Netherlands Organisation for Scientific Research

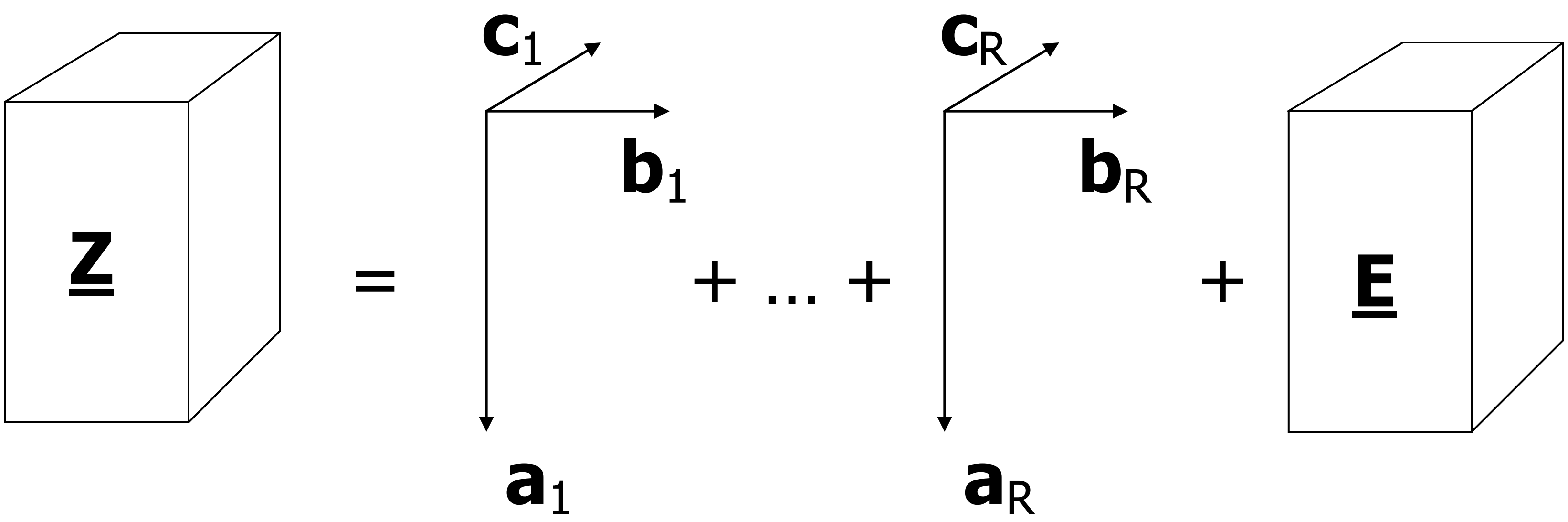
university of groningen

A three-way Jordan canonical form as limit of low-rank tensor approximations

Alwin Stegeman

a.w.stegeman@rug.nl www.alwinstegeman.nl

3-way Canonical Polyadic Decomposition (CPD)



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

Goal: Find $(\mathbf{A},\mathbf{B},\mathbf{C})$ that minimize $\| \underline{\underline{\mathbf{E}}} \|$

$\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_R]$, $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_R]$

