

Candecom / Parafac

How to avoid diverging components

Alwin Stegeman

Heymans Institute
University of Groningen
The Netherlands

a.w.stegeman@rug.nl
<http://www.gmw.rug.nl/~stegeman>

Singular Value Decomposition (SVD)

\mathbf{X} $n \times m$ matrix with $\text{rank}(\mathbf{X}) = R$

The SVD of **\mathbf{X}** is **$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^T$**

where **\mathbf{U}** is $n \times R$ and **$\mathbf{U}^T \mathbf{U} = \mathbf{I}_R$**

\mathbf{V} is $m \times R$ and **$\mathbf{V}^T \mathbf{V} = \mathbf{I}_R$**

$\mathbf{S} = \text{diag}\{s_1, \dots, s_R\}$ with $s_1 \geq \dots \geq s_R > 0$

SVD **$\rightarrow \mathbf{X} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + \dots + s_R \mathbf{u}_R \mathbf{v}_R^T$**

\rightarrow SVD decomposes **\mathbf{X}** into R rank-1 matrices

Theorem: Eckart & Young (1936)

Let \mathbf{Y} be given by the truncated SVD of \mathbf{X} , i.e.

$$\mathbf{Y} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + \dots + s_p \mathbf{u}_p \mathbf{v}_p^T \quad \text{with } p \leq R.$$

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

$$\|\mathbf{X} - \mathbf{Y}\|^2 = \sum_{i,j} (x_{ij} - y_{ij})^2 \text{ is minimal over } \mathbf{Y} \text{ of rank } p \quad \square$$

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

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

Principal Component Analysis (PCA)

\mathbf{X} ($n \times m$) contains n observations of m variables

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

- \mathbf{A} ($n \times R$) contains R uncorrelated latent factors or principal components as columns
- \mathbf{B} ($m \times R$) contains loadings of the variables on the factors

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

$\text{rank}(\mathbf{A} \mathbf{B}^T) = R \rightarrow \mathbf{A} \mathbf{B}^T =$ truncated SVD of \mathbf{X}

