Netherlands Organisation for Scientific Research

university of groningen

Nonexistence of best low-rank approximations for real-valued three-way arrays

Alwin Stegeman

a.w.stegeman@rug.nl

www.gmw.rug.nl/~stegeman

Summarizing Data in Simple Patterns

Information Technology ➔ collection of huge data sets,
often multi-way data $z(i,j,k,\ldots)$

Approximation: Multi-way data $\approx$ simple patterns

- data interpretation (psychometrics, neuro-imaging, data mining)
- separation of chemical compounds (chemometrics)
- separation of mixed signals (signal processing)
- faster calculations (algebraic complexity theory, scientific computing)

Simple structure = rank 1

2-way array = matrix $\mathbf{Z}$ $(I\times J)$ with entries $z(i,j)$

rank 1: $\mathbf{Z} = \mathbf{a}\,\mathbf{b}^T = \mathbf{a} \circ \mathbf{b}$ $\Leftrightarrow$ $z(i,j) = a(i) \cdot b(j)$

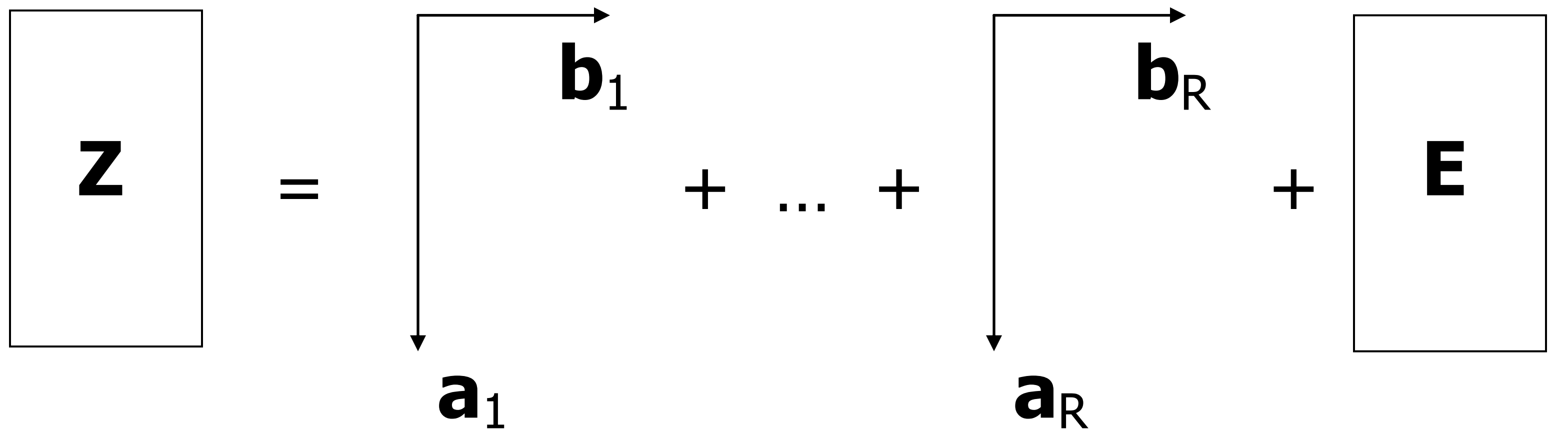
$\text{rank}(\mathbf{Z}) = \min\{R : \mathbf{Z} = \mathbf{a}_1 \circ \mathbf{b}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R\}$

3-way array $\underline{\mathbf{Z}}$ $(I\times J\times K)$ with entries $z(i,j,k)$

rank 1: $\underline{\mathbf{Z}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ $\Leftrightarrow$ $z(i,j,k) = a(i) \cdot b(j) \cdot c(k)$

$\text{rank}(\underline{\mathbf{Z}}) = \min\{R : \underline{\mathbf{Z}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R\}$

