Three-mode Principal Component Analysis (3MPCA) and Item Response Theory

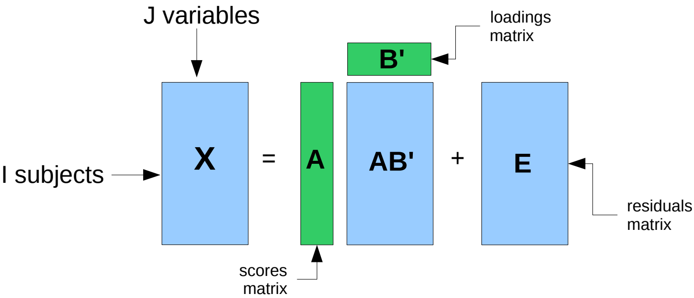
Jorge Tendeiro

28 July 2017

1. Three-way models
   - Models
   - Study 1: Simplicity
   - Study 2: First- and second-order derivatives

2. Item response theory
   - Background
   - Person-fit analysis
   - Unfolding models

Two-way array = data matrix
Scores of subjects (rows) on variables (columns).



$$\boxed{\mathbf{X} = \mathbf{AB}' + \mathbf{E}}$$

$$\boxed{\mathbf{X} = \mathbf{A}\mathbf{B}' + \mathbf{E}}$$

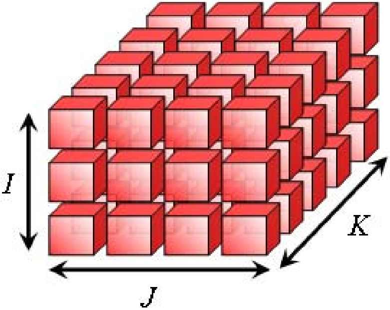
Goal: Representation of variables in low-dimension space.

$$\boxed{\#\text{ columns of } \mathbf{A} \text{ and } \mathbf{B} \quad \ll \quad \#\text{ columns of } \mathbf{X}}$$

$\mathbf{X}$ is approximated by a sum of R rank-1 matrices:

$$\boxed{\mathbf{X} \simeq \mathbf{a}_1\mathbf{b}_1' + \cdots + \mathbf{a}_R\mathbf{b}_R'}$$

...



- Generalize matrix structure to 3D.
- Easy to generalize to n-way.

$\underline{\mathbf{X}}$: $I \times J \times K$ array (I=subjects, J=variables, K=situations)
Number of components: R

Candecomp/Parafac

$$\mathbf{X}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}' + \mathbf{E}_k$$

▸ PCA ▸ ... ▸ ...

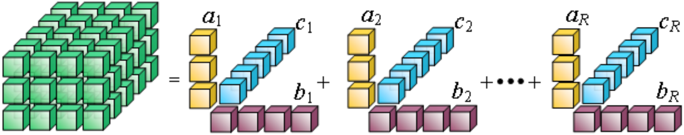
$$\mathbf{X}_k \simeq c_{k1}\mathbf{a}_1\mathbf{b}_1' + \cdots + c_{kR}\mathbf{a}_R\mathbf{b}_R'$$

▸ PCA

$\mathbf{A}$	$(I \times R)$:	"subjects" matrix
$\mathbf{B}$	$(J \times R)$:	"variables" matrix
$\mathbf{C}$	$(K \times R)$:	"situations" matrix
$\mathbf{C}_k$	$(R \times R)$:	$\mathrm{Diag}(\mathbf{c}_{k.})$

Minimize loss function:

$$f_{\mathrm{CP}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{A}\mathbf{C}_k\mathbf{B}'\|^2$$



Parallel proportional profiles (Cattell, 1944):
To simultaneously analyze several matrices together, find a set of common factors ($\mathbf{A}$, $\mathbf{B}$) that can be fitted with different weights ($\mathbf{C}_k, k = 1, \ldots, K$) to many data matrices at the same time.

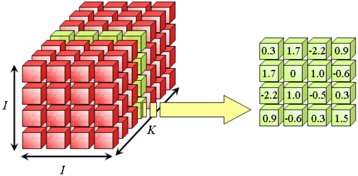
Similarities between PCA and CP:

- CP decomposes an array as a sum of rank one arrays.
- rank($\underline{\mathbf{X}}$)=minimum number of rank one arrays for which CP gives perfect fit

Differences between PCA and CP:

- Only iterative algorithm for CP.
- CP is usually unique.
- Preprocessing three-way data can be hard.

S: $I \times I \times K$ array with symmetric slices
(e.g., set of correlation matrices)



Model

$$\mathbf{S}_k = \mathbf{A}\mathbf{C}_k\mathbf{A}' + \mathbf{E}_k$$

CP ...

INDSCAL is CP with the constraint $\mathbf{A} = \mathbf{B}$.