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

Rotation of the factors/components

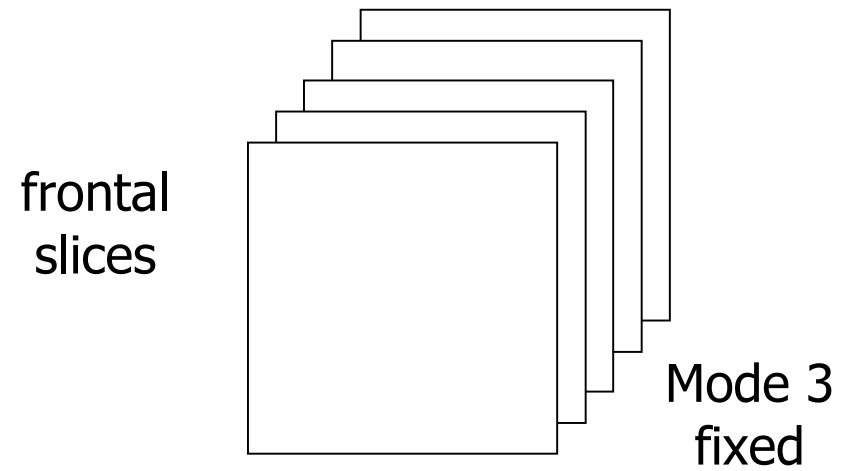
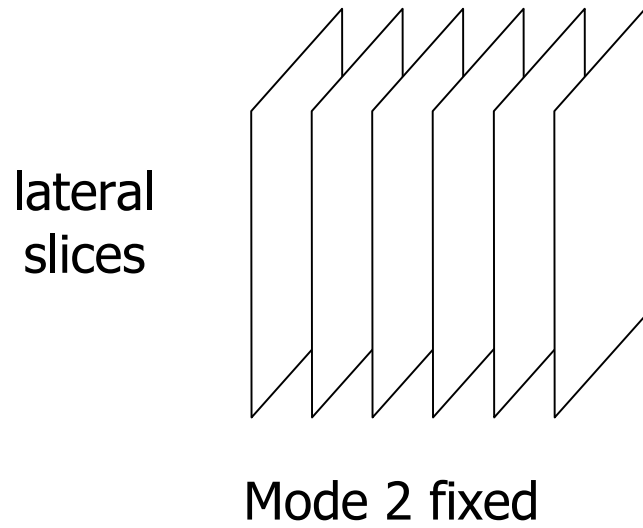
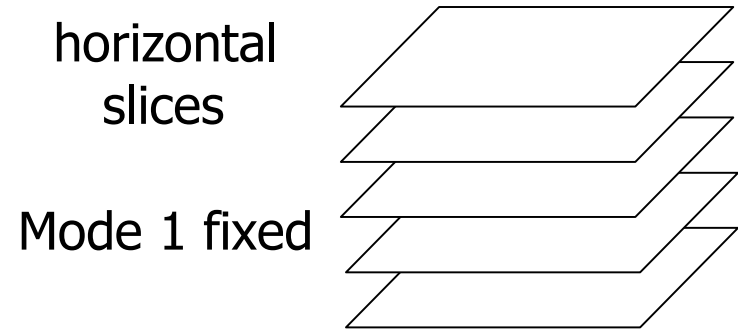
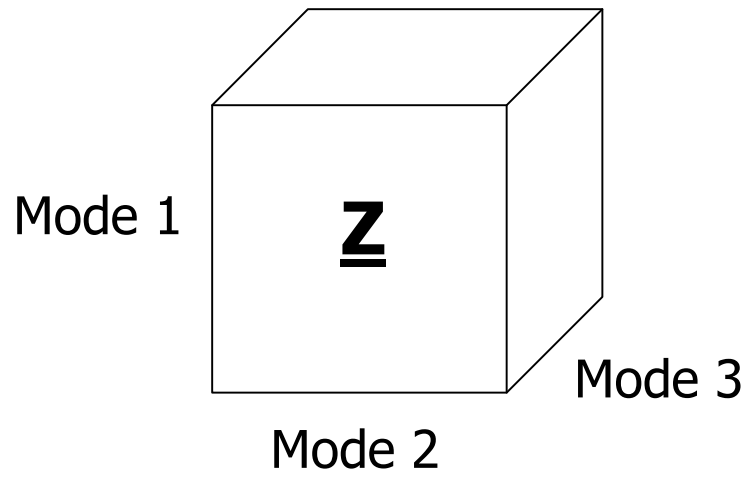
$$\mathbf{A} \mathbf{B}^T = (\mathbf{A} \mathbf{Q})(\mathbf{B} \mathbf{Q})^T \quad \text{for } \mathbf{Q} \text{ with } \mathbf{Q}^T \mathbf{Q} = \mathbf{Q} \mathbf{Q}^T = \mathbf{I}_R$$

- rotation = change of basis for the factor space
- look for \mathbf{Q} with simple structure in loadings $(\mathbf{B} \mathbf{Q})^T$
 \rightarrow factors are easier to interpret

3-way Data Analysis

3-way data → 3-way array $\underline{\mathbf{Z}}$ $n \times m \times p$

$\underline{\mathbf{Z}}$	n	m	p
scores	subjects	tests	time points
scores	air quality indicators	time points	locations
scores	judges	quality indicators	food products
fMRI data	voxels	scans	subjects
spectroscopy	samples	retention times	wavelength
signals	antenna	sensor	time point



Rank of 3-way arrays (3-way rank)

