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 ➔ 2-way array = matrix

notation:			
	x	scalar	1×1
	$\mathbf{x}$	column vector	$n \times 1$
	$\mathbf{x}^T$	row vector	$1 \times n$
	$\mathbf{X}$	matrix	$n \times m$

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

rank($\mathbf{X}$) = row-rank($\mathbf{X}$) = column-rank($\mathbf{X}$) $\leq \min(n,m)$

Examples

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

column 1 + column 2 = column 3
columns 1 and 2 are linearly independent
➔ column-rank(**X**) = 2 ➔ rank(**X**) = 2

$\mathbf{X} = \mathbf{a}\,\mathbf{b}^T = \mathbf{a} \circ \mathbf{b}$ ⬅➔ $x_{ij} = a_i\, b_j$

all columns of **X** are scalar multiples ➔ rank(**X**) = 1
if **X** has a nonzero column

Singular Value Decomposition (SVD)

Let $\mathbf{X}$ be an $n \times m$ matrix with $n \geq m$ and rank($\mathbf{X}$) = R

Then the SVD of $\mathbf{X}$ is $\mathbf{X} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$

where $\mathbf{U}$ is $n \times m$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}_m$
$\mathbf{V}$ is $m \times m$ and $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_m$
$\mathbf{S} = \text{diag}\{s_1, \ldots, s_R, 0, \ldots, 0\}$ is $m \times m$

columns of $\mathbf{U}$ are mutually orthogonal and have length 1
columns of $\mathbf{V}$ are mutually orthogonal and have length 1

singular values of $\mathbf{X}$ are $s_1 \geq \ldots \geq s_R > 0 \geq \ldots \geq 0$

rank($\mathbf{X}$) = R = # positive singular values of $\mathbf{X}$

SVD ➔ $\mathbf{X} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + s_R \mathbf{u}_R \mathbf{v}_R^T$

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

rank($\mathbf{u}_j \mathbf{v}_j^T$) = 1 ➔ SVD decomposes $\mathbf{X}$ into R rank-1 matrices

<u>"economy size SVD"</u>

$\mathbf{U}_R$ 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 $R \times R$

➔ 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 $\mathbf{Y}$ (with $p \leq R$) be given by the truncated SVD of $\mathbf{X}$, i.e.

$$\mathbf{Y} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + s_p \mathbf{u}_p \mathbf{v}_p^T = \mathbf{U}_p \mathbf{S}_p (\mathbf{V}_p)^T .$$

Then $\mathbf{Y}$ is a best rank-p approximation of $\mathbf{X}$, i.e.

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

□

➔ SVD gives all best low-rank approximations of $\mathbf{X}$

➔ rank($\mathbf{X}$) = smallest # rank-1 matrices whose sum equals $\mathbf{X}$

Principal Components Analysis (PCA)

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

PCA model $\mathbf{X} = \mathbf{A}\,\mathbf{B}^{T} + \mathbf{E}$ ⟷ $x_{ij} = \sum_{r=1}^{R} a_{ir}\, b_{jr} + e_{ij}$

- **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 **X** have mean 0 and variance 1
- factors **A** have variance 1 and are uncorrelated

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

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

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

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

$\mathbf{B}^{\mathrm{T}} = n^{-1/2}\,\mathbf{S}_R\,(\mathbf{V}_R)^{\mathrm{T}}$

➔ factors **A** are linear combinations of the data **X**

➔ factors are uncorrelated $\mathbf{A}^T\mathbf{A}/n = (\mathbf{U}_R)^T\mathbf{U}_R = \mathbf{I}_R$

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

$\mathrm{Cov}(\mathbf{X}) = \mathbf{X}^T\mathbf{X}/n = \mathbf{V}\,(\mathbf{S})^2\,\mathbf{V}^T/n$

PCA solution is not unique !!

$\mathbf{A}\,\mathbf{B}^{\mathsf{T}} = (\mathbf{A}\,\mathbf{Q})(\mathbf{Q}^{\mathsf{T}}\mathbf{B}^{\mathsf{T}})$ for $\mathbf{Q}$ with $\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathsf{T}} = \mathbf{I}_R$

➔ $\left\|\mathbf{X} - \mathbf{A}\,\mathbf{B}^{T}\right\|^2 = \left\|\mathbf{X} - \mathbf{A}\mathbf{Q}\,\mathbf{Q}^{T}\mathbf{B}^{T}\right\|^2$

➔ factors are uncorrelated $(\mathbf{A}\,\mathbf{Q})^{\mathsf{T}}(\mathbf{A}\,\mathbf{Q})/n = \mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I}_R$

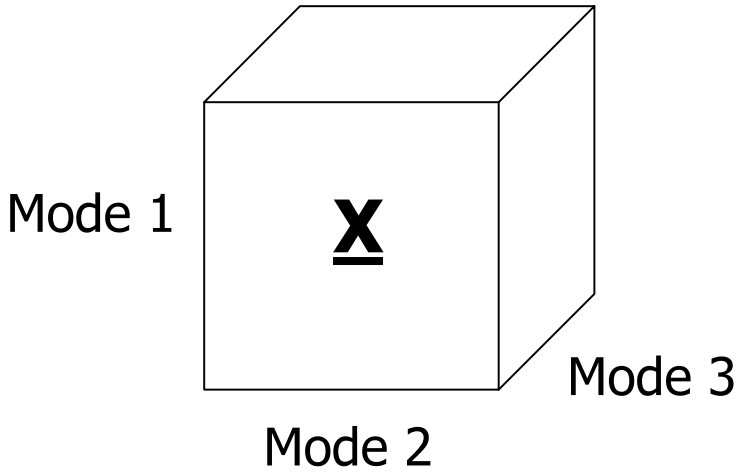
- orthogonal rotation $\mathbf{Q}$ gives a different basis for the factor space spanned by the columns of $\mathbf{A}$
- if the rotation $\mathbf{Q}$ yields simple structure in the loadings matrix ($\mathbf{Q}^{\mathsf{T}}\mathbf{B}^{\mathsf{T}}$), then the factors are easier to interpret