Minimize loss function:

$$f_{\text{IND}}(\mathbf{A}, \mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{S}_k - \mathbf{A}\mathbf{C}_k\mathbf{A}'\|^2$$

Model more general than CP and INDSCAL:

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

- $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$: rank-1 array
- 3MPCA decomposes $\underline{\mathbf{X}}$ as a sum of rank-1 arrays
- rank($\underline{\mathbf{X}}$)$\leqslant PQR$ (usually rank($\underline{\mathbf{X}}$) $\ll PQR$)

3MPCA reduces to CP when the core array has a super-diagonal form:

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

Motivation from PCA:
For $\mathbf{S}$ nonsingular, $\mathbf{X} = \mathbf{A}\mathbf{B}' = (\mathbf{A}\mathbf{S})(\mathbf{S}^{-1}\mathbf{B}')$.

For 3MPCA:
For $\mathbf{S}$, $\mathbf{T}$, and $\mathbf{U}$ nonsingular,

$$\begin{aligned}
\mathbf{G} = [\mathbf{G}_1|\cdots|\mathbf{G}_R] &\longrightarrow \mathbf{S}'\mathbf{G}(\mathbf{U}\otimes\mathbf{T}) \\
\mathbf{A} &\longrightarrow \mathbf{A}(\mathbf{S}')^{-1} \\
\mathbf{B} &\longrightarrow \mathbf{B}(\mathbf{T}')^{-1} \\
\mathbf{C} &\longrightarrow \mathbf{C}(\mathbf{U}')^{-1}
\end{aligned}$$

This is known as a Tucker transformation.

Goal:
Find suitable linear combinations of frontal (and/or lateral and/or horizontal) slices that allow transforming $\underline{\mathbf{X}}$ into an "equivalent" array with many zero entries.

Formally:
Transform $\mathbf{X} = [\mathbf{X}_1|\cdots|\mathbf{X}_K]$ to $\mathbf{SX}(\mathbf{U} \otimes \mathbf{T}) := \mathbf{H} = [\mathbf{H}_1|\cdots|\mathbf{H}_K]$.

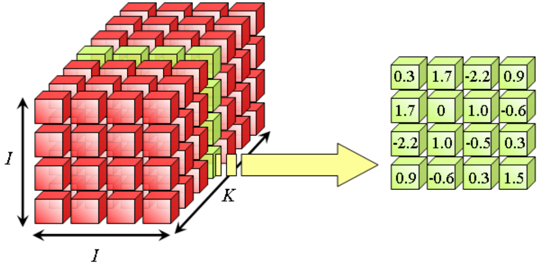
- Many zero entries in $\mathbf{H}$ = few nonzero entries.
- Weight of $\underline{\mathbf{H}}$ = # nonzero entries of $\underline{\mathbf{H}}$.

Why doing this?

1. Facilitate interpretation of 3MPCA decompositions (e.g., rotate the core $\underline{\mathbf{G}}$).
2. Constrained 3MPCA: Distinguish between tautologies and non-trivial models.
3. Mathematical applications: Typical rank, maximal rank.

- Diagonalize frontal slices of $\underline{\mathbf{G}}$ ($P = Q$).
  (Cohen, 1974, 1975; MacCallum, 1976; Kroonenberg, 1983)
- "Super-diagonalize" $\underline{\mathbf{G}}$ ($P = Q = R$; Kiers, 1992).
- Kiers' SIMPLIMAX (1998):
  $\underline{\mathbf{G}} \longrightarrow$ minimize ssq (m smallest elements)
- Other examples may be found in: Murakami, ten Berge, and Kiers (1998), ten Berge and Kiers (1999), and Rocci and ten Berge (2002).

One of my PhD's goals:
Simplify arrays with symmetric slices.



Consider $\underline{\mathbf{X}} = [\mathbf{X}_1|\cdots|\mathbf{X}_K]$: order $I \times I \times K$.
Assume that:

- $\underline{\mathbf{X}}$ is randomly sampled from a continuous distribution with symmetry constraint ($\mathbf{X}'_k = \mathbf{X}_k$, $\forall k$).
- The slices $\mathbf{X}_k$ linearly independent.

A symmetry-preserving transformation of $\underline{\mathbf{X}}$ is of the form

$$\mathbf{H}_l = \mathbf{S}' \left(\sum_k u_{kl} \mathbf{X}_k \right) \mathbf{S}, \quad l = 1, 2, \ldots, K,$$

with $\mathbf{S}_{I\times I}$ and $\mathbf{U}_{K\times K}$ nonsingular.

- The goal is to introduce as many zeros in $\underline{\mathbf{H}}$ as possible.
- I used the Orthogonal Complement Method (OCM; Rocci & ten Berge, 2002), but constrained to symmetry.