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

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

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

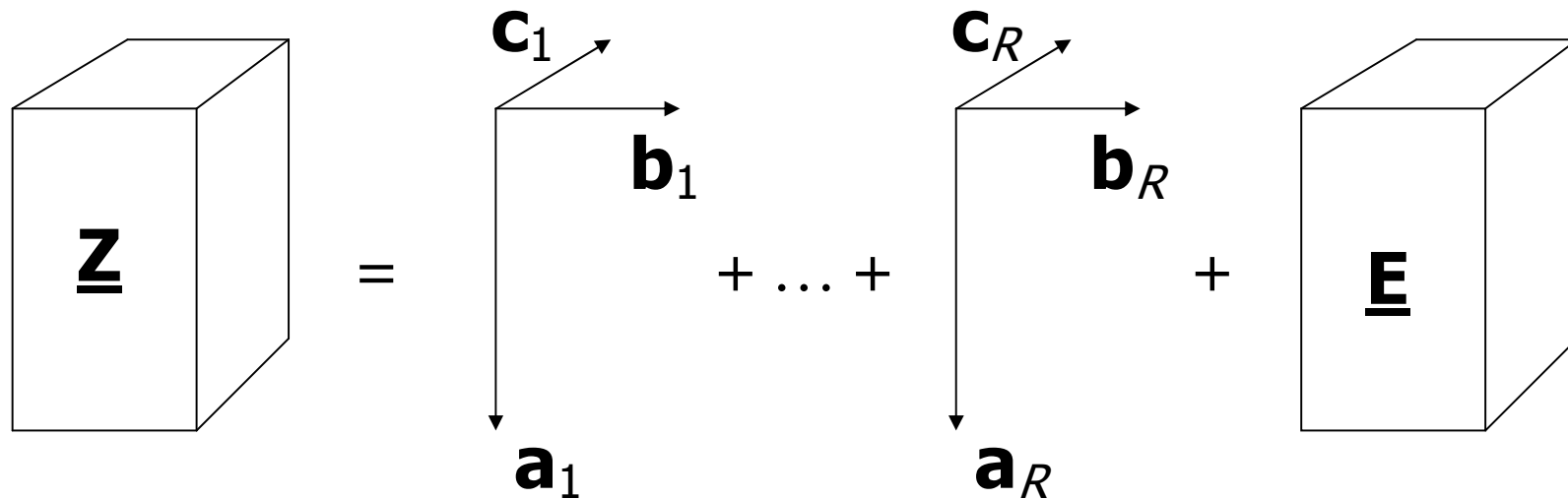
$\underline{\mathbf{Y}} \ n \times m \times p$ may have $\text{rank}(\underline{\mathbf{Y}}) > \max(n, m, p)$

$\underline{\mathbf{Y}}$ random may have $P(\text{rank}(\underline{\mathbf{Y}}) = k+1) > 0$
and $P(\text{rank}(\underline{\mathbf{Y}}) = k) > 0$

Candecomp / Parafac decomposition

$$\underline{\mathbf{Z}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \dots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R + \underline{\mathbf{E}}$$

Parafac decomposes $\underline{\mathbf{Z}}$ into R rank-1 arrays and a residual array $\underline{\mathbf{E}}$ by minimizing $\|\underline{\mathbf{E}}\|^2$



Parafac solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ with

$$\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_R] \quad n \times R$$

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

$$\mathbf{C} = [\mathbf{c}_1 \dots \mathbf{c}_R] \quad p \times R$$

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

$\mathbf{A}, \mathbf{B}, \mathbf{C}$ are called component matrices

Parafac in matrix form: $\mathbf{Z}_k = \mathbf{A} \mathbf{C}_k \mathbf{B}^T + \mathbf{E}_k \quad k = 1, \dots, p$

\mathbf{Z}_k and \mathbf{E}_k are frontal slices k of $\underline{\mathbf{Z}}$ and $\underline{\mathbf{E}}$
 \mathbf{C}_k is diagonal with row k of \mathbf{C} as diagonal

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

$\text{rank}(\underline{\mathbf{Z}}) = \text{smallest } R \text{ for which } \underline{\mathbf{Z}} \text{ has a full Parafac decomposition}$

→ Parafac finds a best rank- R approximation of $\underline{\mathbf{Z}}$

Hitchcock (1927ab): tensor decomposition

Carroll & Chang (1970): Candecomp

Harshman (1970): Parafac

Parafac applications: psychometrics, chemometrics, neuro-imaging, signal processing, algebraic complexity theory, data mining, computer vision...

	Parafac	SVD / PCA
computation	iterative algorithm	closed form
ordered components	no	yes
maximize total explained variance	yes	yes
rotational uniqueness	under mild conditions	no (PCA)
existence for $R < \text{rank}$	not guaranteed	yes

Uniqueness of Parafac solutions

$$\underline{\mathbf{Z}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \dots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R + \underline{\mathbf{E}}$$

Residuals $\underline{\mathbf{E}}$ do not change when :

- changing the order of the summation
- rescaling \mathbf{a}_r by λ_a , \mathbf{b}_r by λ_b and \mathbf{c}_r by λ_c , with $\lambda_a \lambda_b \lambda_c = 1$

$(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and $(\mathbf{A} \mathbf{\Pi} \mathbf{\Lambda}_a, \mathbf{B} \mathbf{\Pi} \mathbf{\Lambda}_b, \mathbf{C} \mathbf{\Pi} \mathbf{\Lambda}_c)$ are equivalent

$\mathbf{\Pi}$ permutation matrix

$\mathbf{\Lambda}_a, \mathbf{\Lambda}_b, \mathbf{\Lambda}_c$ diagonal with $\mathbf{\Lambda}_a \mathbf{\Lambda}_b \mathbf{\Lambda}_c = \mathbf{I}_R$

→ “essential uniqueness”

Kruskal's condition for essential uniqueness

k -rank of \mathbf{A} = max number k such that every set of k columns of \mathbf{A} is linearly independent

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

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

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

Kruskal's (1977) condition: $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2R + 2$

Stegeman & Sidiropoulos (2007)

Diverging Parafac components ("degeneracy")

Sometimes, the Parafac algorithm converges slower and slower and the updates of the Parafac solution display a strange pattern...

