n \times m \times p) \] contains scores of \( n \) subjects on \( m \) variables in \( p \) situations

⇒ same reasons for centering and normalizing, but now there are more possibilities

- center across subjects for all variables and situations combinations \( \Rightarrow x_{ijk} - x_{jk} \)
- center across subjects and variables for all situations \( \Rightarrow x_{ijk} - x_{jk} \)
- normalize within subjects and variables \( \Rightarrow x_{ijk} / \sigma_{ij} \)
To leave the structure of the model intact:

- center across one mode  \( \Rightarrow \)  fiber centering

\[
x_{ijk} - x_{\cdot jk} = \sum_{r=1}^{R} (a_{ir} - a_{\cdot r}) b_{jr} c_{kr} + (e_{ijk} - e_{\cdot jk})
\]

- normalize within one mode  \( \Rightarrow \)  slice normalizing

\[
x_{ijk} / \sigma_j = \sum_{r=1}^{R} a_{ir} (b_{jr} / \sigma_j) c_{kr} + (e_{ijk} / \sigma_j)
\]
Choosing the number $R$ of components

Fit percentage $= \frac{\|X\|^2 - \|E\|^2}{\|X\|^2} \cdot 100$

→ choose $R$ such that adding more components does not significantly increase the fit percentage

In this example, $R = 3$ is a good choice

<table>
<thead>
<tr>
<th>$R$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit %</td>
<td>10.2</td>
<td>16.4</td>
<td>18.7</td>
<td>18.9</td>
<td>19.1</td>
</tr>
</tbody>
</table>
Example of a Parafac analysis

• 5 different kinds of bread are judged on 11 attributes by 8 different judges
• of each bread 2 replicates are judged

\[10 \text{ breads} \times 11 \text{ attributes} \times 8 \text{ judges}\]

• possible scores for each attribute are 0, 1, 2, 3, 4, 5

Parafac analysis $\rightarrow$ “latent variables” $b_r$
loadings for breads in $a_r$
loadings for judges in $c_r$
Preprocessing:  
- centering across breads (mode A)  
- no normalization

Fit percentages for different values of $R$:

<table>
<thead>
<tr>
<th>$R$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit %</td>
<td>35.3</td>
<td>49.2</td>
<td>57.4</td>
<td>62.7</td>
<td>67.2</td>
</tr>
</tbody>
</table>

We choose $R = 2$ (also because visualization of the solution is easy).

We set the columns of $B$ and $C$ to length 1.
Parafac loadings for mode C (judges)
• Kruskal’s condition for uniqueness holds:

$$6 = 2R + 2 \leq k_A + k_B + k_C = 2 + 2 + 2 = 6$$

• the example shows that a 3-way Parafac analysis reveals more structure than a 2-way PCA on a matrix unfolding of the data

$$\begin{array}{ccc}
X_1 & \ldots & X_8 \\
\end{array}$$

10 × 88 matrix

PCA yields an 88 × R loadings matrix

$$R = 2 \Rightarrow 88$$ points for modes B and C together !!
Parafac and Maximum Likelihood

\[ x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \]

with \( e_{ijk} \) uncorrelated and \( \mathcal{N}(0,1) \) distributed

Find \((A, B, C)\) such that the likelihood of \( x_{ijk} \) having this Gaussian distribution is maximized

\[ \Rightarrow \text{equivalent to finding a Parafac solution } (A, B, C) \]

\[ \Rightarrow \text{correlated and/or heteroscedastic } e_{ijk} \text{ is equivalent to Parafac with a weighted least squares objective function} \]
**Imposing constraints on Parafac**

The following constraints can be easily incorporated in Parafac algorithms:

- setting elements of $A$ or $B$ or $C$ to zero
- columns of $A$ or $B$ or $C$ are orthogonal
- columns of $A$ or $B$ or $C$ have zero correlations
- elements of $A$ or $B$ or $C$ are non-negative
- columns of $A$ or $B$ or $C$ must lie in the column space of some “design” matrix
Extension to multi-way Parafac

4-way Parafac: \[ x_{ijkl} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} d_{lr} + e_{ijkl} \]

\[ \mathbf{X} = \mathbf{a}_1 \odot \mathbf{b}_1 \odot \mathbf{c}_1 \odot \mathbf{d}_1 + ... + \mathbf{a}_R \odot \mathbf{b}_R \odot \mathbf{c}_R \odot \mathbf{d}_R + \mathbf{E} \]

- ALS algorithm is equivalent
- extension of Kruskal’s uniqueness condition exists
- same rules for preprocessing
- 4-way Parafac find a best 4-way rank-\( R \) approximation to the 4-way array \( \mathbf{X} \)
Degenerate Parafac solutions

Sometimes, the Parafac algorithm converges slower and slower and the Parafac solution displays a strange pattern.