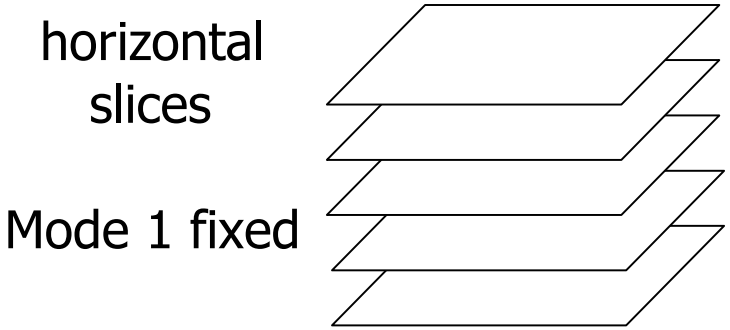
3-way Data Analysis

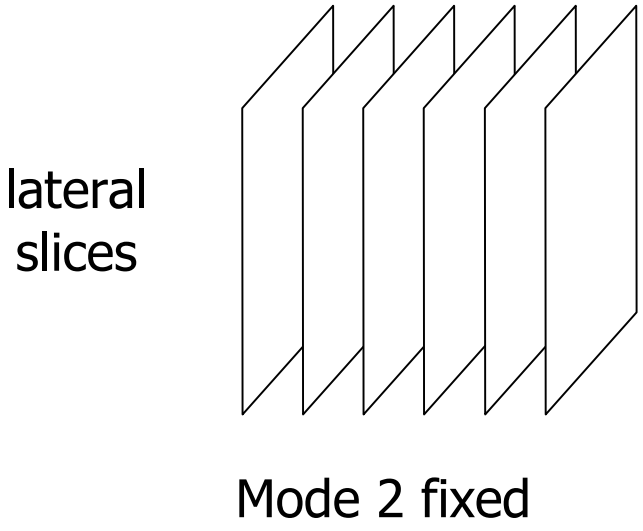
3-way data ➔ 3-way array $\underline{\mathbf{X}}$ $n \times m \times p$

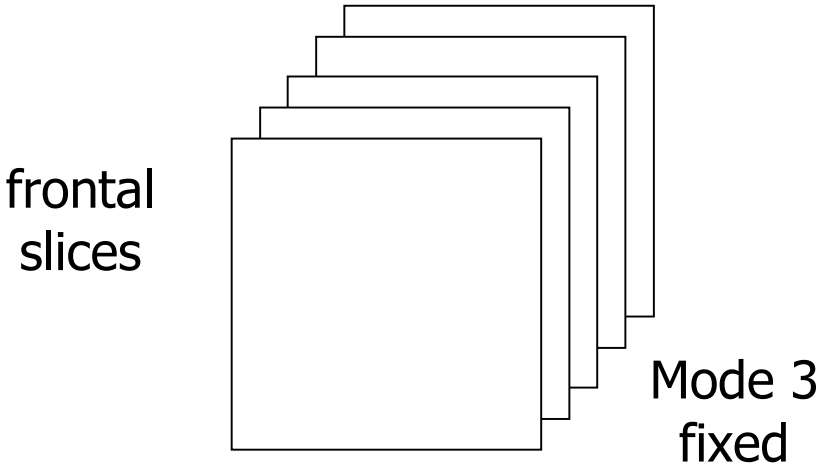
examples

- 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
X
Mode 3
Mode 2

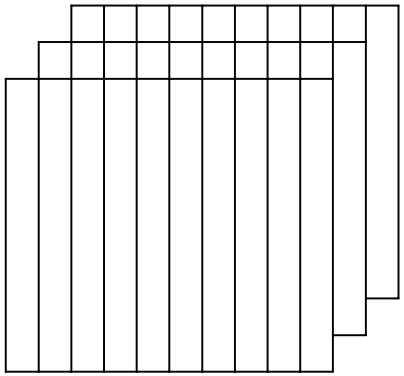

horizontal
slices
Mode 1 fixed


lateral
slices
Mode 2 fixed


frontal
slices
Mode 3
fixed

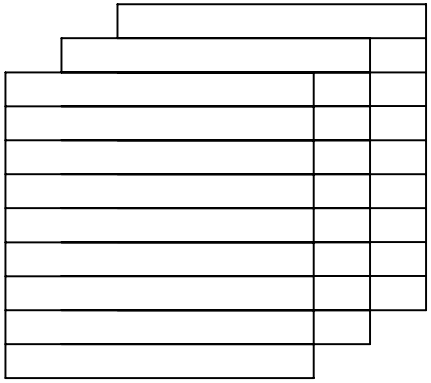
Mode 1 fibers

Modes 2 and 3 fixed



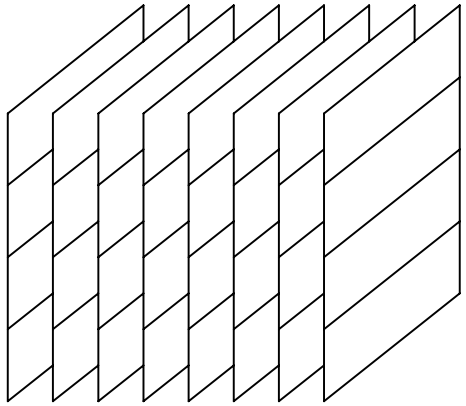
Mode 2 fibers

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 $\underline{\mathbf{X}}$ = rank{ mode t fibers } $t = 1, 2, 3$

analogous to row- and column-rank of matrices

rank($\underline{\mathbf{X}}$) = smallest # rank-1 arrays whose sum equals $\underline{\mathbf{X}}$

rank($\underline{\mathbf{Y}}$) = 1 ⟷ $\underline{\mathbf{Y}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ for non-zero vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$

$$y_{ijk} = a_i \, b_j \, c_k$$

example

$\underline{\mathbf{X}}$ is 2×2×2 with frontal slices $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$