- Two diverging components

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

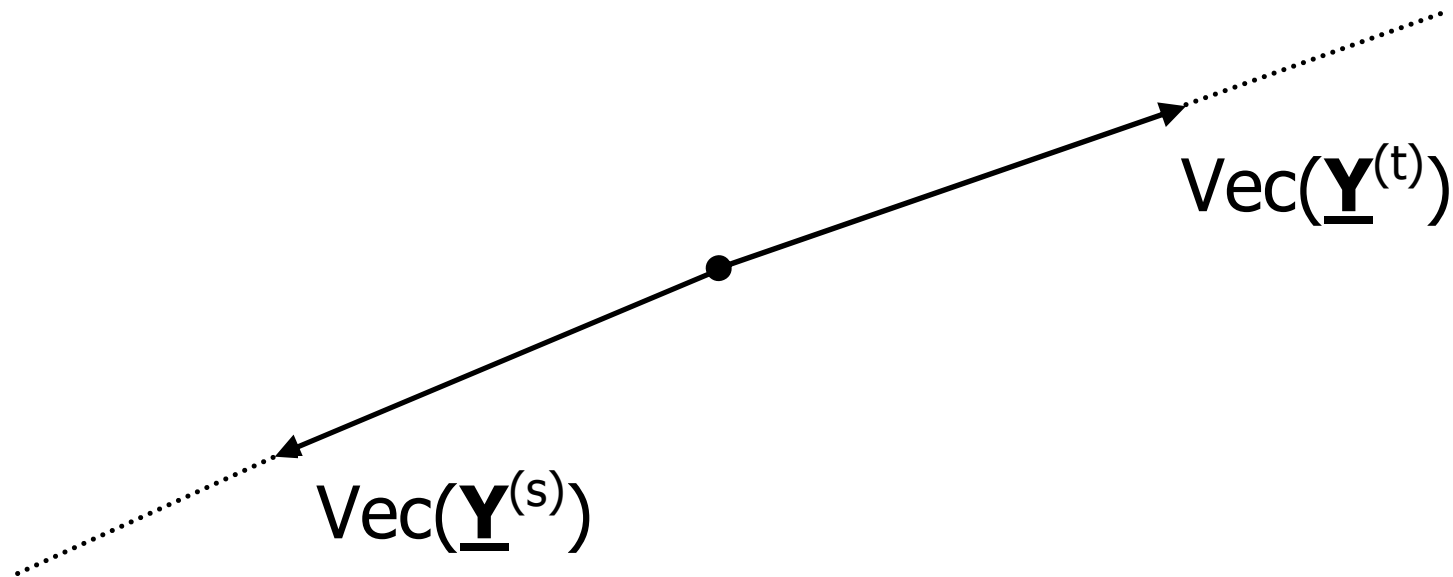
2. $\cos(\mathbf{a}_s, \mathbf{a}_t) \cdot \cos(\mathbf{b}_s, \mathbf{b}_t) \cdot \cos(\mathbf{c}_s, \mathbf{c}_t)$ tends to -1