- Two-factor degeneracy

1. \( a_s \approx \pm a_t \quad b_s \approx \pm b_t \quad c_s \approx \pm c_t \)

2. \( \cos(a_s, a_t) \cdot \cos(b_s, b_t) \cdot \cos(c_s, c_t) \) tends to \(-1\)

3. elements of \( c_s \) and \( c_t \) become arbitrarily large (if \( A \) and \( B \) have length 1 columns)
Two-factor degeneracy

\[ Y^{(s)} = a_s \circ b_s \circ c_s \quad Y^{(t)} = a_t \circ b_t \circ c_t \]

\[ Y^{(s)} + Y^{(t)} \] remains "small" and contributes to a better CP fit
Degenerate Parafac solutions occur when the best rank-$R$ approximation of $\mathbf{X}$ does not exist.

In these cases, the Parafac objective function has no minimum, only an infimum. This explains the slow convergence of the Parafac algorithm.

Degenerate Parafac solutions with 3 or more components involved, may also occur.

Degenerate Parafac solutions do not occur under the restrictions of:

- non-negativity of the elements of $\mathbf{A}$ and $\mathbf{B}$ and $\mathbf{C}$
- orthogonality of the columns of $\mathbf{A}$ and $\mathbf{B}$ and $\mathbf{C}$
References
For each topic, some key references are given.

Singular Value Decomposition
The low-rank approximation of a matrix or multi-way array is called the Eckart-Young problem. For matrices, there always exists a solution (truncated SVD). For three-way arrays, this is not necessarily the case (which explains the occurrence of degenerate Parafac solutions).


Principal Component Analysis
PCA was introduced independently (and somewhat differently) by Pearson (1901) and Hotelling (1933). For a derivation of the optimal A (factors) and B (loadings), see Ten Berge (1993, section 4.2).


Three-way ranks
Existing results and bounds for three-way ranks over the real field are given in the following papers.


**Candecomp/Parafac model**

The model was proposed independently by Harshman (1970) and Carroll & Chang (1970).


Kruskal (1977) proved a useful uniqueness condition for a Parafac solution. A more accessible proof of Kruskal’s condition is given in Stegeman & Sidiropoulos (2005).


Parafac is a special case of the Tucker model, which was introduced by Tucker (1966).


Parafac can be seen as three-way generalizations of the matrix SVD. De Lathauwer et al. (2000) describe a multilinear SVD with orthonormal component matrices and orthogonality restrictions on the core array. However, this SVD has no clear relation to the rank of the array.

**Parafac algorithms**


The Multilinear Engine by Paatero (1999) is a gradient based solver for multilinear models. It includes a script language which makes it easy to incorporate restrictions on the component matrices and to control the algorithm. For more information, contact Pentti Paatero (Helsinki University).


The N-way toolbox for MATLAB by Anderssen & Bro (2000) is freely available and can be used to fit the Parafac model.

**Data compression**
If one mode of the data array is larger then the product of the other two modes (e.g. $n > mp$), then array can be compressed to an $mp \times m \times p$ array. This is explained in Kiers & Harshman (1997).


**Preprocessing**
The do’s and don’t’s of centering and scaling for two-way PCA and three-way Parafac are described in Bro & Smilde (2003).


**Choosing the number of Parafac components**
Bro & Kiers (2003) describe a more sophisticated method to determine a good choice for the number of Parafac components.


**Books on practical 3-way analysis**
Kroonenberg (1983) discusses in detail the Tucker model and provides insight in how to interpret the results from a Tucker analysis. He also includes detailed examples with three-way datasets. Smilde et al. (2004) discuss a wide range of 2-way and 3-way models and consider examples from chemistry.

Parafac analysis of the bread data
The bread data are analyzed in Bro (1998, section 7.2) and can be downloaded from http://www.models.kvl.dk/research/data/Sensory_Bread/index.asp


Parafac and Maximum Likelihood
Vega-Montoto & Wentzell (2003) give ALS algorithms for weighted least squares Parafac (with correlated and/or heteroscedastic errors).


Imposing constraints on Parafac
For Parafac with non-negative component matrices, see Paatero (1997). For the restriction that the column space of a component matrix should lie in the column space of some design matrix, an algorithm can be found in Carroll et al. (1980).

Extension to multi-way Parafac
Sidiropoulos & Bro (2000) have generalized Kruskal’s uniqueness condition to the multi-way Parafac decomposition.


Degenerate Parafac solutions
These type of solutions were first discussed by Harshman & Lundy (1984). In Kruskal et al. (1989) the idea is proposed that degenerate Parafac solutions occur in situations where the Parafac objective function does not have a minimum. Stegeman (2006) proves this claim for $p \times p \times 2$ arrays.