2-way (PCA) decomposition

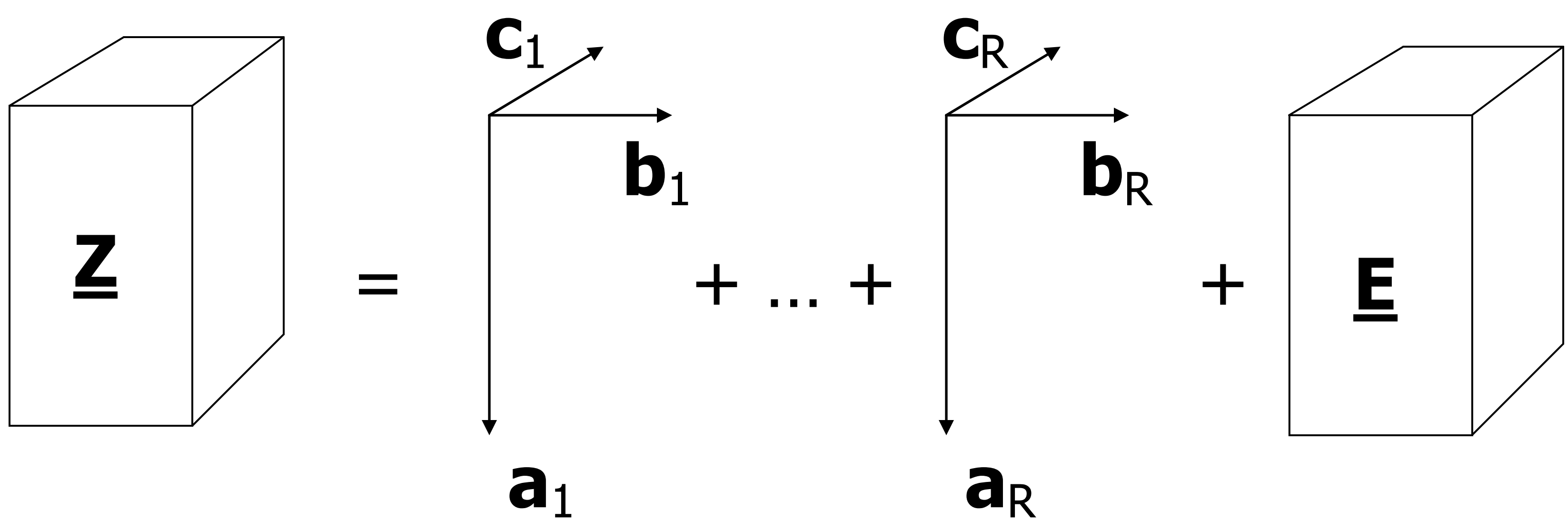


$$\mathbf{Z} = \mathbf{a}_1 \circ \mathbf{b}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R + \mathbf{E}$$

$$= \mathbf{A}\,\mathbf{B}^T + \mathbf{E} \qquad \text{with} \qquad \begin{aligned} \mathbf{A} &= [\mathbf{a}_1 \ldots \mathbf{a}_R] \\ \mathbf{B} &= [\mathbf{b}_1 \ldots \mathbf{b}_R] \end{aligned}$$

Goal: Find (**A**,**B**) that minimize ssq(**E**)

3-way Canonical Polyadic Decomposition (CPD)



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

<u>Goal</u>: Find $(\mathbf{A},\mathbf{B},\mathbf{C})$ that minimize $\text{ssq}(\underline{\mathbf{E}})$

with $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_R]$

	3-way CPD	2-way PCA
computation	iterative algorithm	SVD
best rank-R approximation	yes	yes
rotational uniqueness	under mild conditions *	no
existence for R < rank(data)	not guaranteed	yes

* Kruskal (1977) and many more since 2000

3-way CPD as Optimization Problem

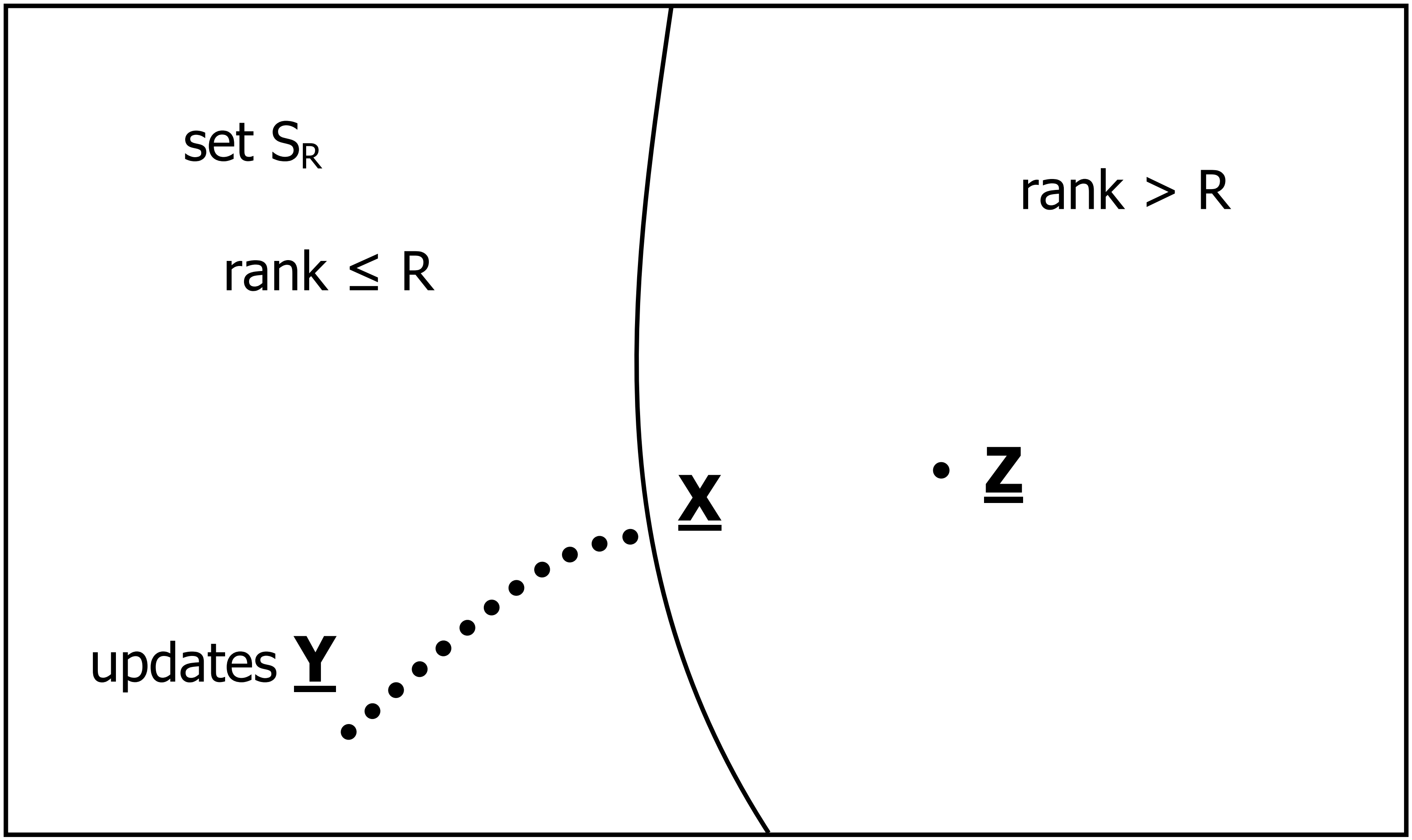
Minimize	$\mathrm{ssq}(\underline{\underline{\mathbf{Z}}} - \underline{\mathbf{Y}})$
over	$S_R = \{ \underline{\mathbf{Y}} : \ \mathrm{rank}(\underline{\mathbf{Y}}) \leq R \}$

