Decomposing a three-way dataset of TV-ratings when this is impossible

Alwin Stegeman

a.w.stegeman@rug.nl

www.alwinstegeman.nl

Summarizing Data in Simple Patterns

Information Technology → collection of huge data sets, often multi-way data z(i,j,k,…)

Approximation: Multi-way data ≈ 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: $\mathbf{Z} = \mathbf{a}\,\mathbf{b}^T = \mathbf{a} \circ \mathbf{b}$ $\Leftrightarrow$ $z(i,j) = a(i) \cdot b(j)$

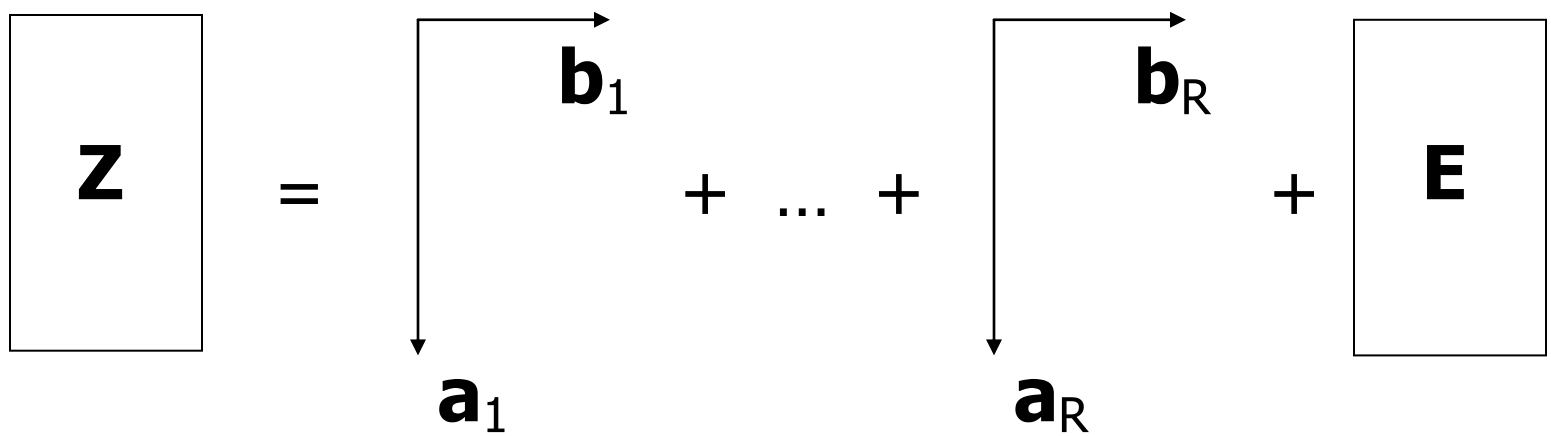
$\mathrm{rank}(\mathbf{Z}) = \min \{R : \mathbf{Z} = \mathbf{a}_1 \circ \mathbf{b}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R \}$

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

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

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

2-way (PCA) decomposition

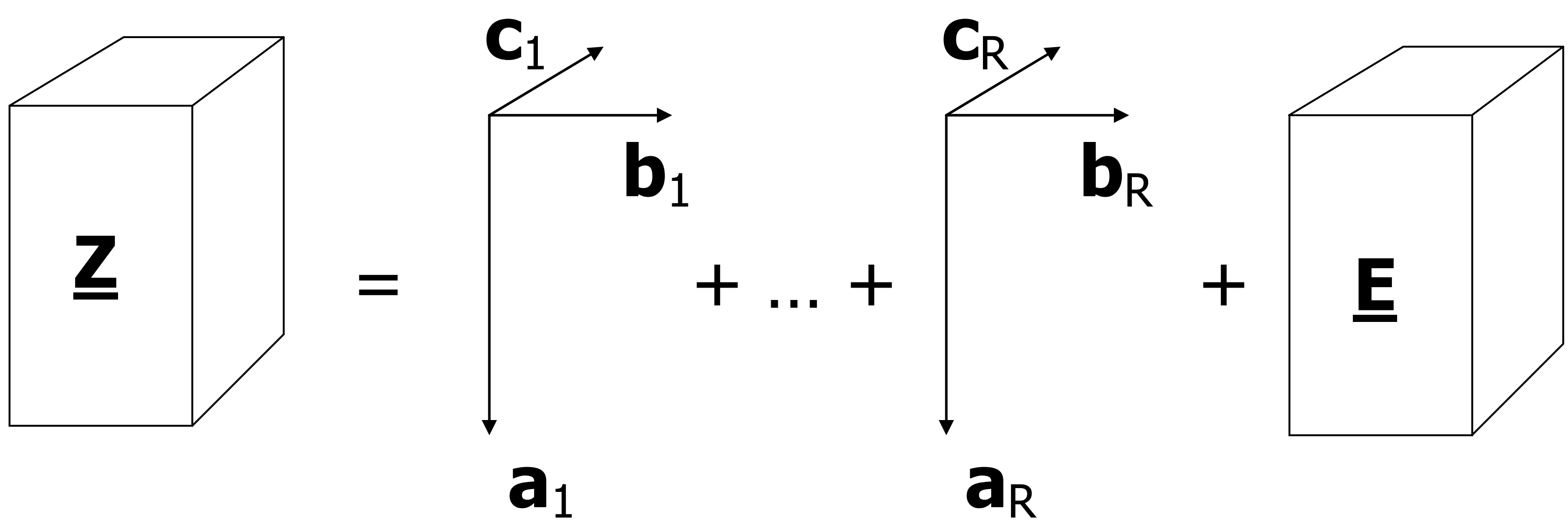


$$\mathbf{Z} = \mathbf{a}_1 \circ \mathbf{b}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R + \mathbf{E}$$

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

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

3-way Candecomp/Parafac (CP)



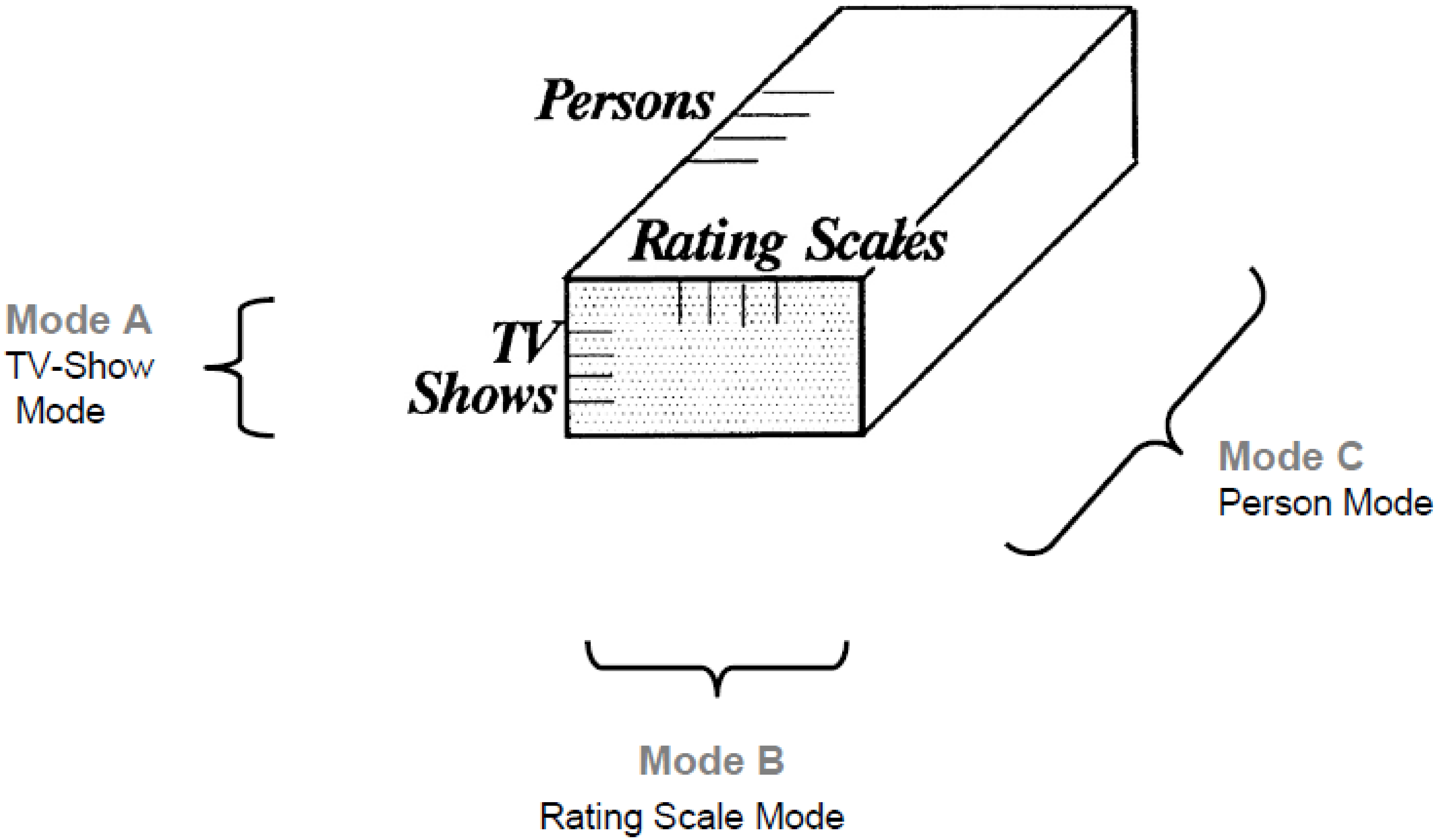
$$\underline{\mathbf{Z}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \ldots + \mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R + \underline{\mathbf{E}}$$