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

example

$\underline{\mathbf{X}}$ is 2×2×2 with frontal slices $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

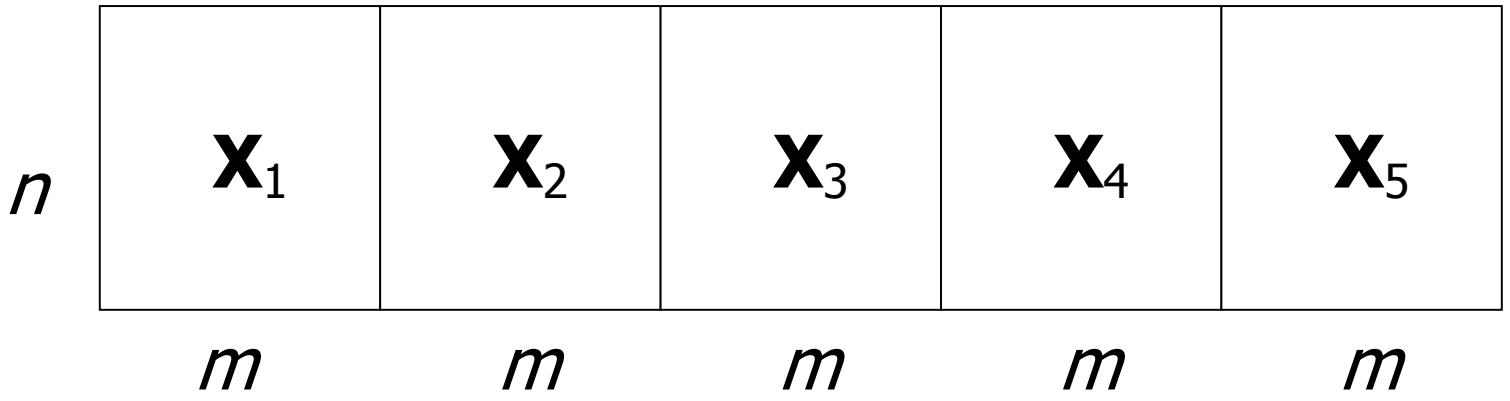
mode 1 rank = mode 2 rank = mode 3 rank = 2
rank($\underline{\mathbf{X}}$) = 3

Differences between 2-way and 3-way rank

- mode t ranks of $\underline{\mathbf{X}}$ may be different
- the rank of $\underline{\mathbf{X}}$ 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 $\underline{\mathbf{X}}$ over the complex field may be different from the rank of $\underline{\mathbf{X}}$ 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 ➔ $n \times 5m$ matrix



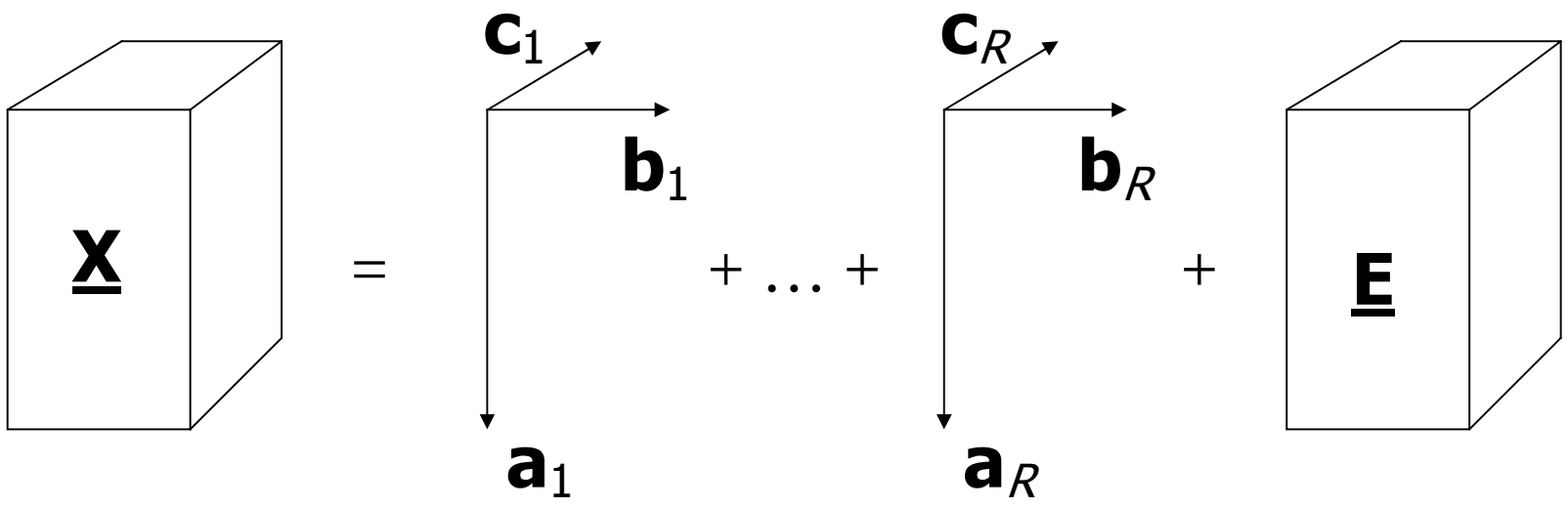
problem: PCA yields a $5m \times R$ loadings matrix, which is difficult to interpret

➔ we need 3-way models to analyze 3-way data

Candecomp / Parafac (CP)

$$\underline{\mathbf{X}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R + \underline{\mathbf{E}}$$

CP decomposes $\underline{\mathbf{X}}$ into R rank-1 arrays and a residual array $\underline{\mathbf{E}}$ by minimizing $\left\|\underline{\mathbf{E}}\right\|^2$



Parafac solution (**A**,**B**,**C**) with $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R]$ $n \times R$

$\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_R]$ $m \times R$

$\underline{\mathbf{X}}$ is $n \times m \times p$ $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_R]$ $p \times R$