Everything you see is Real

$$\mathbb{R}$$

3-way CPD as Optimization Problem

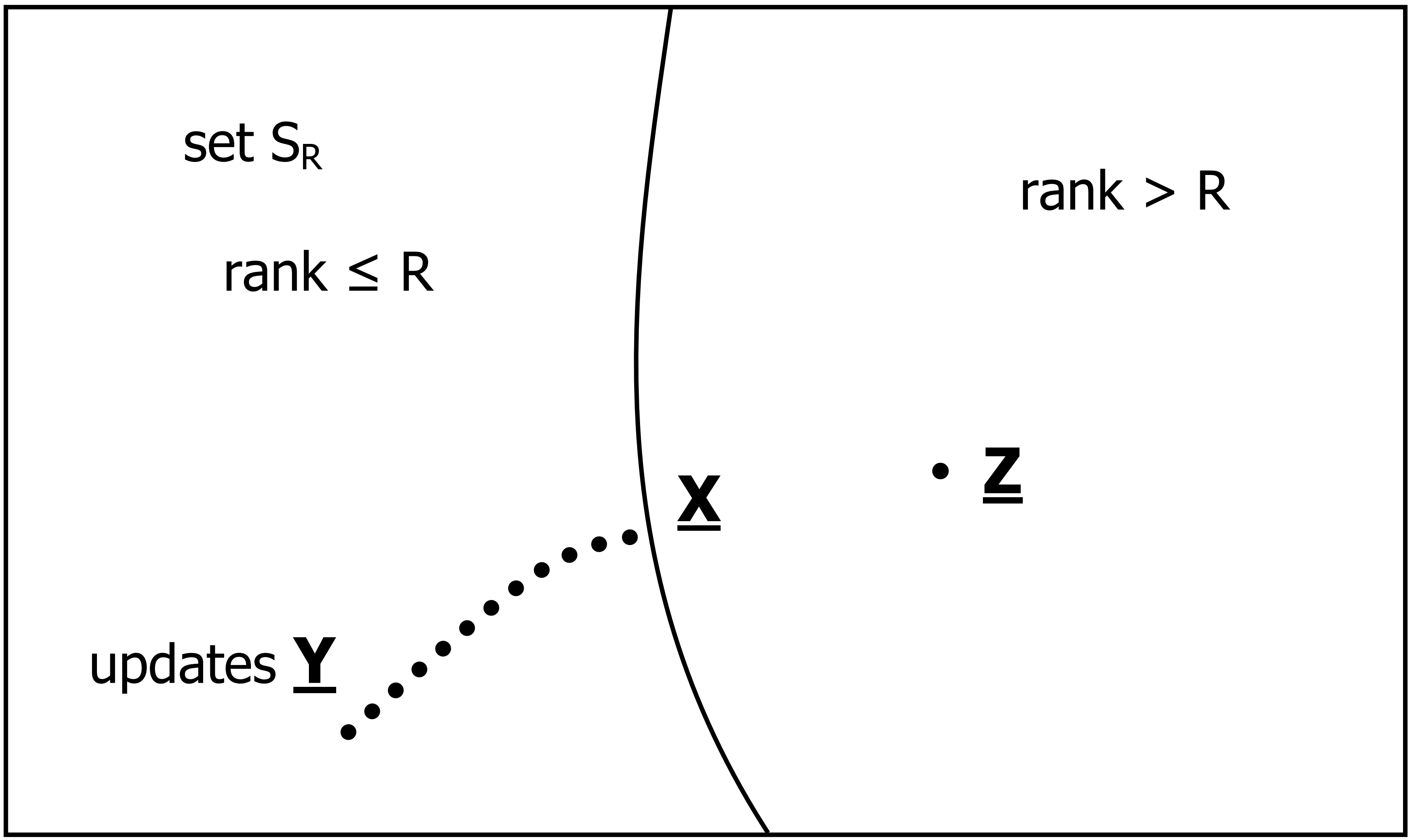
Minimize	$\| \underline{\underline{\mathbf{Z}}} - \underline{\mathbf{Y}} \|$
over	$S_R = \{ \underline{\mathbf{Y}} : \ \text{rank}(\underline{\mathbf{Y}}) \leq R \}$

➔ 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 groups of rank-1 terms $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$ converge to **linear dependency** and **infinite norm**

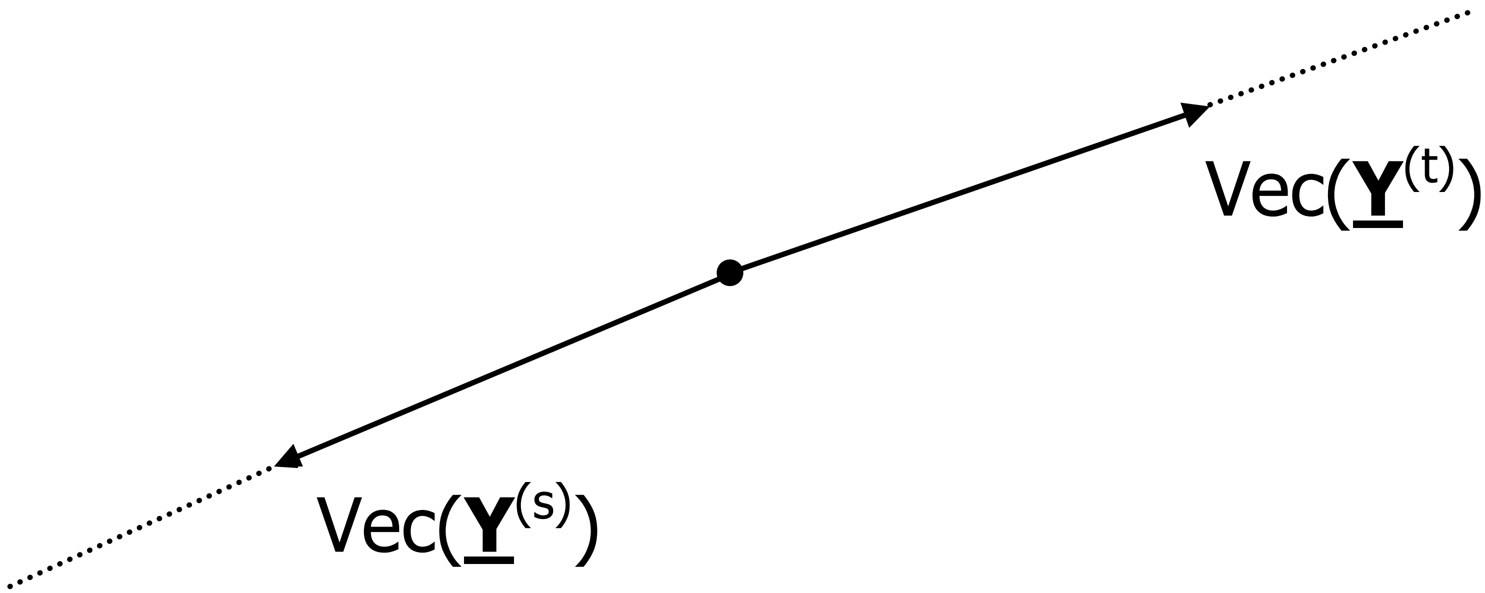
➔ diverging rank-1 terms / components ("degeneracy")