The OCM from Rocci and ten Berge (2002) generalized to symmetric slice arrays:
(Notation: $\mathbf{X}_{\text{vec}^*} = [\text{vec}^*(\mathbf{X}_1)| \ldots |\text{vec}^*(\mathbf{X}_K)]$)

1. Given the array $\mathbf{X}_{\text{vec}^*}$, compute an orthogonal complement $\mathbf{X}^{(c)}_{\text{vec}^*}$.
2. Compute $\mathbf{H}^{(c)}_{\text{vec}^*} = (\mathbf{S}^{-1} \otimes \mathbf{S}^{-1})\mathbf{X}^{(c)}_{\text{vec}^*}\mathbf{V}$ in such a way that $\mathbf{H}^{(c)}_{\text{vec}^*}$ is in simple form.
3. Find the orthogonal complement of $\mathbf{H}^{(c)}_{\text{vec}^*}$ in simple form, say $\mathbf{H}_{\text{vec}^*}$.
4. Find matrix $\mathbf{U}$ such that $\mathbf{H}_{\text{vec}^*} = (\mathbf{S}' \otimes \mathbf{S}')\mathbf{X}_{\text{vec}^*}\mathbf{U}$.

- Maximum number of lin. ind. slices: $K_{max} = \frac{I(I+1)}{2}$.
- The set of frontal slices form a basis for the space of $\mathbb{R}^{K_{max}}$, which is equivalent to the space of all symmetric $I \times I$ matrices.
- Therefore, a simple basis is easy to find (Rocci & ten Berge, 1994): (notation: $\mathbf{e}_i$= column i of $\mathbf{I}_I$)

$$\begin{array}{cc} \mathbf{e}_i\mathbf{e}_i', & i = 1, \dots, I \\ \mathbf{e}_i\mathbf{e}_j' + \mathbf{e}_j\mathbf{e}_i', & 1 \leqslant i < j \leqslant I \end{array}$$

Example: $I = 3$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Note: Only frontal slices were mixed.

- $K_{\text{max}} = 3$, so $K = 1, 2, 3$
- $2 \times 2 \times 3$: done (K_{max} situation)
- $2 \times 2 \times 1$: use EVD

$$\underline{\mathbf{X}} \longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}; \text{ if } \alpha < 0 : \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

- $2 \times 2 \times 2 =$ orthogonal complement of $2 \times 2 \times 1$

$$\underline{\mathbf{X}} \longrightarrow \left[\begin{array}{cc|cc} \alpha & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \end{array}\right]; \text{ if } \alpha < 0 : \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Conclusion for $2 \times 2 \times 2$:

- weight 4 is always possible
- if $\underline{\mathbf{X}}^c$ has eigenvalues of both signs then weight 2 is possible

- $K_{\max} = 6$, so $K = 1, 2, 3, 4, 5, 6$
- $3 \times 3 \times 6$: done ($K_{\max}$ situation)
- $3 \times 3 \times 1$: use EVD

$$\underline{\mathbf{X}} \longrightarrow \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}; \text{ if } d_2 d_3 < 0: \begin{bmatrix} d_1 & 0 & 0 \\ 0 & 0 & 2d_2 \\ 0 & 2d_2 & 0 \end{bmatrix}$$

- $3 \times 3 \times 5$ = orthogonal complement of $3 \times 3 \times 1$

$$\underline{\mathbf{X}} \longrightarrow \left[\begin{array}{ccc|ccc|ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \beta & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc|ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & \alpha & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

Conclusion for $3 \times 3 \times 5$:

- weight 10 is always possible
- if $\underline{\mathbf{X}}^c$ has eigenvalues of both signs then weight 9 is possible

- $3 \times 3 \times 2$: see $\mathrm{EVD}(\mathbf{X}_1^{-1}\mathbf{X}_2)$
  - real eigenvalues

$$\underline{\mathbf{X}} \longrightarrow \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & \beta & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]; \text{ also: } \left[\begin{array}{ccc|ccc} -\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{array}\right]$$

  - complex eigenvalues

$$\underline{\mathbf{X}} \longrightarrow \left[\begin{array}{ccc|ccc} -\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 1 & 0 \end{array}\right]$$

Conclusion for $3 \times 3 \times 2$:

- weight 5 is always possible
- if $\mathbf{X}_1^{-1}\mathbf{X}_2$ has real eigenvalues then weight 4 is possible

- $3 \times 3 \times 4$ = orthogonal complement of $3 \times 3 \times 2$

$$\left[\begin{array}{ccc|ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & \alpha & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta\alpha & 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right]$$

…

($\delta = 1/-1$ in real/complex case)

Conclusion for $3 \times 3 \times 4$:

  - weight 8 is always possible

- $3 \times 3 \times 3$: Simple form in Tendeiro, ten Berge, and Choulakian (2013), shown to work almost surely:

$$\left[\begin{array}{ccc|ccc|ccc} h & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & h & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & h \end{array}\right]$$

Weight 9 always possible (almost surely).