➔ if $\underline{\underline{\mathbf{Z}}} \notin S_R$, then an optimal solution $\underline{\mathbf{X}}$ (if it exists) will be a boundary point of S_R

But : the set S_R is not closed for $R \geq 2$

Bini et al. (1979), Paatero (2000), Lim (2004)
De Silva & Lim (2008)

A misleading picture



Suppose $\underline{\mathbf{Y}} = (\mathbf{A},\mathbf{B},\mathbf{C}) \longrightarrow$ optimal $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}} \notin S_R$

Then some rank-1 terms $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$ converge to

linear dependency and **infinite norm**

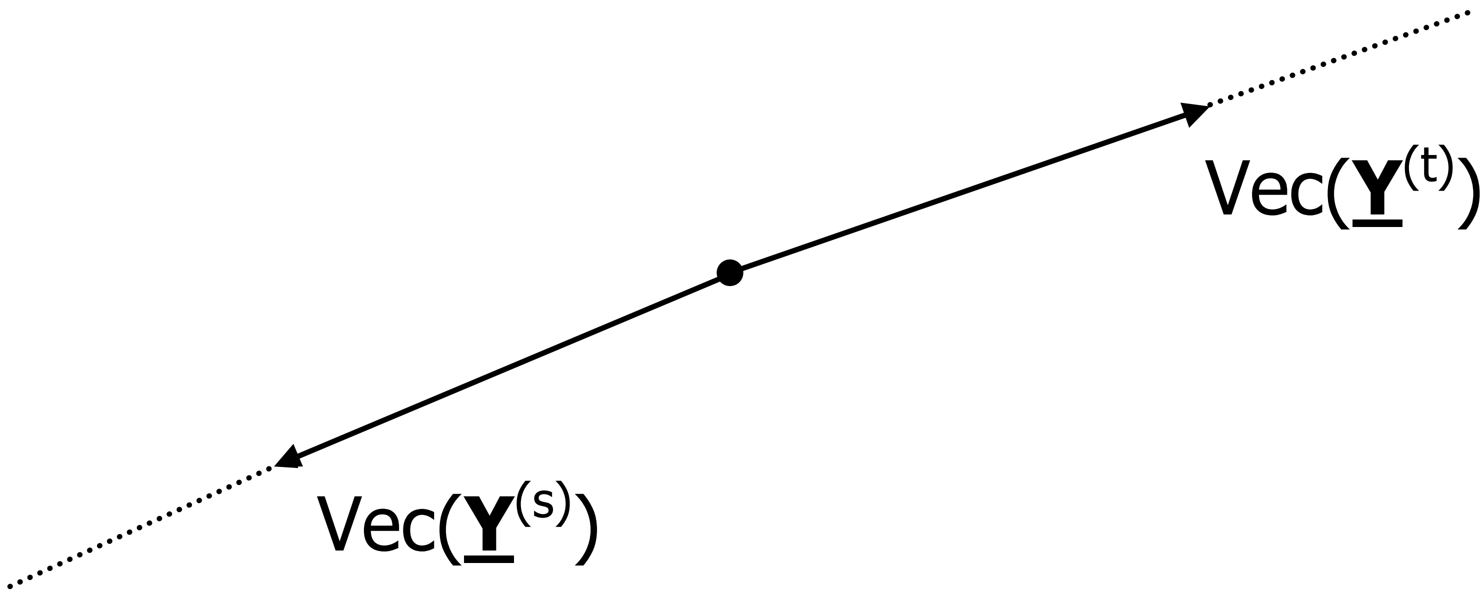
➔ diverging components (“degeneracy”)