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

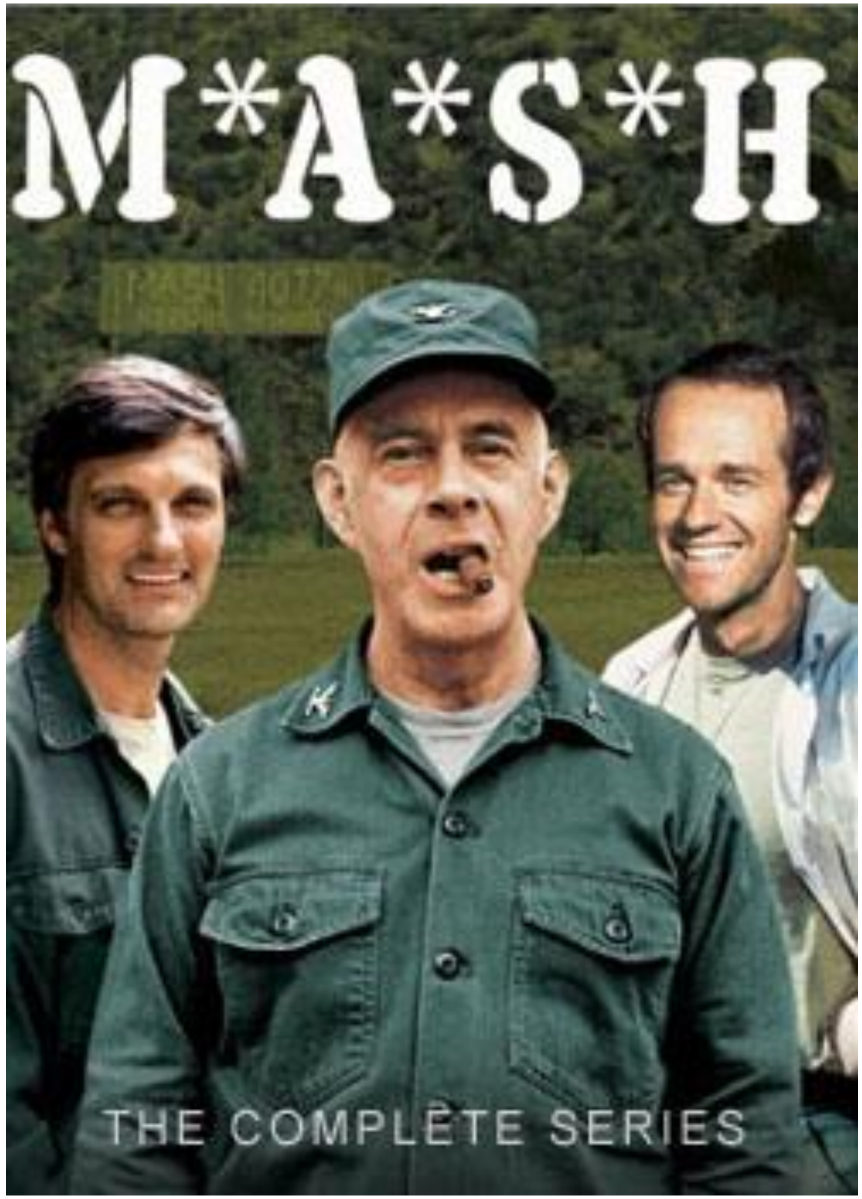
with $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_R]$