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

Two diverging components

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

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



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

- Three diverging components

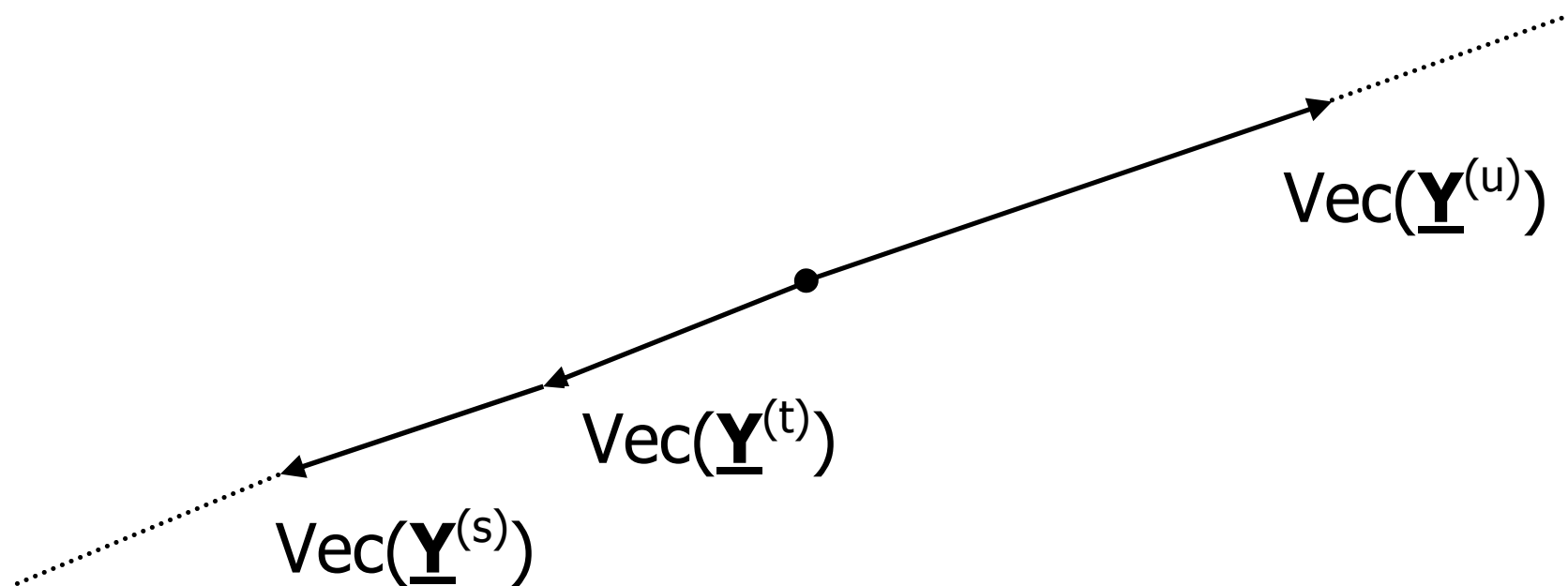
1. $\mathbf{a}_s \approx \pm \mathbf{a}_t \approx \pm \mathbf{a}_u$ $\mathbf{b}_s \approx \pm \mathbf{b}_t \approx \pm \mathbf{b}_u$

2. \mathbf{c}_s , \mathbf{c}_t and \mathbf{c}_u are nearly proportional

3. \mathbf{c}_s , \mathbf{c}_t and \mathbf{c}_u become arbitrarily large
(if \mathbf{A} and \mathbf{B} have length 1 columns)

4. $\mathbf{c}_s \pm \mathbf{c}_t \pm \mathbf{c}_u$ remains “small”

Three diverging components



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

Example for $3 \times 3 \times 2$ with $R = 3$

$$\mathbf{A} = \begin{bmatrix} 0.48 & 0.46 & -0.47 \\ -0.66 & -0.65 & 0.66 \\ -0.57 & -0.60 & 0.58 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0.72 & -0.69 & 0.71 \\ 0.61 & -0.65 & 0.63 \\ 0.33 & -0.31 & 0.32 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 653 & -625 & 1278 \\ 2162 & -2239 & 4398 \end{bmatrix}$$

factor 1 \approx factor 2 \approx $-2 \cdot$ factor 3

random $\underline{\mathbf{z}}$ $p \times q \times 2$	$\text{rank}(\underline{\mathbf{z}})$	R	diverging components?
$p = q$	$p+1$	$R = p$	always
$p = q$	$p+1$	$R < p$	sometimes
$p = q$	p	$R < p$	sometimes
$p > q$	$\min(p, 2q)$	$(p, 2q) > R > q$	no
$p > q$	$\min(p, 2q)$	$R = q$	sometimes
$p > q$	$\min(p, 2q)$	$R < q$	sometimes

Stegeman (2008a)

Diverging Parafac components occur when the best rank- R approximation of $\underline{\mathbf{Z}}$ does not exist

- Parafac objective function has no minimum, only an infimum
- slow convergence of the Parafac algorithm

Conjecture: Kruskal, Harshman, Lundy (1989)

Proof: Krijnen, Dijkstra, Stegeman (2008)

The set of 3-way arrays of rank at most R is not closed

De Silva & Lim (2006)

How to avoid diverging components

De Silva & Lim (2006): Consider the closure of the set of arrays with at most rank R

What is needed?