<u>Also</u> : slow convergence of iterative CPD algorithm

Harshman & Lundy (1984), Kruskal et al. (1989), Krijnen et al. (2008), Stegeman & De Lathauwer (2011)

Two diverging components

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



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

Remarks on diverging components

- CPD sequence (**A**,**B**,**C**) may contain several groups of diverging components

- In each group of diverging components $\cos(\mathbf{a}_s,\mathbf{a}_t)\cdot\cos(\mathbf{b}_s,\mathbf{b}_t)\cdot\cos(\mathbf{c}_s,\mathbf{c}_t)$ is close to ± 1 (a.e.)

- For random data $\underline{\mathbf{Z}}$ diverging components occur very often (up to 60-100%)

- Diverging components cannot be interpreted and must be avoided when interpretation is the goal

Best low-rank approximation of I×J×2 arrays

$\underline{\underline{\mathbf{Z}}}$	rank($\underline{\underline{\mathbf{Z}}}$)	R	Best rank-R ?
I = J	I+1	R = I	zero volume
I = J	I+1	R < I	pos. volume
I = J	I	R < I	pos. volume
I > J	min(I,2J)	(I,2J) > R > J	almost everywhere
I > J	min(I,2J)	R = J	pos. volume
I > J	min(I,2J)	R < J	pos. volume

Stegeman (2013b)

How to avoid diverging CPD components (1)

➔ make sure an optimal CPD solution exists

- **A**, **B** or **C** is constrained to have orthogonal columns (Harshman & Lundy, 1984; Krijnen et al., 2008)
- $\underline{\mathbf{Z}}$ is nonnegative and **A**, **B** and **C** are constrained to be nonnegative (Lim, 2005; Lim & Comon, 2009)
- Constraining $\cos(\mathbf{a}_s,\mathbf{a}_t)\cdot\cos(\mathbf{b}_s,\mathbf{b}_t)\cdot\cos(\mathbf{c}_s,\mathbf{c}_t)$ (Lim & Comon, 2010)
- Add penalty term for deviations from orthogonal **A**,**B**,**C** (Rocci & Giordani, 2013)

How to avoid diverging CPD components (2)

➔ change the CPD problem into: (De Silva & Lim, 2008)

Minimize	ssq(**Z** – **Y**)
over	closure of S_R

What is needed?

➔ Complete characterization of boundary points

➔ Algorithm to find an optimal boundary point

Boundary points and algorithms are known for :

- $I \times J \times K$ and $R=2$
  via constrained HOSVD/Tucker3 of size $2 \times 2 \times 2$
  (Rocci & Giordani, 2010)

- $I \times J \times 2$ and $R \leq \min(I,J)$
  via Generalized Schur Decomposition
  (Stegeman & De Lathauwer, 2009; Stegeman, 2010)

- in both cases we do not need a CPD algorithm !

- in both cases the solution can be transformed to CPD form when no diverging components occur

Finding the optimal boundary point in general

Each group of diverging CPD components has a limit with a specific decomposition form

For R=2 diverging components: $(\mathbf{A},\mathbf{B},\mathbf{C}) \rightarrow (\mathbf{S},\mathbf{T},\mathbf{U})\cdot\underline{\mathbf{G}}$

$$\text{with} \quad \underline{\mathbf{G}} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right] \qquad \text{and} \quad \operatorname{rank}(\underline{\mathbf{G}}) = 3$$

$$(\mathbf{S},\mathbf{T},\mathbf{U})\cdot\underline{\mathbf{G}} = (\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + (\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + (\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2)$$

De Silva & Lim (2008)

R=3 or 4 diverging components: $(\mathbf{A},\mathbf{B},\mathbf{C}) \rightarrow (\mathbf{S},\mathbf{T},\mathbf{U})\cdot\underline{\mathbf{G}}$

$$\underline{\mathbf{G}} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & * & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \text{and } \operatorname{rank}(\underline{\mathbf{G}}) = 5$$

$$\underline{\mathbf{G}} = \left[\begin{array}{cccc|cccc|cccc|cccc} 1 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\text{and } \operatorname{rank}(\underline{\mathbf{G}}) \geq 7$$

Stegeman (2012, 2013a)

For several groups of diverging components or a combination of diverging and nondiverging components:

add the terms to obtain a decomposition of the limit $\underline{\mathbf{X}}$

Algorithm

1. Run a CPD algorithm, obtain solution (**A**,**B**,**C**).
2. When diverging components occur, order them in groups and determine decomposition form of limit $\underline{\mathbf{X}}$.
3. Compute initial values for decomposition of $\underline{\mathbf{X}}$ from (**A**,**B**,**C**).
4. Fit decomposition form of $\underline{\mathbf{X}}$ to data $\underline{\mathbf{Z}}$ using initial values from (**A**,**B**,**C**). Simple ALS algorithm !