Also : slow convergence ("swamp") of CPD algorithm

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

Two diverging rank-1 terms

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



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

Remarks on diverging rank-1 terms

- CPD sequence (**A**,**B**,**C**) may contain several groups of diverging rank-1 terms

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

- For random data $\underline{\mathbf{Z}}$ diverging rank-1 terms may occur very often (up to 60-100%)

- Diverging rank-1 terms 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)	J < R < min(I,2J)	almost everywhere
I > J	min(I,2J)	R = J	pos. volume
I > J	min(I,2J)	R < J	pos. volume

Stegeman (2015)

How to avoid diverging rank-1 terms in CPD (1)

➔ optimal CPD solution exists under constraints:

- **A**, **B** or **C** have orthogonal columns
  (Harshman & Lundy, 1984; Krijnen et al., 2008)

- $\underline{\mathbf{Z}}$ and **A**, **B** and **C** are nonnegative
  (Lim, 2005; Lim & Comon, 2009)

- $\cos(\mathbf{a}_s,\mathbf{a}_t)\cdot\cos(\mathbf{b}_s,\mathbf{b}_t)\cdot\cos(\mathbf{c}_s,\mathbf{c}_t)$ is bounded
  (Lim & Comon, 2010)

- Add penalty terms / Tikhonov regularization
  (Giordani & Rocci, 2013ab; Navasca et al., 2008; Li et al., 2013)

How to avoid diverging rank-1 terms in CPD (2)

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

Minimize	$\| \underline{\mathbf{Z}} - \underline{\mathbf{Y}} \|$
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 $2\times2\times2$

  (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 rank-1 terms occur

Finding the optimal boundary point in general

Assumption: Each group of d diverging rank-1 terms has a limit with rank > d

Theorem: The limit of d=2 diverging rank-1 terms can be written as $(\mathbf{S},\mathbf{T},\mathbf{U})\cdot\underline{\mathbf{G}}$

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

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

De Silva & Lim (2008)

Theorem: If the limit of d=3 diverging rank-1 terms has multilinear rank (3,3,3), then it can be written as (**S**,**T**,**U**)$\cdot$**G** (a.e.) with

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

and rank(**G**) = 5

Stegeman (2012)

Theorem: If the limit of d=4 diverging rank-1 terms has multilinear rank (4,4,4), then it can be written as $(\mathbf{S},\mathbf{T},\mathbf{U})\cdot\underline{\mathbf{G}}$ (a.e.) with

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

and rank($\underline{\mathbf{G}}$) ≥ 7

Stegeman (2013)

Diverging rank-1 terms → Block term decomp.

$$\underline{\mathbf{Y}} = (\mathbf{A}_1,\mathbf{B}_1,\mathbf{C}_1) + \ldots + (\mathbf{A}_m,\mathbf{B}_m,\mathbf{C}_m)$$

$$\downarrow$$

$$\underline{\mathbf{X}} = (\mathbf{S}_1,\mathbf{T}_1,\mathbf{U}_1)\bullet\underline{\mathbf{G}}_1 + \ldots + (\mathbf{S}_m,\mathbf{T}_m,\mathbf{U}_m)\bullet\underline{\mathbf{G}}_m$$

$\underline{\mathbf{G}}_j = 1$ for a nondiverging rank-1 term ($d_j=1$)

$\underline{\mathbf{G}}_j$ ($d_j \times d_j \times d_j$) for d_j diverging rank-1 terms ($d_j>1$)