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

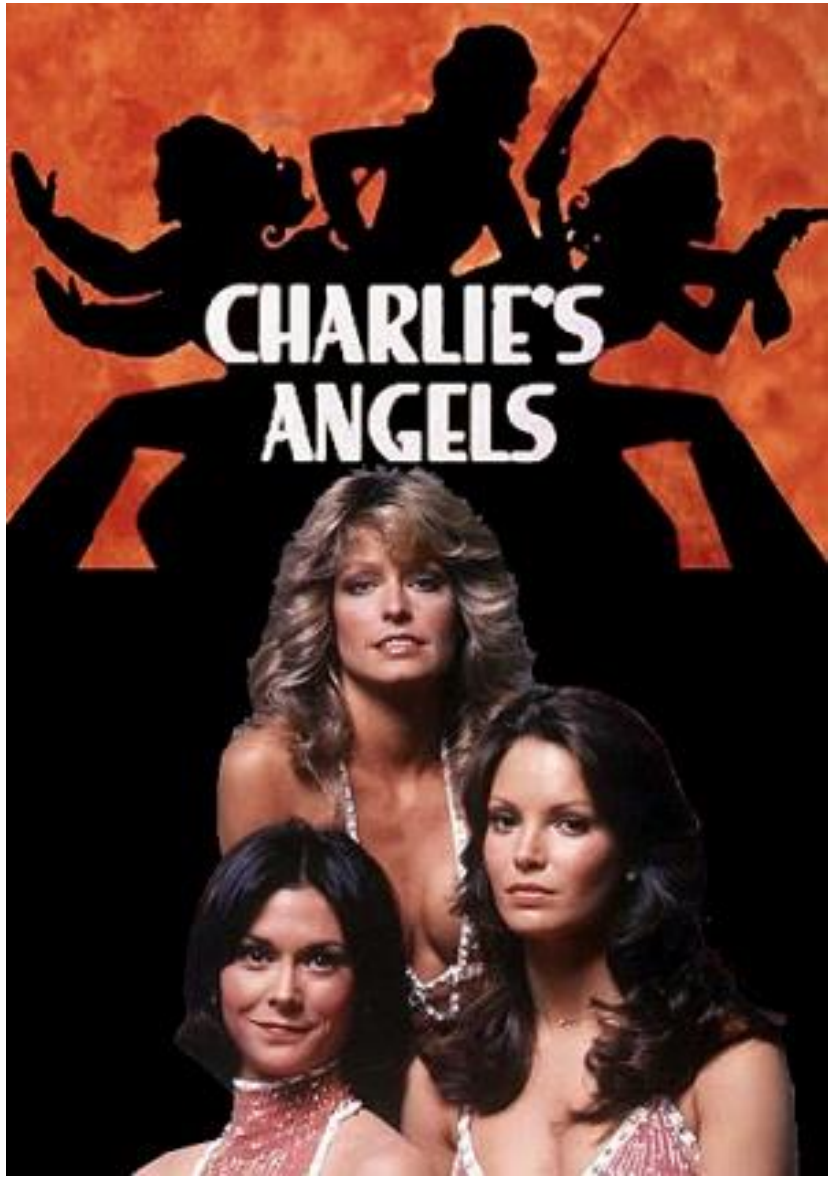
CP analysis of 3-way TV-ratings data



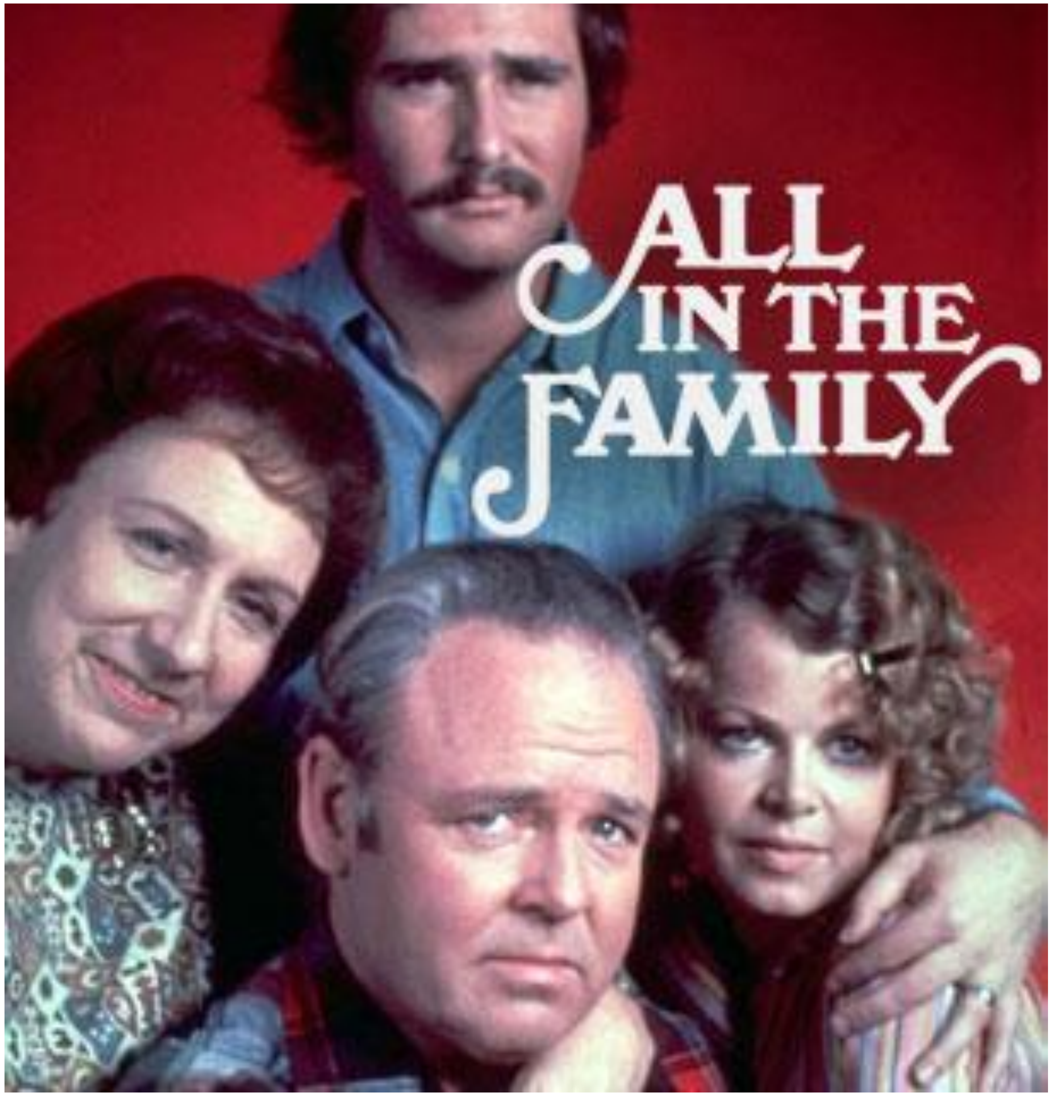
TV show 1 – Mash



TV show 2 – Charlie's Angels



TV show 3 – All in the Family



TV show 4 – 60 Minutes

TV show 5 – The Tonight Show

TV show 6 – Let's Make a Deal

TV show 7 – The Waltons

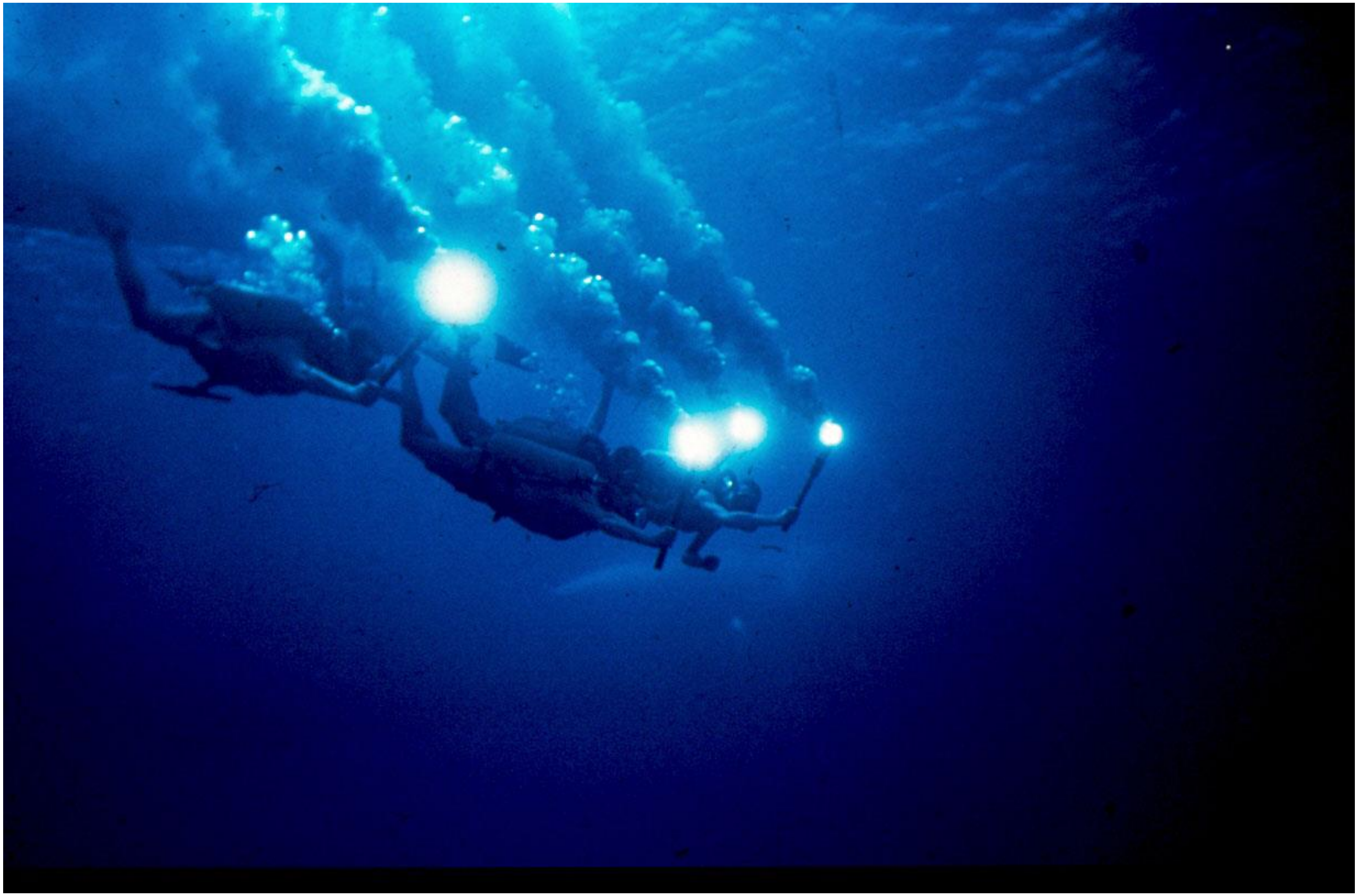
TV show 8 – Saturday Night Live

TV show 9 – News

TV show 10 – Kojak

TV show 11 – Mork and Mindy

TV show 12 – Jacques Cousteau



TV show 13 – Football

TV show 14 – Little House on the Prairie

TV show 15 – Wild Kingdom

Rating Scales 1-8

-6, -5, …, -1 , 0, 1, …, 5, 6

1. Thrilling . . . . Boring
2. Intelligent . . . Idiotic
3. Erotic . . . . . . Not Erotic
4. Sensitive . . . . Insensitive
5. Interesting . . Uninteresting
6. Fast . . . . Slow
7. Intellectually Stimulating . . . Intellectually Dull
8. Violent . . . . Peaceful