Stegeman (2012, 2013a), Kiers & Smilde (1998)

Numerical Example: 6×6×6 and R=6

CPD ALS with tolerance 1e-9 terminates after 19.645 iters

$\underline{\mathbf{Y}} = (\mathbf{A},\mathbf{B},\mathbf{C})$ has 2+3 diverging components

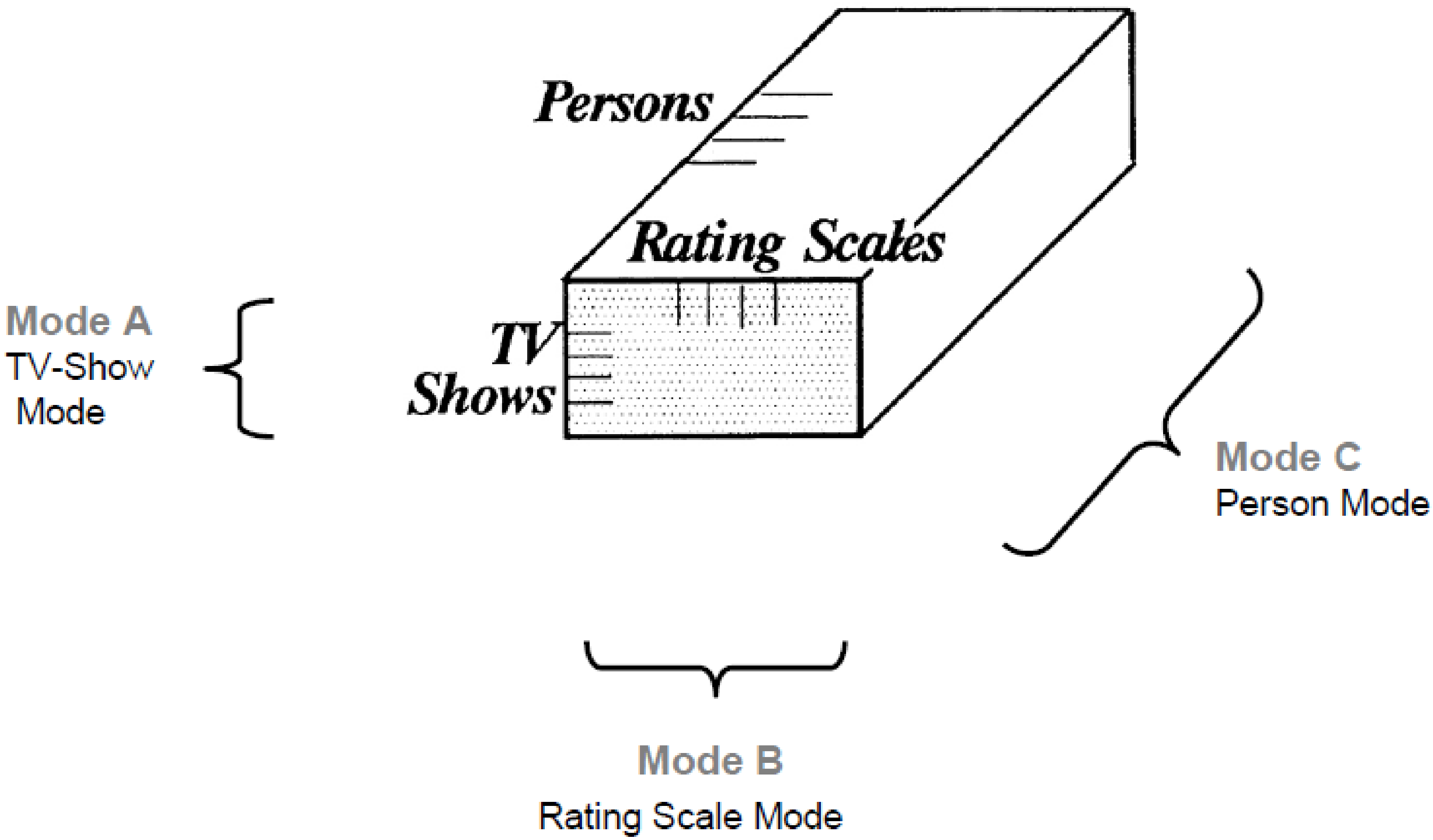
$\mathrm{ssq}(\underline{\mathbf{Z}} - \underline{\mathbf{Y}}) = 54.5370$

fit model $\underline{\mathbf{Z}} = (\mathbf{s}_1,\mathbf{t}_1,\mathbf{u}_1) + (\mathbf{S}_2,\mathbf{T}_2,\mathbf{U}_2)\cdot\underline{\mathbf{G}}_2 +$
$(\mathbf{S}_3,\mathbf{T}_3,\mathbf{U}_3)\cdot\underline{\mathbf{G}}_3 + \underline{\mathbf{E}}$