- $K_{\text{max}} = 10$, so $K = 1, 2, \cdots, 8, 9, 10$

Summarizing, simple forms worked out for:

- $4 \times 4 \times 10$: K_{max} situation
- $4 \times 4 \times 1$: Use EVD (weight 4)
- $4 \times 4 \times 9 =$ orthogonal complement of $4 \times 4 \times 1$
  - Weights 16, 7, or 18.
- $4 \times 4 \times 2$: See $\text{EVD}(\mathbf{X}_1^{-1}\mathbf{X}_2)$
  - real eigenvalues: weight 6
  - one pair of complex eigenvalues: weight 7
  - two pairs of complex eigenvalues: weight 8
- $4 \times 4 \times 8 =$ orthogonal complement of $4 \times 4 \times 2$
  - any symmetric slice $4 \times 4 \times 8$ array can almost surely be simplified into one out of two weight 18 arrays

Question: Can simpler targets be found for the cases previously presented?

Answer:

- $3 \times 3 \times K$ for $K = 1, 2, 4, 5, 6$: NO (proved)
- $4 \times 4 \times K$ for $K = 8, 9$: NO(?) (simulation)

- $\underline{\mathbf{X}}$: symmetric slice $3 \times 3 \times 4$ array
- ten Berge et al. (2004)

$$\boxed{\text{typical rank } (\underline{\mathbf{X}}) = \{4, 5\}}$$

- rank=4?, rank=5?
  Check if roots of a certain fourth degree polynomial are real and distinct.

Using ▸ $3 \times 3 \times 4$ simple form we concluded that:

- rank $(\underline{\mathbf{X}})$=4 iif $\delta = 1$ and $\alpha > 0$ (and rank is 5 otherwise)
- a CP decomposition is now straightforward

Example: rank=4

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & \sqrt{\alpha} & -\sqrt{\alpha} \\ \sqrt{\alpha} & -\sqrt{\alpha} & 0 & 0 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5\sqrt{\alpha - 1} & -0.5\sqrt{\alpha - 1} \\ 0.5\sqrt{\alpha - 1} & -0.5\sqrt{\alpha - 1} & 0 & 0 \end{bmatrix}$$

- Simplification achieved for some types of arrays with symmetric frontal slices; closed form rotation matrices available.
- Maximal simplicity achieved (mathematically proved or empirically verified via SIMPLIMAX).
- Typical rank considerations come as nice follow-ups.

Considerations:

- 3MPCA core arrays are not "randomly sampled from a continuous distribution", but do behave as such.
- Valid contribution for Matrix Theory: Simultaneous reduction of more than a pair of matrices to sparse forms is scarce.

Developments:

- Extend results and procedures to other orders.
- Address issues like: Maximal simplicity, typical rank.

Kruskal, Harshman, Lundy (1983, 1985):

$$\underline{\mathbf{X}} = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \end{array}\right]$$

Random starts of CP with $r = 2$ components invariably give $f = 2$.

It must be the global minimum.

No!

ten Berge, Kiers, De Leeuw (1988)

$$\inf(f) = 1$$

It must be LOCAL minima.

No!

What we found

All solutions with $f = 2$ are NOT minima (not even local!).

How can you reach such a conclusion?

Goal

Given a scalar function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, how to find extremes (minima, maxima)?

Three types of points to consider:

1. Points in the border of the domain of f;
   Example: $f(x) = x^3$, x between -1 and 1.
2. Points where f is not twice continuously differentiable;
   Example: $f(x) = |x|$, for $x = 0$.
3. **Points where f is twice continuously differentiable.**

$\uparrow$

our goal

How to optimize $f : \mathbb{R}^n \longrightarrow \mathbb{R}$:

Step 1 Compute partial derivatives = 1st order derivatives for each variable, while the others are "constant".
Find stationary points (SPs) by solving the system of equations

$$f_i := \frac{\partial f(x_1, \ldots, x_n)}{\partial x_i} = 0, \quad i = 1, \ldots, n.$$

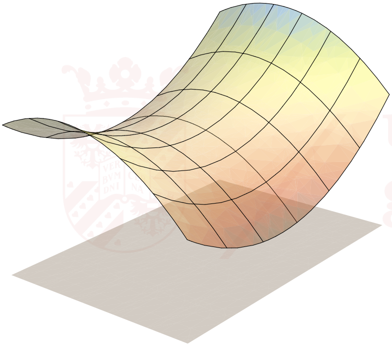
Step 2 Analyze 2nd order derivatives = eigenvalues of the **Hessian** matrix:

$$\mathrm{Hess} = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{bmatrix}$$