Rating Scales 9-16

-6, -5, …, -1, 0, 1, …, 5, 6

9. Caring . . . . . . Callous
10. Satirical . . . . . Not Satirical
11. Informative . . . Uninformative
12. Touching . . . . "Leaves Me Cold"
13. Deep . . . . Shallow
14. Tasteful . . . . . Crude
15. Real . . . . Fantasy
16. Funny . . . . . . Not Funny

TV-ratings data

30 persons have rated 15 TV shows on 16 rating scales

Preprocessing:

- Centering across rating scales
- Centering across TV shows
- Normalizing within persons

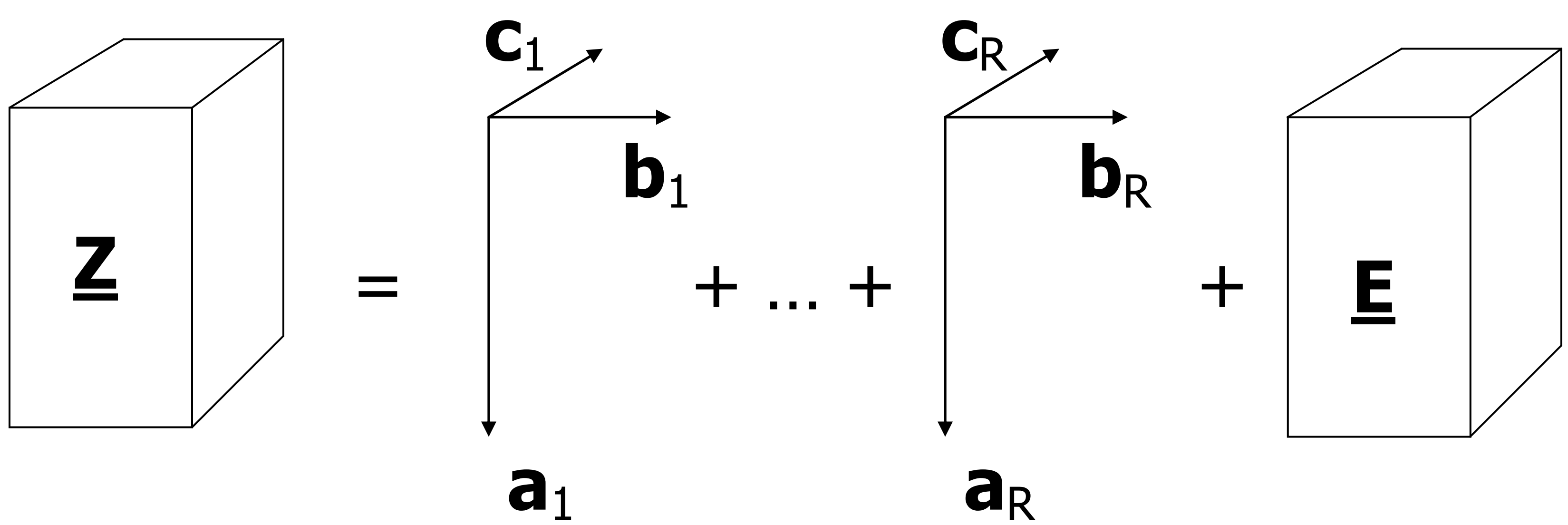
TV data also analyzed by Lundy et al. (1989) and Harshman (2004)
Analysis presented is by Stegeman (2014)

Output of the CP analysis with R components

Matrix **A** (15×R): columns are TV show components

Matrix **B** (16×R): columns are rating scales loadings

Matrix **C** (30×R): columns are person loadings



Scaling the CP solution

Column of **A**: mean squared component score = 1

Column of **B**: mean squared loading = 1

Column of **C**: sum of squared loadings = 4

$$\underline{\underline{\mathbf{Z}}} = g_1 (\mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1) + \ldots + g_R (\mathbf{a}_R \circ \mathbf{b}_R \circ \mathbf{c}_R) + \underline{\underline{\mathbf{E}}}$$

weight g_r indicates strength of component r

columns are sign changed such that **C** has positive loadings

Fit of the CP solution

Fit % = 100 – 100 ssq($\underline{\mathbf{E}}$) / ssq($\underline{\mathbf{Z}}$) (range 0 to 100)

Congruence coefficient of two components

$cc_A(1,2) = \dfrac{\mathbf{a}_1^T \mathbf{a}_2}{\sqrt{\text{ssq}(\mathbf{a}_1)}\sqrt{\text{ssq}(\mathbf{a}_2)}}$ (range -1 to +1)

$cc(1,2) = cc_A(1,2)\, cc_B(1,2)\, cc_C(1,2)$

The CP solution with 2 components

Overall:	fit = 41.96 %	cc(1,2) = 0.002
Component 1:	fit = 28.46 %	g_1 = 1.46
Component 2:	fit = 13.59 %	g_2 = 1.01

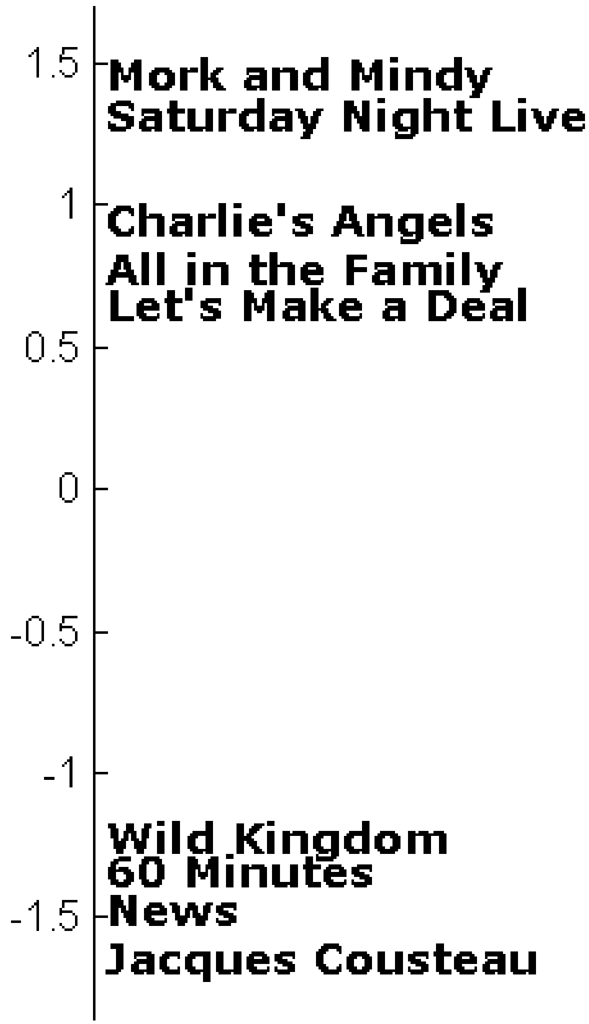
Interpretation:

Component 1 = "Humor"