$\mathrm{ssq}(\underline{\mathbf{Z}} - \underline{\mathbf{X}}) = 54.5336$, tolerance 1e-12, 137 iters

condition numbers of **S**, **T**, **U** are: 21.8, 6.3, 61.0

Application: CPD analysis of TV-ratings (1981)



15 TV shows x 16 Rating Scales x 30 Persons

15 TV shows


M*A*S*H
THE COMPLETE SERIES
CHARLIE'S ANGELS
ALL IN THE FAMILY
The TONIGHT SHOW WITH JAY LENO
LET'S MAKE A DEAL
NOS JOURNAAL
"WHO LOVES YA, BABY?"
TELLY SAVALAS
KOJAK
MORK & MINDY
J. ALLEN
32
98
MUTUAL of OMAHA'S WILD KINGDOM
WILDKINGDOM.COM

16 Rating Scales

1. Thrilling / Boring
2. Intelligent / Idiotic
3. Erotic / Not Erotic
4. Sensitive / Not
5. Interesting / Not
6. Fast / Slow
7. Intellectually Stimulating / Dull
8. Violent / Peaceful
9. Caring / Callous
10. Satirical / Not
11. Informative / Not
12. Touching / Leaves Me Cold
13. Deep / Shallow
14. Tasteful / Crude
15. Real / Fantasy
16. Funny / Not

CPD with R=3 yields two diverging components

Fit = 50.76 %	Comp.3 = “Humor” (24.38 %)
Comps. 1 & 2 have	triple cosine = -0.996
	ssq $\approx$ 100 $\cdot$ ssq(comp.3)

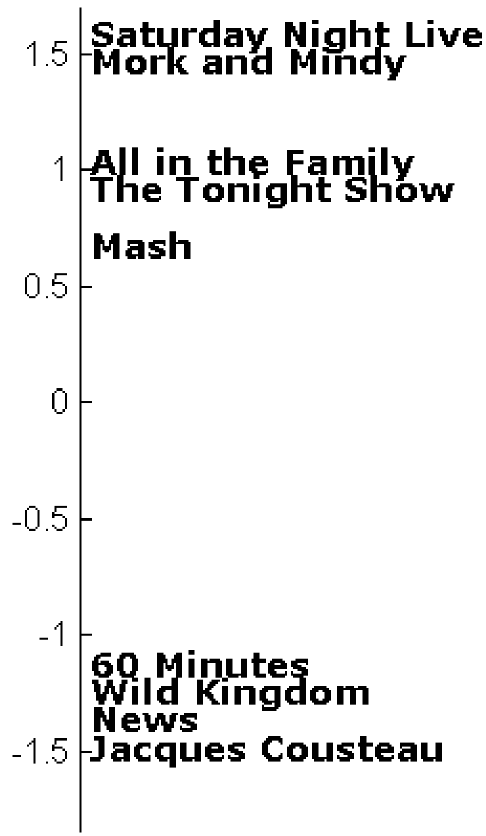
Try CPD with R=3 and orthogonal TV show components:

Fit = 50.22 %	“Humor”	27.19 %
	“Sensitivity”	13.04 %
	“Violence”	9.99 %

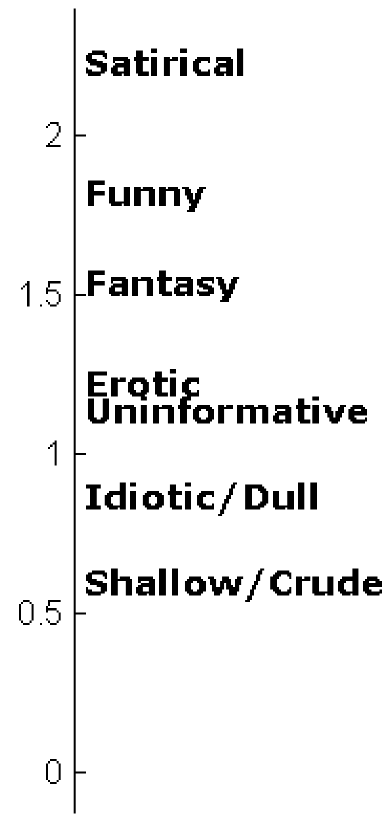
Lundy et al. (1989), Harshman (2004)