A,**B**,**C** are called component matrices

CP in matrix form: $\mathbf{X}_k = \mathbf{A}\ \mathbf{C}_k\ \mathbf{B}^T + \mathbf{E}_k \quad k = 1, \ldots, p$

$\mathbf{X}_k$ and $\mathbf{E}_k$ are the k-th frontal slices of $\underline{\mathbf{X}}$ and $\underline{\mathbf{E}}$

$\mathbf{C}_k$ is diagonal $R \times R$ with row k of **C** as diagonal

CP in element form: $x_{ijk} = \sum_{r=1}^{R} a_{ir}\, b_{jr}\, c_{kr} + e_{ijk}$

rank($\underline{\mathbf{X}}$) = smallest R for which $\underline{\mathbf{X}}$ has a full CP decomposition

➔ CP finds a best rank-R approximation of $\underline{\mathbf{X}}$

Differences with matrix SVD

- CP solution cannot be found analytically – iterative algorithm is needed
- a best rank-R approximation of $\underline{\mathbf{X}}$ 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 $\underline{\mathbf{X}}$

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 $\mathbf{a}_r$ may be seen as factors and $\mathbf{b}_r$ and $\mathbf{c}_r$ 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

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

Parafac ➔ $R = P = Q$ 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 restictions $g_{rpq} = 0$ are used in chemistry
- uniqueness of a hybrid model depends on the pattern of restrictions on g_{rpq}

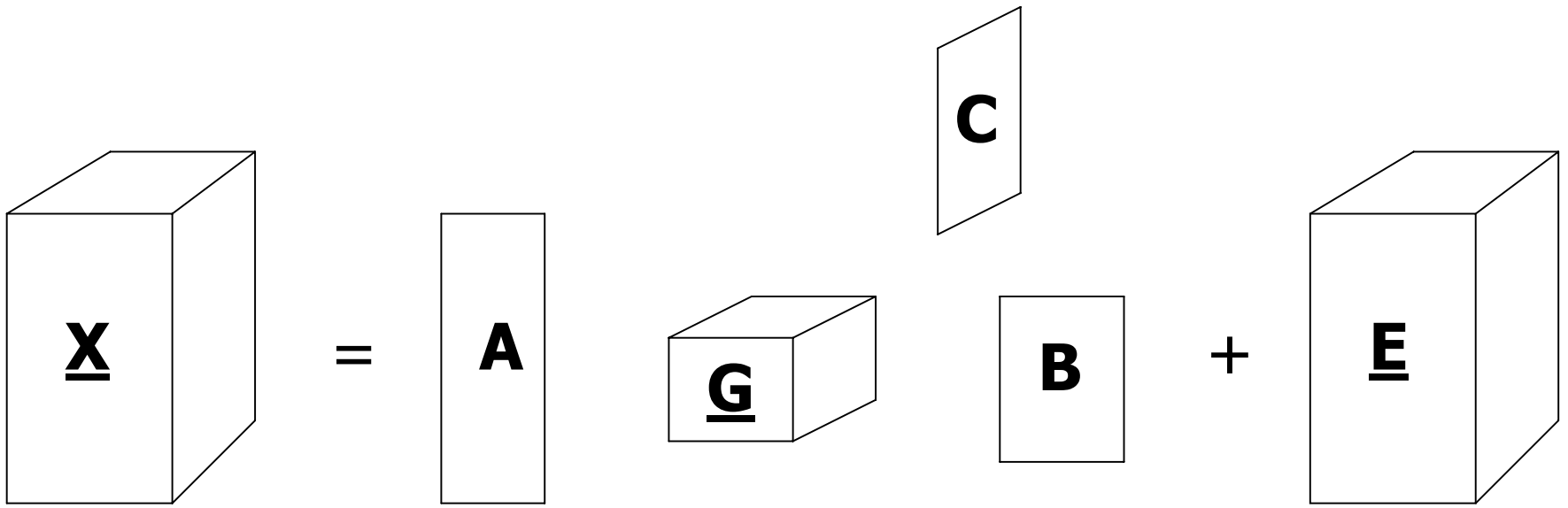
$\underline{\mathbf{X}}$ is $n \times m \times p$

Tucker solution $(\mathbf{A},\mathbf{B},\mathbf{C},\underline{\mathbf{G}})$ with $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R]$ $n \times R$

$\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_P]$ $m \times P$

$\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_Q]$ $p \times Q$

$\underline{\mathbf{G}}$ is $R \times P \times Q$ with elements g_{rpq} ➔ core array $\underline{\mathbf{G}}$



Uniqueness of Parafac solutions

$$\underline{\mathbf{X}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R + \underline{\mathbf{E}}$$

The fitted model part and residuals do not change if we

- change the order of the summation
- multiply $\mathbf{a}_r$ by λ_a, $\mathbf{b}_r$ by λ_b and $\mathbf{c}_r$ by λ_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 $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{\Lambda}_a, \mathbf{\Lambda}_b, \mathbf{\Lambda}_c$ are

diagonal matrices such that $\mathbf{\Lambda}_a\,\mathbf{\Lambda}_b\,\mathbf{\Lambda}_c = \mathbf{I}_R$

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

k-rank of **A** = max number k such that every set of k columns of **A** is linearly independent

$k_{\mathbf{A}} = 0$ ➔ **A** has an all-zero column

$k_{\mathbf{A}} = 1$ ➔ **A** has no all-zero columns, but it has two proportional columns

$k_{\mathbf{A}} = 2$ ➔ **A** has no all-zero or proportional columns, but there are 3 linearly dependent columns

Kruskal's condition for essential uniqueness:

$$2R + 2 \leq k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}}$$

A Parafac algorithm: Alternating Least Squares

$$\text{Minimize} \quad \sum_{i,j,k} \left(x_{ijk} - \sum_{r=1}^{R} a_{ir}\, b_{jr}\, c_{kr} \right)^2$$