→ Complete characterization of boundary points

What is known?

$p \times p \times 2$	rank $p + 1$	$R = p$	Stegeman (2006)
$p \times q \times 2$	typical rank	all R	Stegeman (2008a)
several cases of $p \times q \times 3$			Stegeman (2008b)

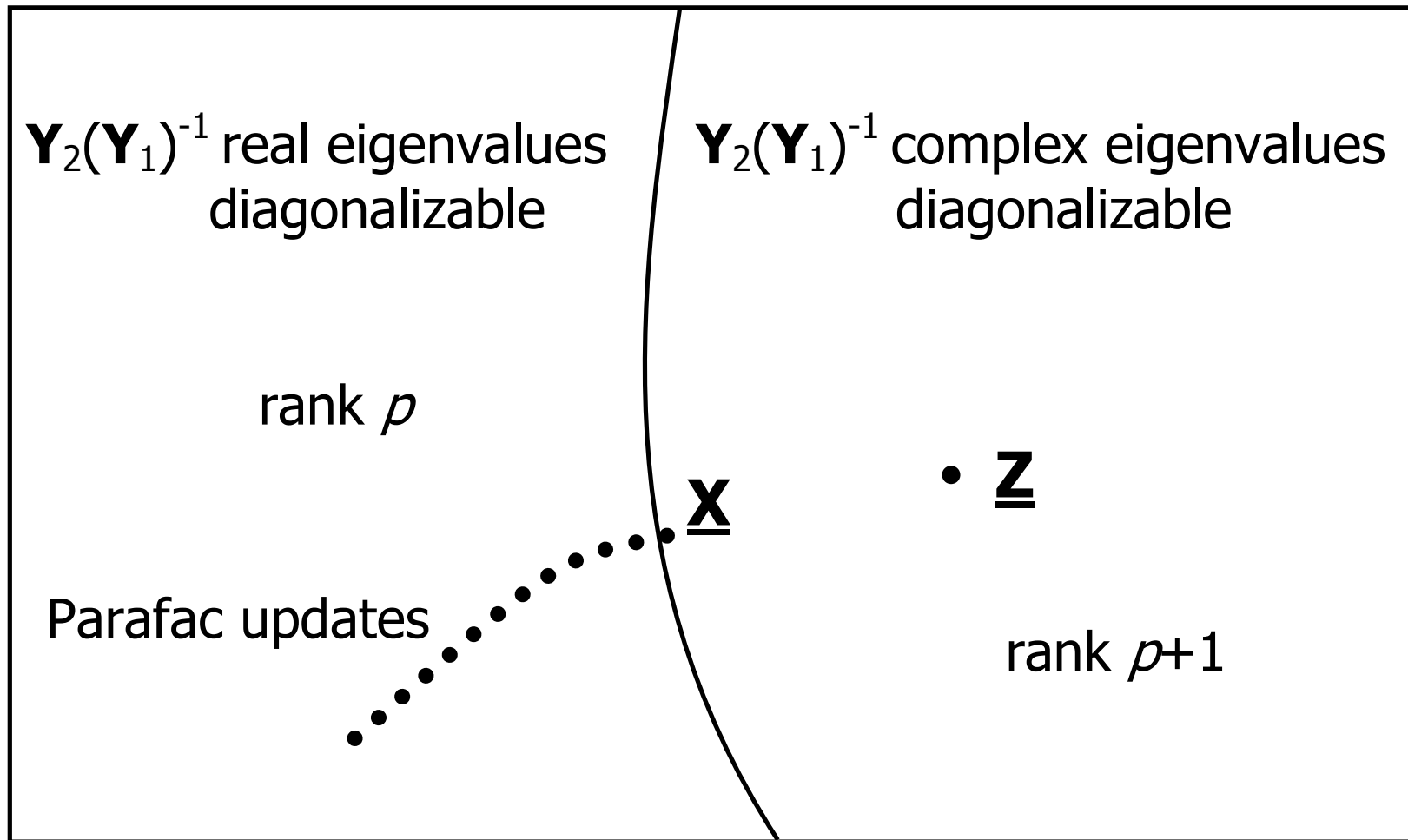
Consider $p \times p \times 2$ arrays with $R = p$