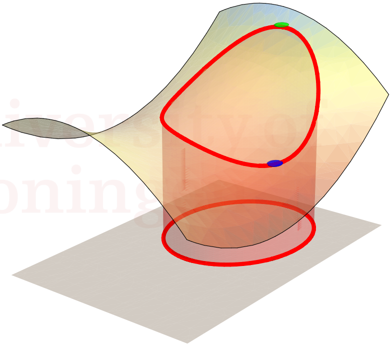
Decision rule:

- Hess is positive definite $\implies$ SP is minimum
- Hess is negative definite $\implies$ SP is maximum
- Hess is indefinite $\implies$ SP is saddle point

Method of Lagrange multipliers
Useful to find maxima/minima of a function subject to constraints.



Unconstrained.
No minimum, no maximum.



Constrained to red points.
Minimum, Maximum.

S: $I \times I \times K$ array with $I \times I$ symmetric frontal slices

- Carroll and Chang (1970) suggested to use CP ... to fit the INDSCAL ... model:
$$\mathbf{S}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}' + \mathbf{E}_k,$$
and then "hope" that **A** is columnwise proportional to **B** (**A** and **B** equivalent).

- **A** and **B** seem equivalent in practical applications. However, contrived counterexamples do exist.

Result: $\mathbf{A} \neq \mathbf{B}$ is possible at global minima if slices are indefinite (ten Berge and Kiers, 1991).

Example:

$$\underline{\mathbf{S}} = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 2 \\ 0 & -1 & 0 & 0 & -2 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \end{array}\right]$$

$$\mathbf{A}^* = \begin{bmatrix} \sqrt{1/3} & \sqrt{0.5} \\ -\sqrt{1/3} & 0 \\ \sqrt{1/3} & -\sqrt{0.5} \end{bmatrix}, \mathbf{B}^* = \begin{bmatrix} \sqrt{1/3} & \sqrt{0.5} \\ \sqrt{1/3} & 0 \\ \sqrt{1/3} & -\sqrt{0.5} \end{bmatrix}, \mathbf{C}^* = \begin{bmatrix} 2 & 2 \\ 0 & -2 \end{bmatrix}.$$

This solution minimizes CP's loss function:

$$f_{\mathrm{CP}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \|\mathbf{S}_1 - \mathbf{A}\mathbf{C}_1\mathbf{B}'\|^2 + \|\mathbf{S}_2 - \mathbf{A}\mathbf{C}_2\mathbf{B}'\|^2 \geqslant 5$$

and

$$f_{\mathrm{CP}}(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*) \equiv 5$$

Result: $\mathbf{A} \neq \mathbf{B}$ is only possible at non-optimal points if slices are non-negative definite (ten Berge and Kiers, 1991).

Example:

$$\underline{\mathbf{S}} = \left[\begin{array}{ccc|ccc} 3 & 1 & 0 & 3 & -1 & 0 \\ 1 & 3 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\mathbf{A}^* = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{B}^* = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{C}^* = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This solution does not minimize CP's loss function:

$$f_{\mathrm{CP}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \|\mathbf{S}_1 - \mathbf{A}\mathbf{C}_1\mathbf{B}'\|^2 + \|\mathbf{S}_2 - \mathbf{A}\mathbf{C}_2\mathbf{B}'\|^2 \geqslant 21$$

but

$$f_{\mathrm{CP}}(\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*) \equiv 39$$

These points are, in fact, saddle points (Bennani Dosse and ten Berge, 2008).

What happens for $R > 1$?

Question:
What is the general situation when $R > 1$?

Approach to find an answer:

- Use simulation (run CP lots of times).
- Analyze the first and second-order differential structures of the loss function of CP

$$f_{\text{CP}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{A}\mathbf{C}_k\mathbf{B}'\|^2$$

But how to do this? Number of variables is too big.
Example: $2 \times 2 \times 2$ array, $R = 2$ components

$$\boxed{f_{\text{CP}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \text{ has 12 variables}}$$

$$f_{\mathrm{CP}}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{A}\mathbf{C}_k\mathbf{B}'\|^2 \tag{1}$$

Procedure:

- Parameter elimination:
  Express **C** as a function of **A** and **B** (valid at stationary points):

  $$\text{row } i \text{ of } \mathbf{C} = (\mathbf{A}'\mathbf{A} * \mathbf{B}'\mathbf{B})^{-1}\mathrm{diag}(\mathbf{A}'\mathbf{X}_i\mathbf{B})$$

- Simplify target function (1).
- Use matrix differential calculus: The variables to differentiate for are *matrices* **A**, **B**.
- Constrain **A**,**B**:
  - Columns of unit length (identification constraint).
  - Orthonormal.