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 $\left\|\underline{\mathbf{E}}_{\text{old}}\right\|^2 - \left\|\underline{\mathbf{E}}_{\text{new}}\right\|^2 > \varepsilon$, then step 1, else STOP

➔ each iteration decreases the objective function $\left\|\underline{\mathbf{E}}\right\|^2$

➔ to avoid local minima: try different starting values

Preprocessing for 2-way PCA

$\mathbf{X}$ $(n \times m)$ 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

$\underline{\mathbf{X}}$ $(n \times m \times p)$ 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 ➔ $x_{ijk} - x_{\bullet jk}$
- center across subjects and variables for all situations ➔ $x_{ijk} - x_{\bullet\bullet k}$
- normalize within subjects and variables ➔ x_{ijk} / σ_{ij}

To leave the structure of the model intact:

- center across one mode ➔ fiber centering

$$x_{ijk} - x_{\bullet jk} = \sum_{r=1}^{R} (a_{ir} - a_{\bullet r})\, b_{jr}\, c_{kr} + (e_{ijk} - e_{\bullet jk})$$

- normalize within one mode ➔ 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

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

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

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

R	1	2	3	4	5
Fit %	10.2	16.4	18.7	18.9	19.1

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 breads × 11 attributes × 8 judges

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

Parafac analysis ➔ "latent variables" $\mathbf{b}_r$
loadings for breads in $\mathbf{a}_r$
loadings for judges in $\mathbf{c}_r$

Preprocessing:

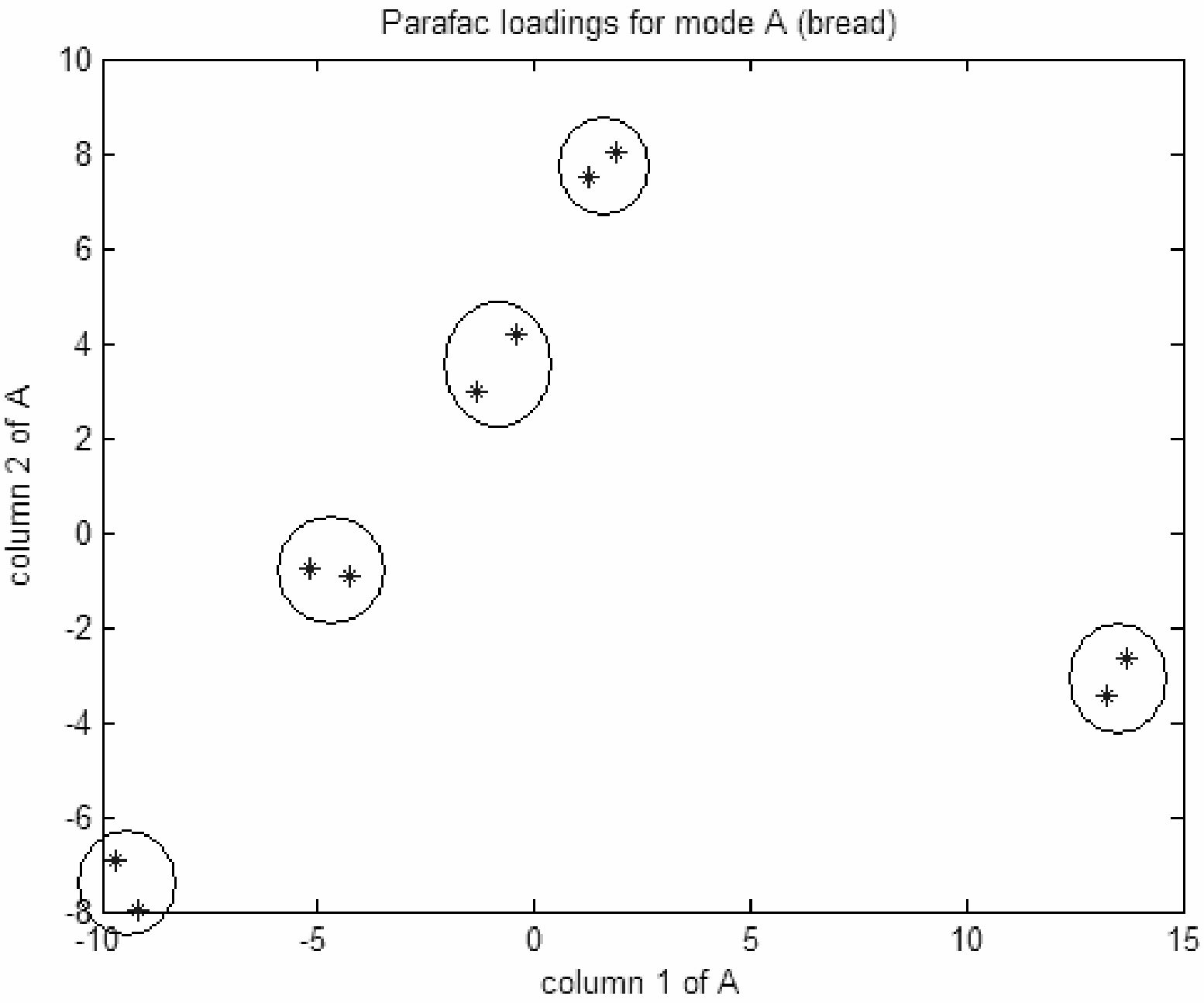
- centering across breads (mode A)
- no normalization