$\mathbf{Y}_2(\mathbf{Y}_1)^{-1}$	eigenvalues all real	some complex eigenvalues
diagonalizable	$\text{rank}(\underline{\mathbf{Y}}) = p$ positive volume	$\text{rank}(\underline{\mathbf{Y}}) \geq p+1$ positive volume
not diag.	$\text{rank}(\underline{\mathbf{Y}}) \geq p+1$ zero volume	$\text{rank}(\underline{\mathbf{Y}}) \geq p+1$ zero volume

Ten Berge (1991), Ja' Ja' (1979)

typical rank = $\{p, p+1\}$ (Ten Berge & Kiers, 1999)

Space of $p \times p \times 2$ arrays



Boundary array **X**

$\mathbf{X}_2(\mathbf{X}_1)^{-1}$ has all eigenvalues real, but not all distinct

$\mathbf{X}_2(\mathbf{X}_1)^{-1}$ diagonalizable \rightarrow rank p (type 1)

$\mathbf{X}_2(\mathbf{X}_1)^{-1}$ not diagonalizable \rightarrow rank $\geq p+1$ (type 2)

dimensionality (type 1) < dimensionality (type 2)

within type 2, $\mathbf{X}_2(\mathbf{X}_1)^{-1}$ having only eigenvalues with geometric multiplicity 1 has highest dimensionality

For random **Z** only these boundary arrays **X** are found

$\text{rank}(\underline{\mathbf{Y}}) = p \rightarrow$ eigendecomposition $\mathbf{Y}_2(\mathbf{Y}_1)^{-1} = \mathbf{K} \mathbf{\Lambda} \mathbf{K}^{-1}$

A rank- p decomposition of $\underline{\mathbf{Y}}$ is $\mathbf{Y}_1 = \mathbf{K} \mathbf{I}_p (\mathbf{K}^{-1} \mathbf{Y}_1)$

$$\mathbf{Y}_2 = \mathbf{K} \mathbf{\Lambda} (\mathbf{K}^{-1} \mathbf{Y}_1)$$

$$\mathbf{A} = \mathbf{K} \quad \mathbf{B} = (\mathbf{K}^{-1} \mathbf{Y}_1)^\top \quad \mathbf{C} = \begin{bmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_p \end{bmatrix}$$

Kruskal: $k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} = 2p + 2 \rightarrow$ uniqueness

When running Parafac algorithm :

$$\underline{\mathbf{Y}} \dashrightarrow \underline{\mathbf{X}} \quad \text{and} \quad \mathbf{Y}_2(\mathbf{Y}_1)^{-1} = \mathbf{K} \mathbf{\Lambda} \mathbf{K}^{-1} \dashrightarrow \mathbf{X}_2(\mathbf{X}_1)^{-1}$$

For random $\underline{\mathbf{Z}}$:

$\mathbf{X}_2(\mathbf{X}_1)^{-1}$ has (at least) one eigenvalue with algebraic multiplicity larger than 1 and geometric multiplicity 1

Parafac algorithm terminates with $\underline{\mathbf{Y}}$ close to $\underline{\mathbf{X}}$:

- $\mathbf{A} = \mathbf{K}$ has eigenvectors nearly identical up to sign
- $\mathbf{B} = (\mathbf{K}^{-1} \mathbf{Y}_1)^T$ has the same columns nearly linearly dependent (proportional) and large magnitudes
- $\mathbf{C} = \begin{bmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_p \end{bmatrix}$ has these columns nearly identical

How to avoid diverging components

$p \times q \times 2$ arrays: Generalized Schur Decomposition

$$\text{GSD} \quad \mathbf{Z}_k = \mathbf{Q}_a \mathbf{R}_k \mathbf{Q}_b^T + \mathbf{E}_k \quad k = 1, 2$$

\mathbf{Q}_a is $p \times R$ with $\mathbf{Q}_a^T \mathbf{Q}_a = \mathbf{I}_R$

\mathbf{Q}_b is $q \times R$ with $\mathbf{Q}_b^T \mathbf{Q}_b = \mathbf{I}_R$

\mathbf{R}_k is upper triangular $k = 1, 2$

- GSD optimal solution $\underline{\mathbf{X}}$ always exists
- GSD algorithm gives $\underline{\mathbf{X}}$ and its GSD $(\mathbf{Q}_a \mathbf{Q}_b \mathbf{R}_1 \mathbf{R}_2)$
- the set of $p \times q \times 2$ arrays with a full GSD is the closure of the set of $p \times q \times 2$ arrays with rank at most R