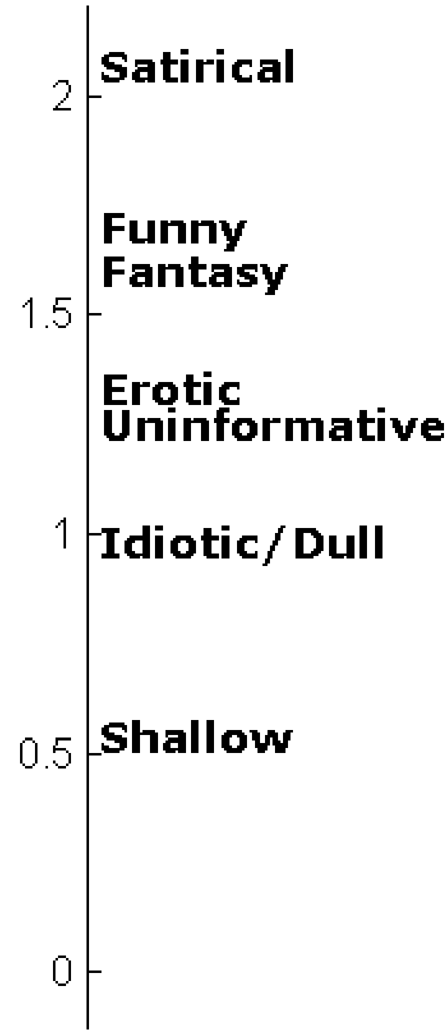
Component 2 = "Sensitivity"

Component 1 = “Humor”

TV shows mode

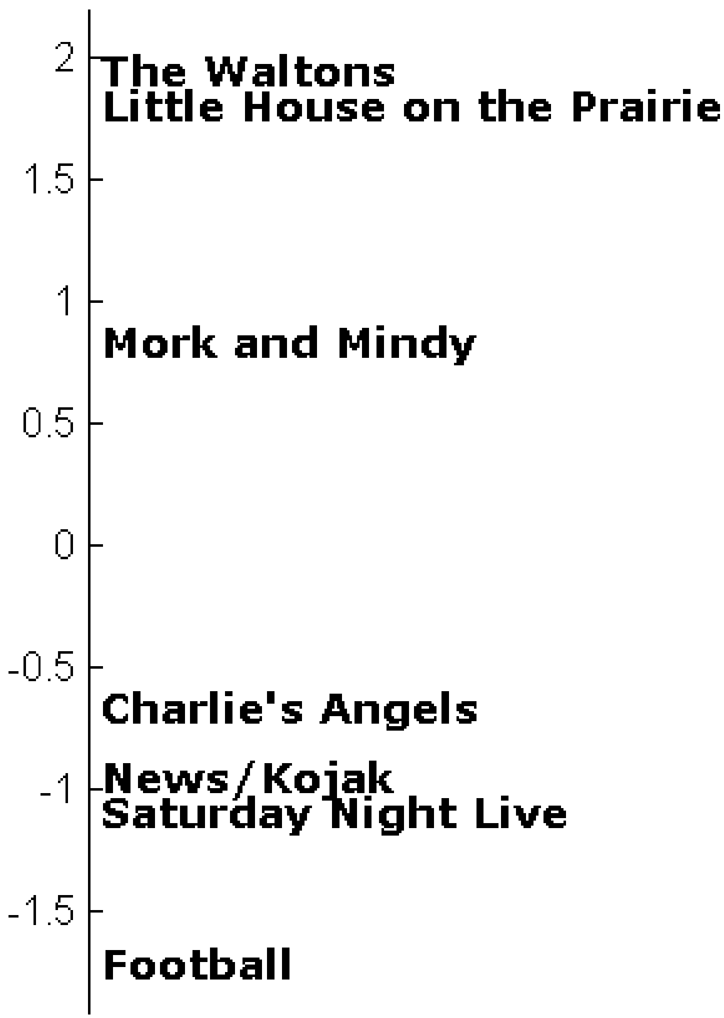


Scales mode

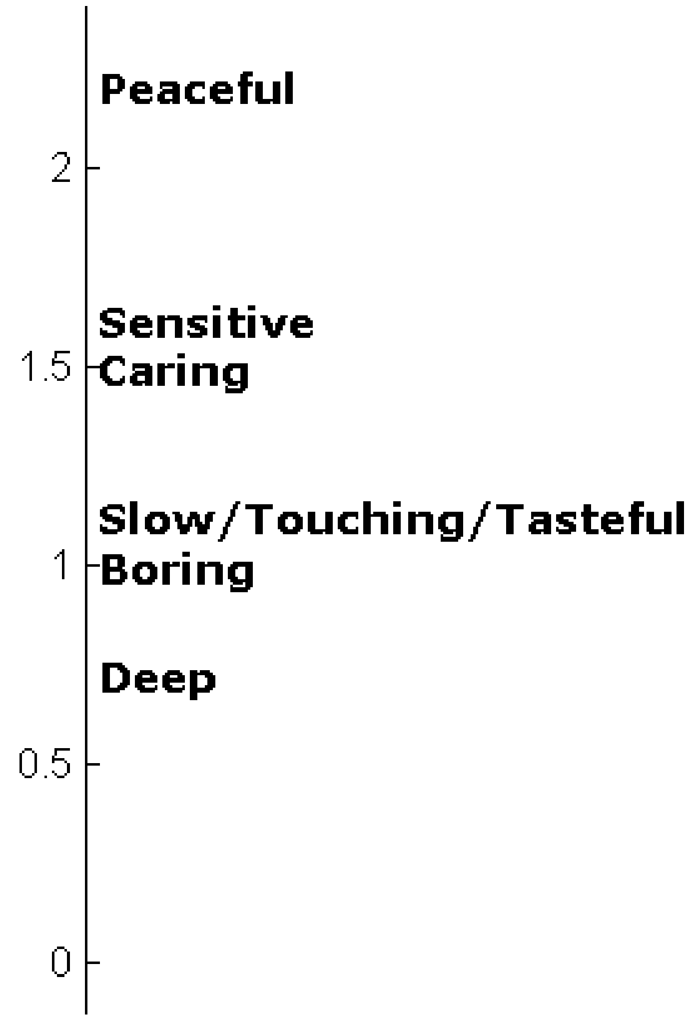


Component 2 = “Sensitivity”