Apply the same procedure to INDSCAL's loss function:

$$f_{\mathrm{IND}}(\mathbf{A}, \mathbf{C}) = \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{A}\mathbf{C}_k\mathbf{A}'\|^2$$

What was done – for both f_{CP} and f_{IND}:

- Jacobian and Hessian matrices computed in closed form.
- Second-order sufficient condition is now available to label SPs.

SVD-approach (ten Berge, 1988)

- INDSCAL model under orthogonality constraints
- Claim: The algorithm sometimes stops at local optima

But saddle points are possible (for contrived examples).

Example:

$$\underline{\mathbf{S}} = \left[\begin{array}{ccc|ccc} 3 & 1 & 0 & 3 & -1 & 0 \\ 1 & 3 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Global minimum ($f = 1$)

$$\mathbf{A}^* = \begin{bmatrix} \sqrt{0.5} & -\sqrt{0.5} \\ \sqrt{0.5} & \sqrt{0.5} \\ 0 & 0 \end{bmatrix}, \mathbf{C}^* = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}.$$

Non-optimal SPs: Saddle points

$$\mathbf{A}^* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{C}^* = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \qquad f = 5$$

$$\mathbf{A}^* = \begin{bmatrix} \sqrt{0.5} & 0 \\ \sqrt{0.5} & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{C}^* = \begin{bmatrix} 4 & 0 \\ 2 & 1 \end{bmatrix} \qquad f = 20$$

$$\mathbf{A}^* = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{C}^* = \begin{bmatrix} 0 & 3 \\ 1 & 3 \end{bmatrix} \qquad f = 22$$

What happens for randomly generated data?

Simulation study

- 150 $3 \times 3 \times 3$ symmetric slice arrays with Gramian slices
- SVD-approach with $R = 2$ components;
  10 different initializations per array
- The second-order differential structure was analysed in each case.

Results

- Saddle points did not occur: There are no indications that the SVD-approach converges to saddle points for randomly generated data.
- Local optima occurred for $\sim 8\%$ of the arrays.

- Eleven cases considered.
  ($R > 1$ component, arrays with symmetric 3×3 slices)
- Two types of arrays: Gramian *vs* non-Gramian slices.
- 100 runs per array.

Results:

- $\mathbf{A} \neq \mathbf{B}$ did not occur for Gramian slices.
- $\mathbf{A} \neq \mathbf{B}$ did occur for indefinite slices only in "sick" cases (degenerate).
- Saddle points happen rarely.

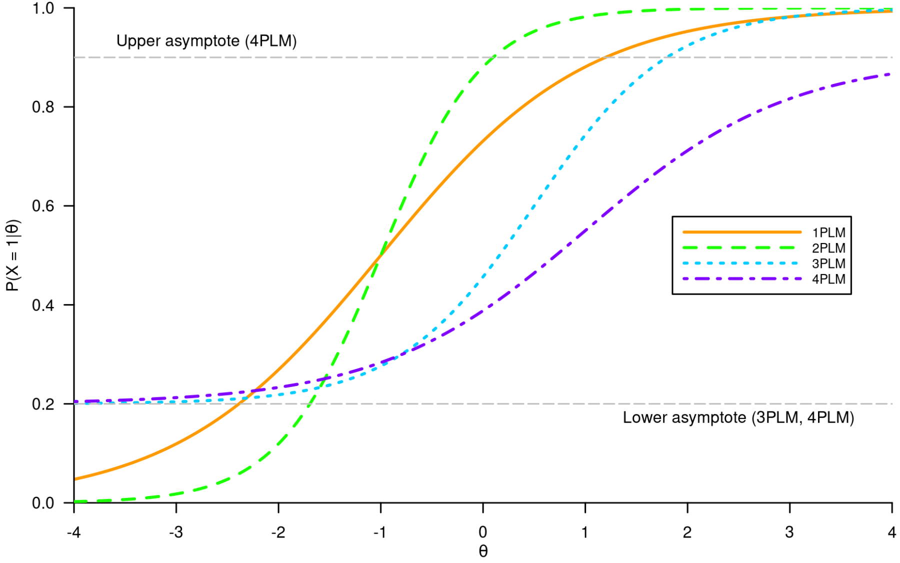
- Loss functions of CP and INDSCAL were transformed into "simpler" optimization functions.
- First and second order derivatives were derived.
- These tools allow to identify saddle points.
  If there is a saddle point: Rerun the algorithm!
- Simulations showed that saddle points do not occur frequently, but they do occur with positive probability.