→ optimal GSD solution $\underline{\mathbf{X}}$ is limit point of Parafac updates $\underline{\mathbf{Y}}$

After computing $\underline{\mathbf{X}}$ and its GSD $(\mathbf{Q}_a \mathbf{Q}_b \mathbf{R}_1 \mathbf{R}_2)$:

→ Jordan form of $\mathbf{R}_2(\mathbf{R}_1)^{-1}$ gives a sparse (Tucker3) representation of the limit point of the diverging Parafac components

→ nondiverging Parafac components can be obtained from the GSD (without computing the Jordan form)

Stegeman & De Lathauwer (2008)

GSD for $p \times p \times 2$ arrays and $R=p$

$$\text{rank}(\underline{\mathbf{Y}}) = p \quad \rightarrow \quad \mathbf{Y}_k = \mathbf{A} \mathbf{C}_k \mathbf{B}^T \quad k = 1, 2$$

$$\mathbf{A} = \mathbf{Q}_a \mathbf{R}_a \quad \text{QR-decomposition}$$

$$\mathbf{B} = \mathbf{Q}_b \mathbf{L}_b \quad \text{QL-decomposition}$$

$$\underline{\mathbf{Y}} \text{ has full GSD} \quad \mathbf{Y}_k = \mathbf{Q}_a (\mathbf{R}_a \mathbf{C}_k \mathbf{L}_b^T) \mathbf{Q}_b^T \quad k = 1, 2$$

Boundary array $\underline{\mathbf{X}}$ with $\mathbf{X}_2(\mathbf{X}_1)^{-1}$ having all eigenvalues real, not all distinct, and not diagonalizable

Jordan form $\mathbf{X}_2(\mathbf{X}_1)^{-1} = \mathbf{P} \mathbf{J} \mathbf{P}^{-1}$

$$\mathbf{P} = \mathbf{Q}_a \mathbf{R}_a$$

QR-decomposition

$$\mathbf{X}_1^T \mathbf{Q}_a = \mathbf{Q}_b \mathbf{L}_b$$

QL-decomposition

$\underline{\mathbf{X}}$ has full GSD

$$\mathbf{X}_1 = \mathbf{Q}_a \mathbf{L}_b^T \mathbf{Q}_b^T$$

$$\mathbf{X}_2 = \mathbf{Q}_a (\mathbf{R}_a \mathbf{J} (\mathbf{R}_a)^{-1} \mathbf{L}_b^T) \mathbf{Q}_b^T$$

References

- 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.
- De Silva, V. & Lim, L-H. (2006). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SCCM Technical Report*, **06-06**, Stanford.
- Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211-218.
- 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.
- Hitchcock, F.L. (1927a). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, **6**, 164-189.
- Hitchcock, F.L. (1927b). Multiple invariants and generalized rank of a p -way matrix or tensor. *Journal of Mathematics and Physics*, **7**, 39-70.
- Ja' Ja', J. (1979). Optimal evaluation of pairs of bilinear forms. *SIAM Journal on Computing*, **8**, 443-462.

- Krijnen, W.P., Dijkstra, T.K. & Stegeman, A. (2008). On the non-existence of optimal solutions and the occurrence of “degeneracy” in the Candecomp/Parafac model. *Psychometrika*, to appear.
- 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.
- Kruskal, J.B., Harshman, R.A. & Lundy, M.E. (1989). How 3-MFA data can cause degenerate Parafac solutions, among other relationships. In: *Multway Data Analysis*, Coppi R. & Bolasco, S. (editors), North-Holland, 115-121.
- Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p+1$ or higher. *Psychometrika*, **71**, 483-501.
- Stegeman, A. & Sidiropoulos, N.D. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and its Applications*, **420**, 540-552.
- Stegeman, A. & De Lathauwer, L. (2008). A method to avoid “degeneracy” in the Candecomp/Parafac model for generic $I \times J \times 2$ arrays. Submitted to *SIAM Journal on Matrix Analysis and Applications*.

- Stegeman, A. (2008a). Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Applications*, to appear.
- Stegeman, A. (2008b). Degeneracy in Candecomp/Parafac and Indscal explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, to appear.
- Ten Berge, J.M.F. (1991). Kruskal's polynomial for $2 \times 2 \times 2$ arrays and a generalization to $2 \times n \times n$ arrays. *Psychometrika*, **56**, 631-636.
- 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.