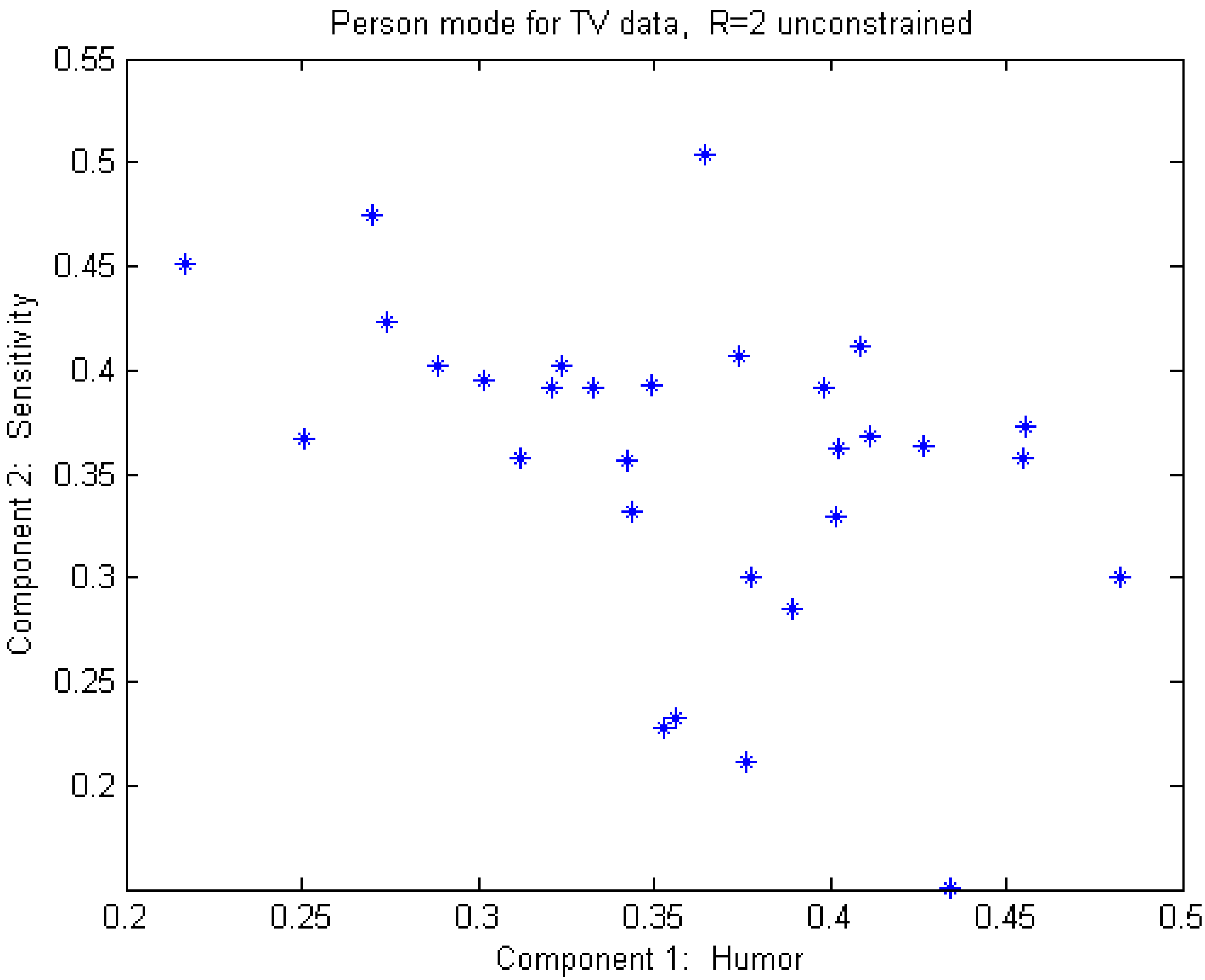
TV shows mode



Scales mode



Components 1 and 2 – persons plot



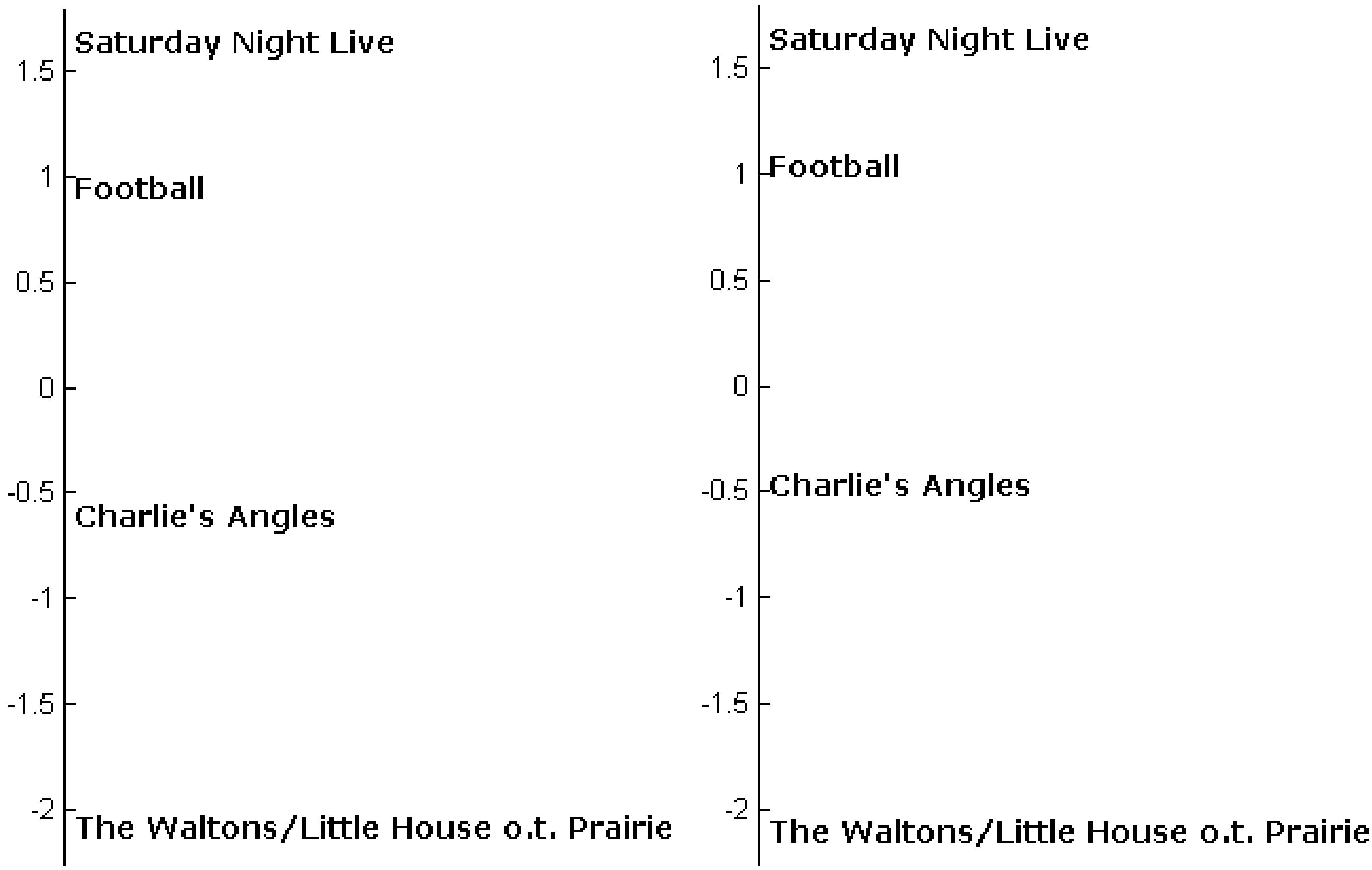
The CP solution with 3 components

Overall:	fit = 50.76 %	cc(1,2) = -0.996
		cc(1,3) = -0.13
		cc(2,3) = 0.12
Comps. 1+2:	fit = 20.11 %	g_1 = 15.23
		g_2 = 15.39
Component 3:	fit = 24.38 %	g_3 = 1.52

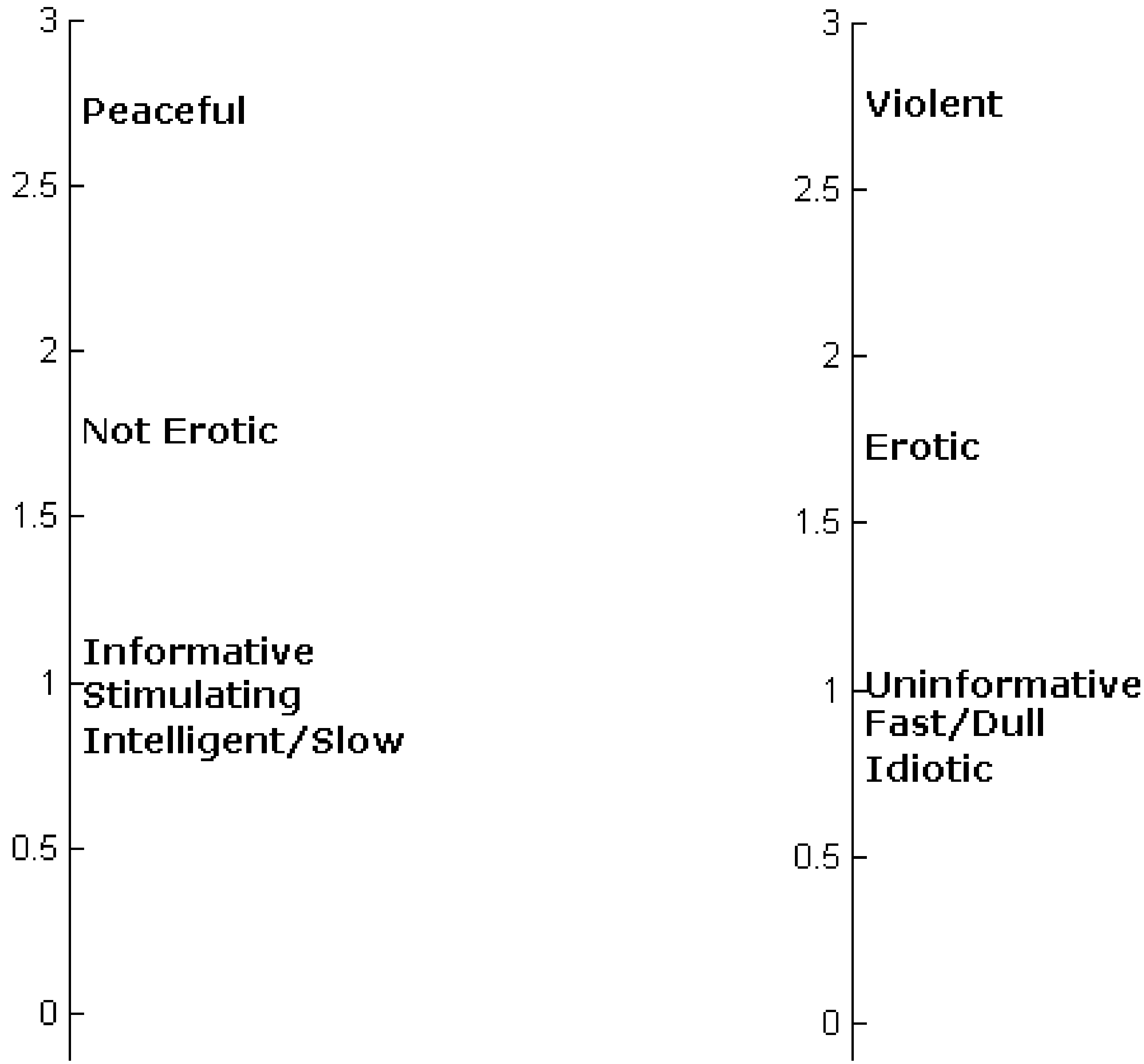
Interpretation: Components 1 & 2 = ???