De Lathauwer (2008)

Algorithm

1. Run a CPD algorithm, obtain solution (**A**,**B**,**C**)

2. When diverging rank-1 terms occur, order them in groups and determine decomposition form of limit **X**

3. Compute initial values for decomposition of **X** from (**A**,**B**,**C**)

4. Fit decomposition form of **X** to data **Z** using initial values from (**A**,**B**,**C**). Simple ALS algorithm !

Stegeman (2012, 2013), 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

$\| \underline{\underline{\mathbf{Z}}} - \underline{\mathbf{Y}} \|^2 = 54.5370$

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

$\| \underline{\underline{\mathbf{Z}}} - \underline{\mathbf{X}} \|^2 = 54.5336$, tolerance 1e-12, 137 iters

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

Diverging rank-1 terms in a matrix problem

Minimize	$\| \mathbf{Z} - \mathbf{Y} \|$
over	$D_R = \{ \mathbf{Y} = \mathbf{A}\,\mathbf{C}\,\mathbf{A}^{-1},\ \mathbf{C} \text{ diagonal} \}$

Theorem: For generic $\mathbf{Z}$ with some complex eigenvalues:

(i) the set D_R ($R\times R$ matrices) is not closed for $R \geq 2$

(ii) no optimal solution exists

(iii) pairs of diverging rank-1 terms occur in $\mathbf{A}\,\mathbf{C}\,\mathbf{A}^{-1}$
when converging to optimal boundary point $\mathbf{X}$

Stegeman (2013)

Link to the (real) Jordan canonical form

- Each pair of diverging rank-1 terms corresponds to identical eigenvalues of **X** with only one eigenvector

- Optimal boundary point **X** satisfies the real Jordan form $\mathbf{P}\,\mathbf{J}\,\mathbf{P}^{-1}$, with $\mathbf{J} = \text{blockdiag}(\mathbf{J}_1, \ldots, \mathbf{J}_m)$ where

$\mathbf{J}_j = \lambda_j$ for a nondiverging rank-1 term

$$\mathbf{J}_j = \begin{bmatrix} \lambda_j & 1 \\ 0 & \lambda_j \end{bmatrix}$$ for each pair of diverging rank-1 terms

Stegeman (2013)

<u>3-way Jordan form of the CPD limit point</u>

- CPD $\underline{\mathbf{Y}} = (\mathbf{A},\mathbf{B},\mathbf{C}) \cdot \underline{\mathbf{I}}_R$ implies diagonalization when $\mathbf{A},\mathbf{B},\mathbf{C}$ have rank R

- Optimal boundary point $\underline{\mathbf{X}}$ has decomposition $(\mathbf{S},\mathbf{T},\mathbf{U}) \cdot \underline{\mathbf{G}}$, with $\underline{\mathbf{G}}$ = blockdiag($\underline{\mathbf{G}}_1, \ldots, \underline{\mathbf{G}}_m$) where

  $\underline{\mathbf{G}}_j = 1$ for a nondiverging rank-1 term

  $\underline{\mathbf{G}}_j = d_j \times d_j \times d_j$ canonical form for d_j diverging rank-1 terms

Final Remarks

- Avoid diverging rank-1 terms in CPD by
  (i) imposing constraints, or by
  (ii) including the boundary of the rank-R set

For method (ii) as presented, we have

- Uniqueness properties of decomposition of **X** for $\max(d_j)=2$ in Stegeman (2012, 2014)
- Simulations with random **Z** in Stegeman (2012, 2013)
- Application to TV-ratings data in Stegeman (2014)
- Matlab codes online at www.alwinstegeman.nl

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 Lathauwer, L. (2008). Decompositions of a higher-order tensor in block terms – Part II: definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1033-1066.
- 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.
- Giordani, P. & Rocci, R. (2013a). Constrained Candecomp/Parafac via the Lasso. *Psychometrika*, **78**, 669-684.
- Giordani, P. & Rocci, R. (2013b). Candecomp/Parafac with ridge regularization. *Chemometrics and Intelligent Laboratory Systems*, **129**, 3-9.
- 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.
- Li, N., Kindermann, S. & Navasca, C. (2013). Some convergence results on the regularized alternating least-squares method for tensor decomposition. *Linear Algebra and its Applications*, **438**, 796-812.
- 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.
- Navasca, C., De Lathauwer, L. & Kindermann, S. (2008). Swamp reducing technique for tensor decomposition. Proceedings of the 16th European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland.
- 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.
- 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. (2013). 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. (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. (2015). On the (non)existence of best low-rank approximations of generic IxJx2 arrays. Technical Report, arXiv:1309.5727.
- 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.