Component 1 = "Humor"

TV shows mode



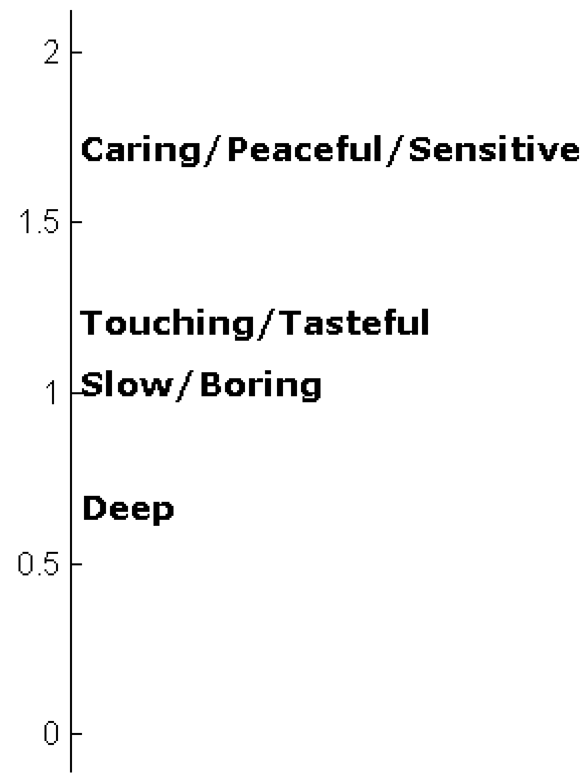
Scales mode



Component 2 = “Sensitivity”

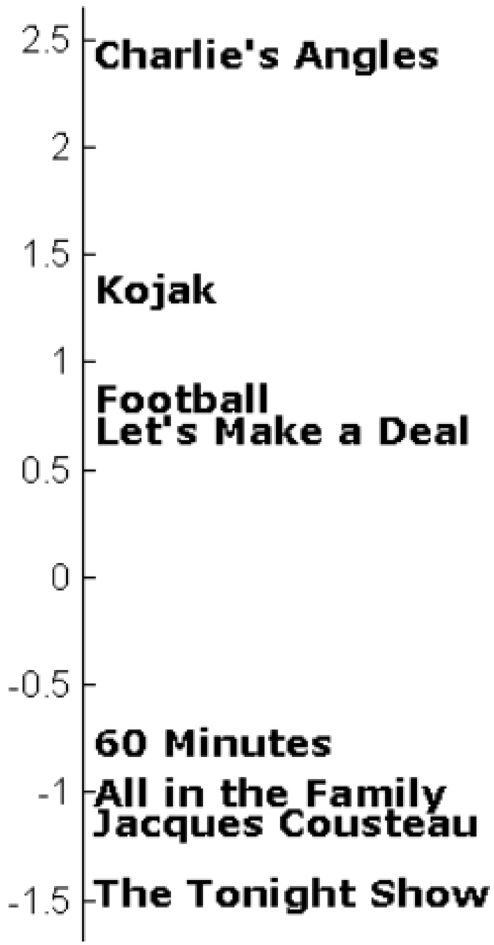
TV shows mode

Scales mode

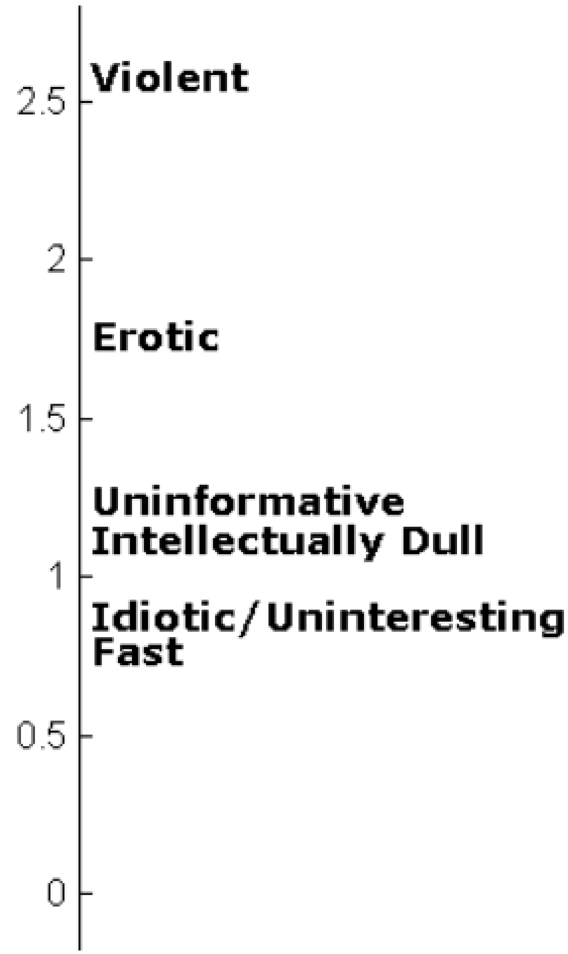


Component 3 = "Violence"