- IRT models are nowadays very popular psychometric tools.
- Advantages over classical test theory (see, e.g., Embretson & Reise, 2000):
  - Scoring examinees and items is (theoretically) sample invariant (up to a linear transformation).
  - Both persons and items placed on the same scale.
  - Precision of measurement varies along the latent trait.
  - There is a multitude of models that apply to varying measurement situations.
  - Adaptive testing easy to implement.
  - ...

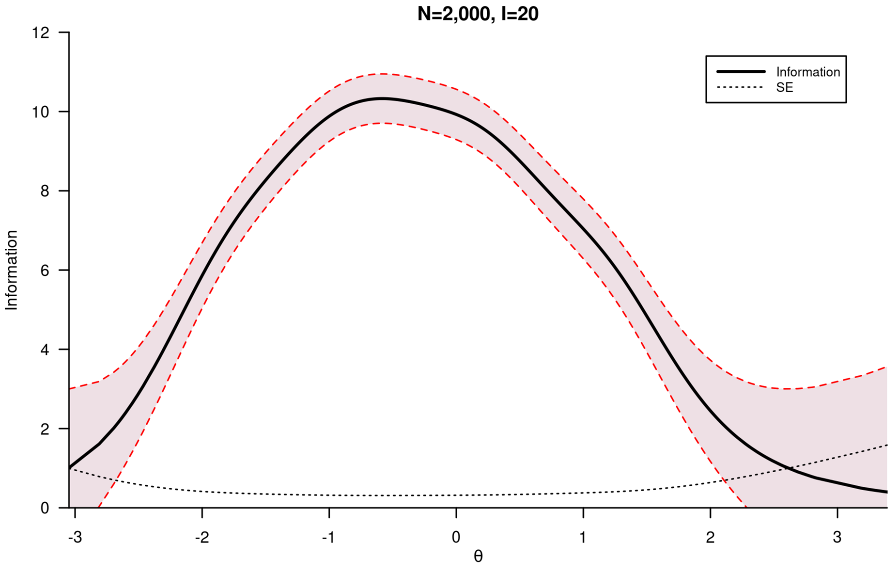
Popular models:

- For dichotomous data: 1PLM (Rasch model), 2PLM, 3PLM, 4PLM.
- For polytomows data: NRM, GRM, PCM, RSM, GGUM

(but much, much more!!)



1.0
0.8
0.6
0.4
0.2
0.0
P(X = 1|θ)
Upper asymptote (4PLM)
Lower asymptote (3PLM, 4PLM)
1PLM
2PLM
3PLM
4PLM
-4
-3
-2
-1
0
1
2
3
4
θ


N=2,000, I=20
Information
SE
Information
θ
12
10
8
6
4
2
0
-3
-2
-1
0
1
2
3

I have explored two topics within IRT, namely person-fit analysis and (recently) unfolding models.

Below I provide a brief account on each topic.

- A lot of effort in IRT is put of assessing model fit
  (e.g., consider the recent IMPS 2017 meeting).
- Various methods/statistics exist, both at the item ("columns" direction) and the scale levels.
- However, person-level misfitting response patterns ("rows direction") are also worrisome:
  - They affect the estimation of person trait.
  - They may also affect the estimation of item parameters.
  - They distort rank ordering of persons (problematic, e.g., in selection settings).

	Easy $\longrightarrow$ Difficult									
Student	It1	It2	It3	It4	It5	It6	It7	It8	It9	It10
A	1	1	1	1	1	1	1	0	0	0
B	1	1	0	1	0	0	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Warm-up??	0	0	0	1	1	1	0	1	0	1
Cheater??	1	1	0	1	0	0	1	1	1	1
Random??	1	0	1	0	0	0	1	0	0	1

- Statistical methods have been developed to identify such aberrant response patterns (see Sijtsma & Meijer, 2001 for an overview).
- I've written various papers on statistical techniques aimed at identifying aberrant response patterns (in the context of CATs, cognitive measurement, personality measurement, using parametric and nonparametric models, etc.).
- I compiled the most used person-fit statistics in an R package (`PerFit`).

Some caveats:

- Power is typically low (due to typically few items available per person).
- Person-fit analysis may assist test forensics.
  However, they can not provide, on their own, conclusive answers.