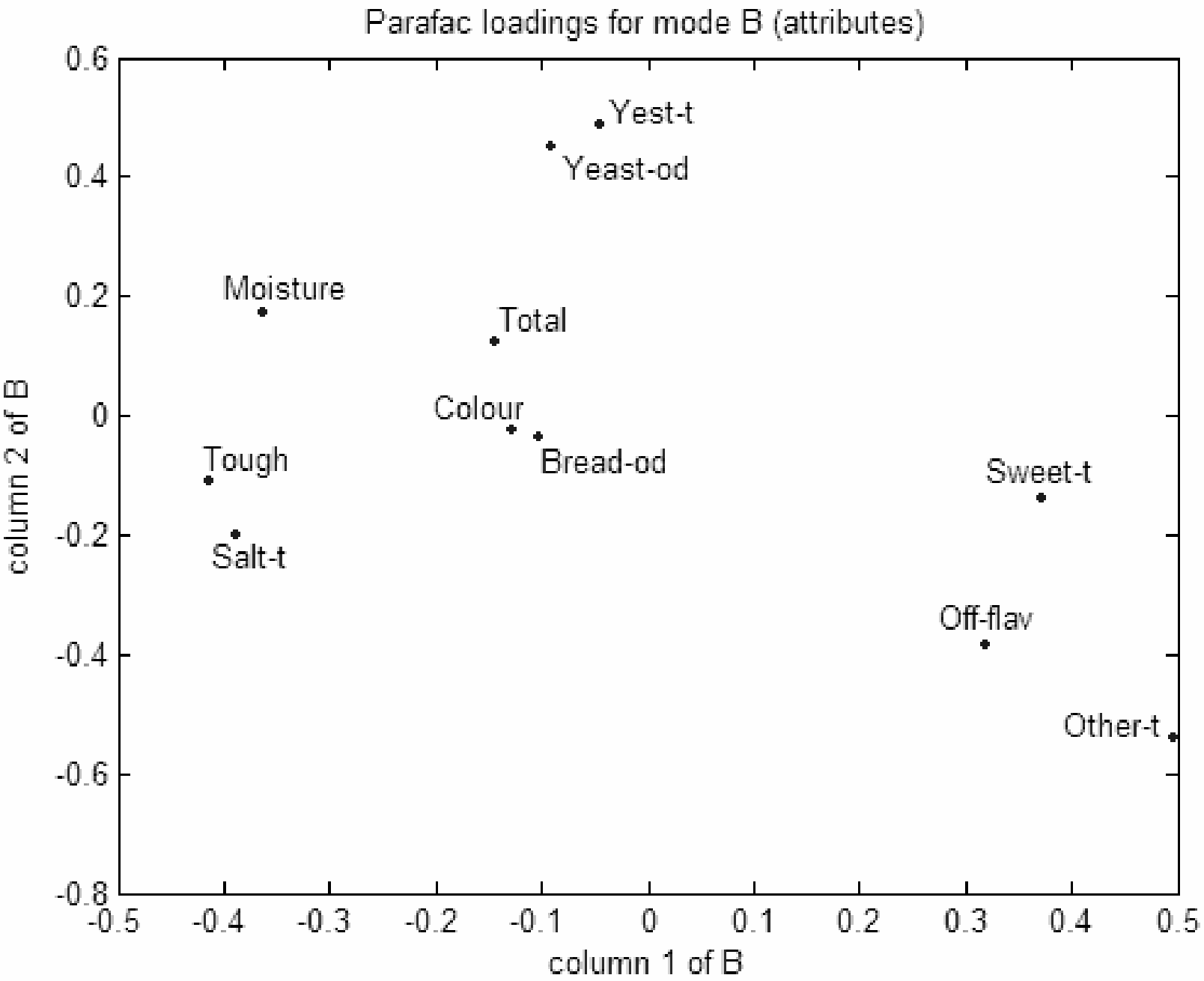
Fit percentages for different values of R :

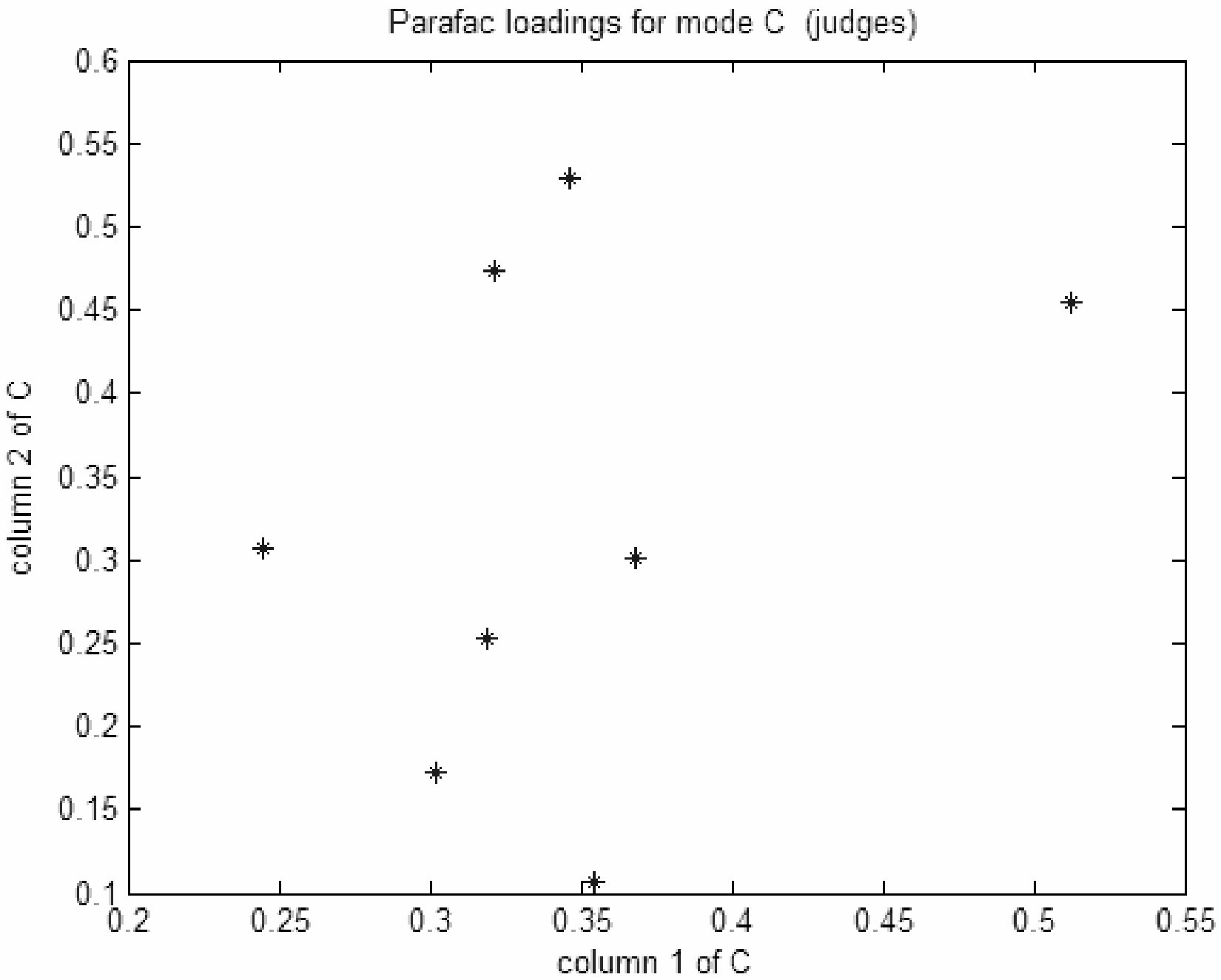
R	1	2	3	4	5
Fit %	35.3	49.2	57.4	62.7	67.2