TV shows mode



Scales mode



Decomposition of the limit $\mathbf{X}$ in 4 terms

Fit = 50.7571 % (50.7569 for R=3 unconstrained)

$$\underline{\mathbf{X}} = (\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + (\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + (\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2) + (\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$$

$(\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1)$:	"Violence"	7.62 %
$(\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1)$:	"Sensitivity"	10.75 %
$(\mathbf{s}_1 \circ \mathbf{t}_2 \circ \mathbf{u}_2)$:	interaction	1.55 %
$(\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$:	"Humor"	24.37 %

Stegeman (2014)

Final Remarks

- Avoid diverging CPD components by either imposing constraints (orthogonality, nonnegativity) or including the boundary of the rank-R set
- When constraints are not appropriate, finding the optimal boundary point $\underline{\mathbf{X}}$ and its decomposition is a good alternative
- For the TV-ratings data, the decomposition of the optimal boundary point $\underline{\mathbf{X}}$ yields interpretable oblique components

References

- Bini, D., Capovani, M., Romani, F. & Lotti, G. (1979). *$O(n^{2.7799})$* complexity for *$n \times n$* approximate matrix multiplication. *Information Processing Letters*, **8**, 234-235.
- De Silva, V. & Lim, L-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1084-1127.
- Harshman, R.A., & Lundy, M.E. (1984). Data preprocessing and the extended Parafac model. In: *Research Methods for Multimode Data Analysis*, H.G. Law et al., Editors, Praeger, New York, pp. 216-284.
- Harshman, R.A. (2004). The problem and nature of degenerate solutions or decompositions of 3-way arrays. Talk at the Tensor Decompositions Workshop, Palo Alto, CA, American Institute of Mathematics.
- Kiers, H.A.L. & Smilde, A.K. (1998). Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. *Journal of Chemometrics*, **12**, 125-147.
- 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*, **73**, 431-439.

- 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: *Multiway Data Analysis*, Coppi R. & Bolasco, S. (editors), North-Holland, 115-121.
- Lim, L.-H. (2004). What's possible and what's impossible in tensor decompositions/approximations. Talk at the Tensor Decompositions Workshop, Palo Alto, CA, American Institute of Mathematics.
- Lim, L.-H. (2005). Optimal solutions to non-negative Parafac/ multilinear NMF always exist. Talk at the Workshop of Tensor Decompositions and Applications, CIRM, Luminy, Marseille, France.
- Lim, L.-H. & Comon, P. (2009). Nonnegative approximations of nonnegative tensors. *Journal of Chemometrics*, **23**, 432-441.
- Lim, L.-H. & Comon, P. (2010). Multiarray signal processing: tensor decompositions meet compressed sensing. *Comptes-Rendus de l'Académie des Sciences, Mécanique*, **388**, 311-320.
- Lundy, M.E., Harshman, R.A., & Kruskal, J.B. (1989). A two-stage procedure incorporating good features of both trilinear and quadrilinear models. In: *Multiway Data Analysis*, R. Coppi & S. Bolasco (Editors), North-Holland, pp. 123-130.

- Paatero, P. (2000). Construction and analysis of degenerate Parafac models. *Journal of Chemometrics*, **14**, 285-299.
- Rocci, R. & Giordani, P. (2010). A weak degeneracy revealing decomposition for the Candecomp/Parafac model. *Journal of Chemometrics*, **24**, 57-66.
- Rocci, R. & Giordani, P. (2013). Constrained Candecomp/Parafac via the Lasso. *Psychometrika*, **78**, 669-684.
- Stegeman, A. (2010). The Generalized Schur Decomposition and the rank-R set of real IxJx2 arrays. Technical Report, arXiv:1011.3432.
- Stegeman, A. (2012). Candecomp/Parafac – from diverging components to a decomposition in block terms. *SIAM Journal on Matrix Analysis and Applications*, **33**, 291-316.
- Stegeman, A. (2013a). A three-way Jordan canonical form as limit of low-rank tensor approximations. *SIAM Journal on Matrix Analysis and Applications*, **34**, 624-650.
- Stegeman, A. (2013b). On the (non)existence of best low-rank approximations of generic IxJx2 arrays. Technical Report, arXiv:1309.5727.
- Stegeman, A. (2014). Finding the limit of diverging components in three-way Candecomp/Parafac – a demonstration of its practical merits. *Computational Statistics and Data Analysis*, **75**, 203-216.

- Stegeman, A. & De Lathauwer, L. (2009). A method to avoid diverging components in the Candecomp/Parafac model for generic IxJx2 arrays. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1614-1638.
- Stegeman, A. & De Lathauwer, L. (2011). Are diverging CP components always nearly proportional? Technical Report, arXiv:1110.1988.