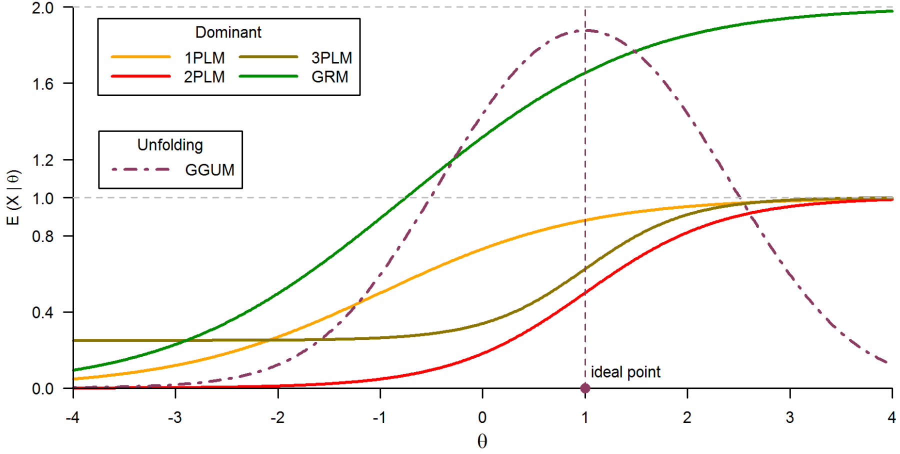
- The mostly used IRT models are cumulative:
  The larger θ, the larger the item's expected score.
  This follows the widely spread Likert's philosophy (early 1930s).
- A different class of models, commonly referred to as unfolding, condition the expected item score on the distance between the item and the person's location on the latent trait.
- This principle was first established by Thurstone in the late 1920s. It is sensible in contexts measuring attitudes or preferences (involving self-introspection).

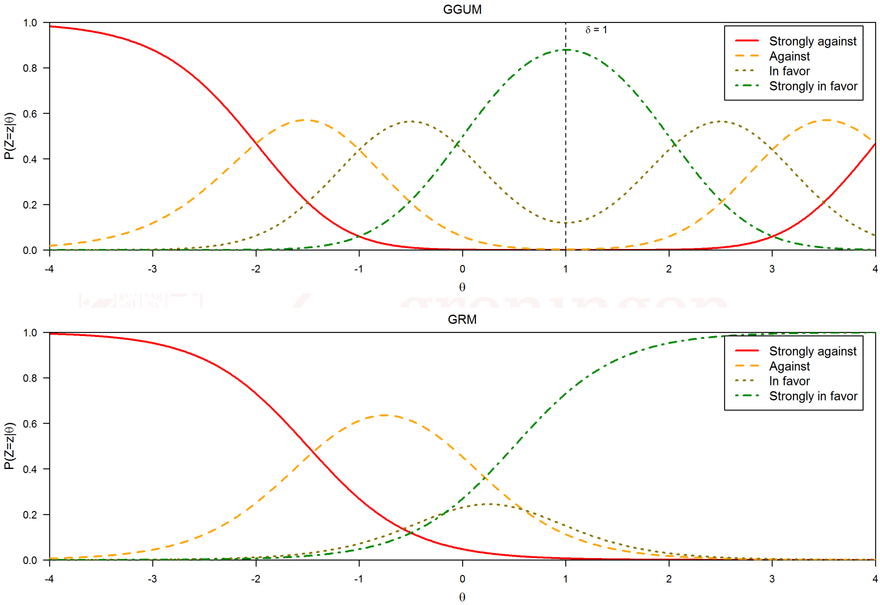
The most popular unfolding model in use nowadays is the GGUM (generalized graded unfolding model; Roberts et al., 1996, 2000):

$$P(Z_i = z|\theta_n) = \frac{f(z) + f(M-z)}{\sum_{w=0}^{C}\left[f(w) + f(M-w)\right]},$$

with

- $f(w) = \exp\left\{\alpha_i\left[w(\theta_n - \delta_i) - \sum_{k=0}^{w}\tau_{ik}\right]\right\},\ w = 0, \ldots, M.$
- $C =$ number of observable response categories
- $M = 2C + 1$
- α_i: Discrimination of item i
- δ_i: Difficulty of item i
- τ_{ik}: Threshold parameters of item i


Dominant
1PLM
3PLM
2PLM
GRM
Unfolding
GGUM
ideal point
E (X | θ)
θ
2.0
1.6
1.2
1.0
0.8
0.4
0.0
-4
-3
-2
-1
0
1
2
3
4


GGUM
δ = 1
Strongly against
Against
In favor
Strongly in favor
P(Z=z|θ)
θ
GRM
Strongly against
Against
In favor
Strongly in favor
P(Z=z|θ)
θ

About the unfolding mechanism:

- What matters the most is the perceived distance between the item's statement and the person's standing.
- Conceivably, more of the latent trait may decrease the probability of endorsement.
- A person may disagree with an item statement's because she either believes too strongly in favor of it ('too far' to the right) or against it ('too far' to the left).

Altogether, this conceptualization of item endorsement is fundamentally distinct from the much more common dominant process.

My current work with the GGUM includes:

- Extending person-fit strategies to this type of model.
- Explore empirical applications, because literature on this is scarce and inconclusive (ongoing).
- Revamp the classic algorithm (MML) using R (in preparation).
- Explore Bayesian estimation (starting soon).
- ...

ありがとう。

j.n.tendeiro@rug.nl