Component 3 = "Humor"

Components 1 and 2 – TV shows mode



Components 1 and 2 – Scales mode



Comparing the solutions for R=2 and R=3

congruence coefficients of R=2 components (columns) and R=3 components (rows):

	"Humor"	"Sensitivity"
Comp. 1	-0.15	-0.41
Comp. 2	0.15	0.46
"Humor"	0.93	0.01

Some Theory

- Diverging components occur when CP does not have an optimal solution, i.e., best rank-R approximation does not exist (Krijnen et al., 2008; De Silva & Lim, 2008)

- CP has an optimal solution if the columns of **A** (or **B** or **C**) are restricted to be orthogonal
(Harshman & Lundy, 1984; Krijnen et al., 2008)

- CP has an optimal solution if the data is nonnegative and **A**,**B**,**C** are restricted to be nonnegative
(Lim, 2005; Lim & Comon, 2009)

R=3 components and orthogonal TV shows mode

Overall:	fit = 50.22 %	cc(r1,r2) = 0
Component 1:	fit = 27.19 %	g_1 = 1.43
Component 2:	fit = 13.04 %	g_2 = 0.99
Component 3:	fit = 9.99 %	g_3 = 0.87

Interpretation:

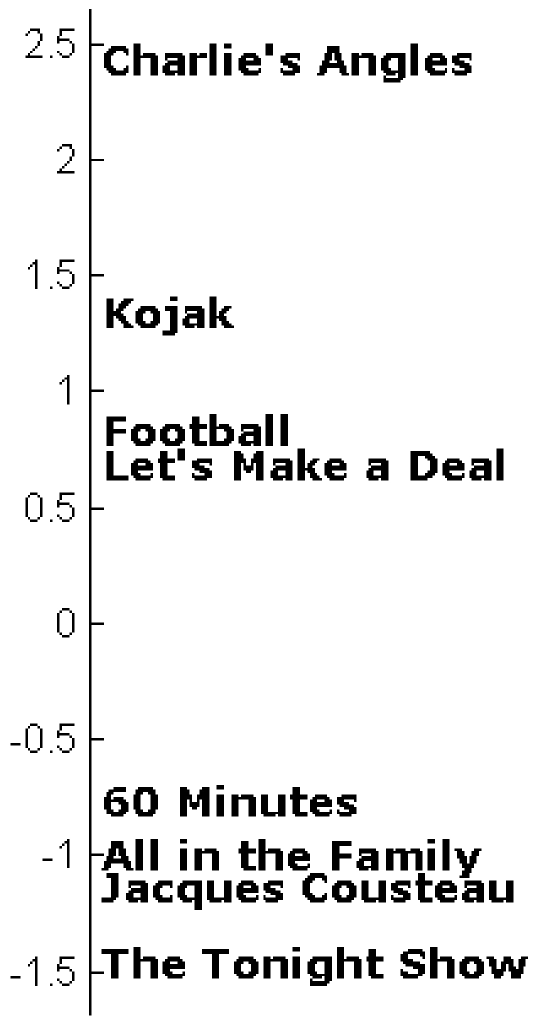
Component 1 = “Humor”

Component 2 = “Sensitivity”

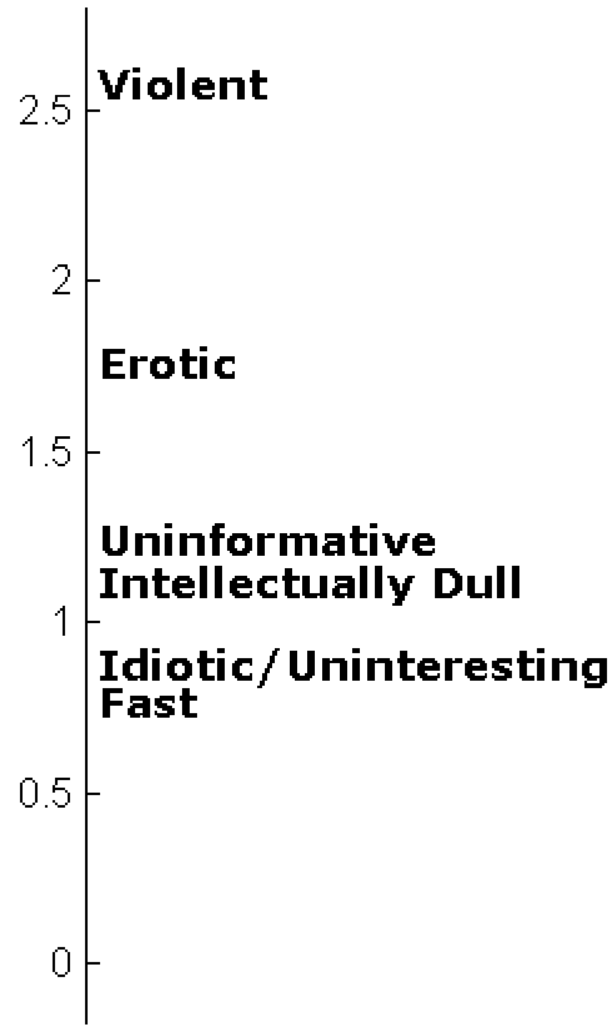
Component 3 = “Violence”

Component 3 = "Violence"