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 A (bread)
column 2 of A
column 1 of A
10
8
6
4
2
0
-2
-4
-6
-8
-10
-5
0
5
10
15


Parafac loadings for mode B (attributes)
Yest-t
Yeast-od
Moisture
Total
Colour
Bread-od
Tough
Salt-t
Sweet-t
Off-flav
Other-t
column 2 of B
column 1 of B
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5


Parafac loadings for mode C (judges)
column 2 of C
column 1 of C
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55

- Kruskal's condition for uniqueness holds:

$$6 = 2R + 2 \leq k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} = 2 + 2 + 2 = 6$$

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

$\mathbf{X}_1$	…	$\mathbf{X}_8$

10×88 matrix

PCA yields an $88 \times R$ loadings matrix

$R = 2$ ➔ 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 N(0,1) distributed

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

➔ equivalent to finding a Parafac solution (**A**,**B**,**C**)

➔ correlated and/or heteroscedastic e_{ijk} 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: $x_{ijkl} = \sum_{r=1}^{R} a_{ir}\, b_{jr}\, c_{kr}\, d_{lr} + e_{ijkl}$

$$\underline{\mathbf{X}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 \circ \mathbf{d}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R \circ \mathbf{d}_R + \underline{\mathbf{E}}$$

- 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 $\underline{\mathbf{X}}$

Degenerate Parafac solutions

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

- Two-factor degeneracy

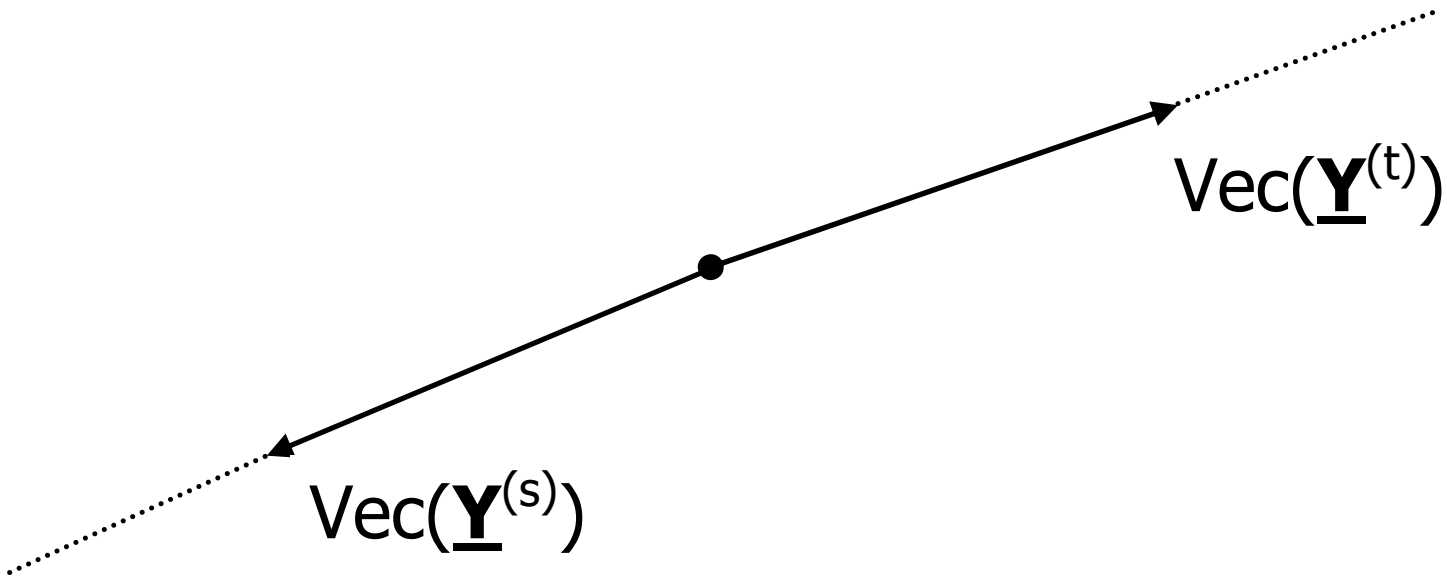
  1. $\mathbf{a}_s \approx \pm\, \mathbf{a}_t \qquad \mathbf{b}_s \approx \pm\, \mathbf{b}_t \qquad \mathbf{c}_s \approx \pm\, \mathbf{c}_t$

  2. $\cos(\mathbf{a}_s, \mathbf{a}_t) \cdot \cos(\mathbf{b}_s, \mathbf{b}_t) \cdot \cos(\mathbf{c}_s, \mathbf{c}_t)$ tends to -1

  3. elements of $\mathbf{c}_s$ and $\mathbf{c}_t$ become arbitrarily large (if **A** and **B** have length 1 columns)

Two-factor degeneracy

$\underline{\mathbf{Y}}^{(s)} = \mathbf{a}_s \circ \mathbf{b}_s \circ \mathbf{c}_s$ $\qquad$ $\underline{\mathbf{Y}}^{(t)} = \mathbf{a}_t \circ \mathbf{b}_t \circ \mathbf{c}_t$



$\underline{\mathbf{Y}}^{(s)} + \underline{\mathbf{Y}}^{(t)}$ remains "small" and contributes to a better CP fit

Degenerate Parafac solutions occur when the best rank-*R* approximation of $\underline{\mathbf{X}}$ 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 **A** and **B** and **C**
- orthogonality of the columns of **A** and **B** and **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).

- Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika, **1**, 211-218.

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).

- Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* **2** (6), 559–572.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology,* **24***,* 417-441.
- Ten Berge, J.M.F. (1993). *Least Squares Optimization in Multivariate Analysis*. DSWO Press, Leiden, The Netherlands.

Three-way ranks

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

- Kruskal, J.B. (1989). Rank, decomposition and uniqueness for 3-way and N-way arrays. In: *Multiway Data Analysis*, Coppi, R. & Bolasco, S. (editors), Elsevier, Amsterdam, 7-18.

- Ten Berge, J.M.F. & Kiers, H.A.L. (1999). Simplicity of core arrays in three-way principal component analysis and the typical rank of $p \times q \times 2$ arrays. *Linear Algebra and its Applications*, **294**, 169-179.
- Ten Berge, J.M.F. (2000). The typical rank of tall three-way arrays. *Psychometrika*, **65**, 525-532.
- Ten Berge, J.M.F. (2004). Simplicity and typical rank of three-way arrays, with applications to Tucker-3 analysis with simple cores. *Journal of Chemometrics*, **18**, 17-21.

Candecomp/Parafac model

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

- Carroll, J.D. & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckart-Young decomposition. *Psychometrika*, **35**, 283-319.
- Harshman, R.A. (1970). Foundations of the Parafac procedure: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, **16**, 1-84.

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).

- Kruskal, J.B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, **18**, 95-138.
- Stegeman, A. & Sidiropoulos, N.D. (2005). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. Submitted.

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

- Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, **31**, 279-311.

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.

- De Lathauwer, L., De Moor, B. & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, **21**, 1253-1278.

Parafac algorithms

The Alternating Least Squares algorithm is treated nicely in Ten Berge (1993, p.58).

- Ten Berge, J.M.F. (1993). *Least Squares Optimization in Multivariate Analysis*. DSWO Press, Leiden, The Netherlands.

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).

- Paatero, P. (1999). The Multilinear Engine – a table-driven least squares program for solving multilinear problems, including the *n*-way Parallel Factor Analysis model. *Journal of Computational Statistics and Graphical Statistics*, **8**, 854-888.

A comparison of the performance of Parafac algorithms, is made in Tomasi & Bro (2006).

- Tomasi, G. & Bro, R. (2006). A comparison of algorithms for fitting the Parafac model. *Computational Statistics & Data Analysis*, **50**, 1700-1734.

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

- Andersson, C.A. & Bro, R. (2000). The N-way Toolbox for MATLAB. *Chemometrics & Intelligent Laboratory Systems*, **52**, 1-4. http://www.models.kvl.dk/source/nwaytoolbox/

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).

- Kiers, H.A.L. & Harshman, R.A. (1997). Relating two proposed methods for speedup of algorithms for fitting two-and three-way principal component and related multilinear models. *Chemometrics & Intelligent Laboratory Systems*, **36**, 31-40.

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).

- Bro, R. & Smilde, A.K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, **17**, 16-33.

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.

- Bro, R. & Kiers, H.A.L. (2003). A new efficient method for determining the number of components in Parafac models. *Journal of Chemometrics*, **17**, 274-286

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.

- Kroonenberg, P.M. (1983). *Three-Mode Principal Component Analysis*. DSWO Press, Leiden, The Netherlands.
- Smilde, A., Bro, R. & Geladi, P. (2004). Multi-Way Analysis, Applications in the Chemical Sciences. Wiley, Chichester, UK.

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

- Bro, R. (1998). *Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications*. Ph.D. thesis, University of Amsterdam (NL) & Royal Veterinary and Agricultural University (DK). Available online at: http://www.models.kvl.dk/users/rasmus/brothesis.pdf

Parafac and Maximum Likelihood

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

- Vega-Montoto, L. & Wentzell, P.D. (2003). Maximum likelihood parallel factor analysis (MLPARAFAC). *Journal of Chemometrics*, **17**, 237-253.

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).

- Carroll, J.D., Pruzansky, S. & Kruskal, J.B. (1980). Candelinc: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, **45**, 3-24.
- Paatero, P. (1997). A weighted non-negative least squares algorithm for three-way 'Parafac' factor analysis. *Chemometrics & Intelligent Laboratory Systems*, **38**, 223-242.

Extension to multi-way Parafac

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

- Sidiropoulos, N.D. & Bro, R. (2000). On the uniqueness of multilinear decomposition of *N*-way arrays. *Journal of Chemometrics*, **14**, 229-239.

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.

- Harshman, R.A. & Lundy, M.E. (1984). Data preprocessing and the extended Parafac model. In: *Research Methods for Multimode Data Analysis*, Law, H.G., Snyder Jr, C.W., Hattie, J.A. & McDonald, R.P. (editors), Praeger, New York, 216-284.
- Kruskal, J.B., Harshman, R.A. & Lundy, M.E. (1989). How 3-MFA data can cause degenerate Parafac solutions, among other relationships. In: *Multiway Data Analysis*, Coppi, R. & Bolasco, S. (editors), Elsevier, Amsterdam, 115-122.
- Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p+1$ or higher. *Psychometrika*, to appear.