TV shows mode



Scales mode

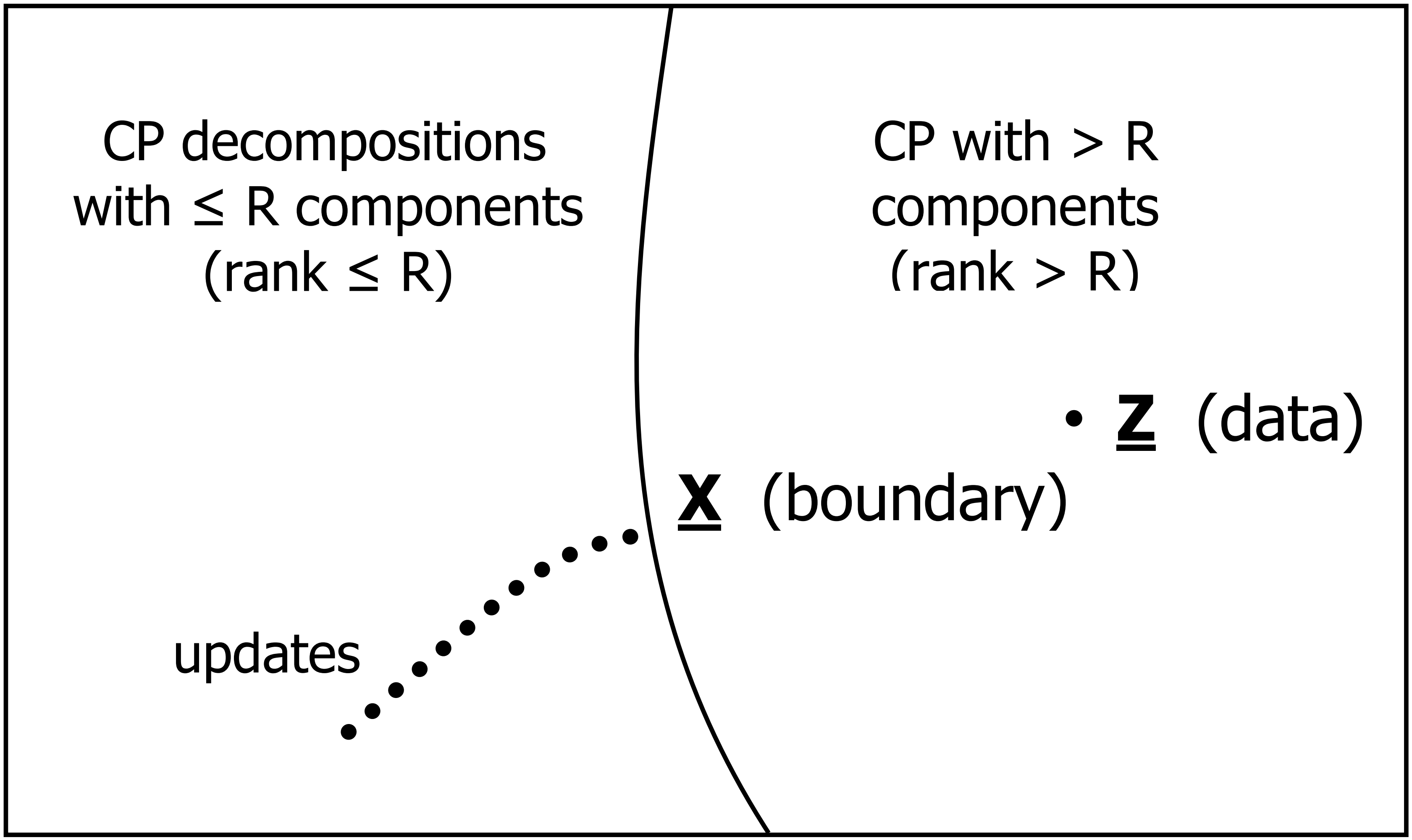


Comparing the solutions obtained so far

	R=2		R=3		
R=3 orth.	“H”	“S”	Comp.1	Comp.2	“H”
“Humor”	0.96	0.00	-0.20	0.19	0.95
“Sensitivity”	0.00	0.94	-0.30	0.36	0.03
“Violence”	0.11	0.08	0.30	-0.24	-0.03
“H” (R=2)			-0.15	0.15	0.93
“S” (R=2)			-0.41	0.46	0.01

➔ the two diverging components relate to “S” and “V”

Some more Theory



- CP does not have an optimal solution if optimal boundary point $\underline{\mathbf{X}}$ does not have rank ≤ R

- In that case, the decomposition of $\underline{\mathbf{X}}$ contains one or more interaction terms, e.g., $\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2$

➔ How to find $\underline{\mathbf{X}}$ and its decomposition?

Algorithms exist for:

- I×J×2 arrays and R ≤ min(I,J)
  (Stegeman & De Lathauwer, 2009)

- I×J×K arrays and R=2
  (Rocci & Giordani, 2010)

Two-stage method for I×J×K arrays and general R

First fit CP. In case of diverging components, do this:

- For combinations of nondiverging and groups of 2,3,4 diverging components, the form of the decomposition of the limit point **X** has been proven (Stegeman, 2012,2013)

- This form of decomposition is fitted to the data **Z** with initial values obtained from the diverging CP decomposition (Stegeman, 2012,2013)

- This yields **X** and its decomposition with interaction terms (Stegeman, 2012,2013)

The form of the limit of two diverging components is:

$$g_{111}\,(\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + g_{221}\,(\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + g_{212}\,(\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2)$$

For the TV data with R=3, we fit the decomposition:

$$g_{111}\,(\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + g_{221}\,(\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + g_{212}\,(\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2)$$

$$+ g_{333}\,(\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$$

Decomposition of the limit point in 4 terms

Overall:	fit = 50.7571 %	(50.7569 for R=3)
g_{111} ($\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1$):	fit = 7.62 %	g_{111} = 0.99
g_{221} ($\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1$):	fit = 10.75 %	g_{221} = 0.95
g_{212} ($\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2$):	fit = 1.55 %	g_{122} = 0.33
g_{333} ($\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3$):	fit = 24.37 %	g_{333} = 1.52

Interpretation: $\mathbf{s}_1$ and $\mathbf{t}_1$ = "Violence"

$\mathbf{s}_2$ and $\mathbf{t}_2$ = "Sensitivity"

$\mathbf{s}_3$ and $\mathbf{t}_3$ = "Humor"

Comparison to R=3 solution with orth. TV shows

	"Humor"	"Sensitivity"	"Violence"
g_{111} ($\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1$)	-0.07	0.13	0.86
g_{221} ($\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1$)	0.05	0.81	0.02
g_{212} ($\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2$)	-0.03	-0.02	-0.03
g_{333} ($\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3$)	0.95	0.04	-0.03

Interpretation of the decomposition in 4 terms

$\mathbf{s}_r$ = TV show component r
$\mathbf{t}_r$ = Rating scale loadings r
$\mathbf{u}_r$ = Idealized person r

	TV shows	scales	id. person	weight g
$(\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1)$	Violent	Violence	1	0.99
$(\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1)$	Sensitive	Sensitivity	1	0.95
$(\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2)$	Sensitive	Violence	2	0.33
$(\mathbf{s}_3 \circ \mathbf{t}_3 \circ \mathbf{u}_3)$	Humorous	Humor	3	1.52

Note: $\mathbf{u}_2 > 0$ or $\mathbf{u}_2 < 0$ varies per person

Remarks on TV-ratings analysis

- The decomposition of the limit point resembles the R=3 CP solution with orthogonal scales.
- However, orthogonality between "Sensitivity" and "Violence" is not intuitive.
- The sign of the interaction term between "Sensitive" TV shows and "Violence" scales differs per person.
- The decomposition $g_{111}\ (\mathbf{s}_1 \circ \mathbf{t}_1 \circ \mathbf{u}_1) + g_{221}\ (\mathbf{s}_2 \circ \mathbf{t}_2 \circ \mathbf{u}_1) + g_{212}\ (\mathbf{s}_2 \circ \mathbf{t}_1 \circ \mathbf{u}_2)$ is only partially unique, but this does not affect interpretation.

General Remarks

- General results on (non)existence of best rank-R approximations only exist for:

  - R=1 best approx. always exists
  - $\underline{\underline{\mathbf{Z}}}$ 2×2×2 and R=2

    (De Silva & Lim, 2008)

  - $\underline{\underline{\mathbf{Z}}}$ generic I×J×2 and R≥2

    (Stegeman, 2006, 2008, 2015)

- The form of the decomposition of the limit point $\underline{\underline{\mathbf{X}}}$ is a generalization of the Jordan canonical form for matrices

  (Stegeman, 2013)

References

- 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.
- 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.
- 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.

- 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.
- Rocci, R. & Giordani, P. (2010). A weak degeneracy revealing decomposition for the Candecomp/Parafac model. *Journal of Chemometrics*, **24**, 57-66.
- Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for p×p×2 arrays of rank p+1 or higher. *Psychometrika*, **71**, 483-501.
- Stegeman, A. (2008). Low-rank approximation of generic p×q×2 arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Applications*, **30**, 988-1007
- 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 & Data Analysis*, **75**, 203-216.
- Stegeman, A. (2015). On the (non)existence of best low-rank approximations of generic IxJx2 arrays. Submitted. arXiv: 1309.5727
- Stegeman, A. & De Lathauwer, L. (2009). A method to avoid diverging components in the Candecomp/Parafac model for generic I xJ x2 arrays. *SIAM Journal on Matrix Analysis and Applications*, **30**, 1614-1638.