Simultaneous Parameter Estimation in Exploratory Factor Analysis by Weighted Least Squares

Steffen Unkel and Nickolay T. Trendafilov

Department of Mathematics and Statistics,
The Open University, Milton Keynes, United Kingdom

- $\mathbf{Z} \in \mathbb{R}^{n \times p}$: Given data matrix of p observed (mean-centered and scaled) variables on n cases.

- Exploratory Factor Analysis (EFA) model [Mulaik (1972)]:

$$\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi} \quad ,$$

$\mathbf{F}$: $n \times k$ matrix of common factors ($k \ll p$);
$\mathbf{\Lambda}$: $p \times k$ matrix of loadings with $\text{rank}(\mathbf{\Lambda}) = k$;
$\mathbf{U}$: $n \times p$ matrix of unique factors;
$\mathbf{\Psi}$: $p \times p$ diagonal matrix of uniquenesses.

- $\mathbf{Z} \in \mathbb{R}^{n \times p}$: Given data matrix of p observed (mean-centered and scaled) variables on n cases.

- Exploratory Factor Analysis (EFA) model [Mulaik (1972)]:

$$\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi} \quad ,$$

$\mathbf{F}$: $n \times k$ matrix of common factors ($k \ll p$);
$\mathbf{\Lambda}$: $p \times k$ matrix of loadings with rank$(\mathbf{\Lambda}) = k$;
$\mathbf{U}$: $n \times p$ matrix of unique factors;
$\mathbf{\Psi}$: $p \times p$ diagonal matrix of uniquenesses.

- Assumptions:
$\mathbf{F}$ and $\mathbf{U}$ are mean-centered;
$\mathbf{F}'\mathbf{F} = \mathbf{I}_k, \quad \mathbf{U}'\mathbf{U} = \mathbf{I}_p, \quad \mathbf{U}'\mathbf{F} = \mathbf{0}_{p \times k}.$

- $\mathbf{Z} \in \mathbb{R}^{n \times p}$: Given data matrix of p observed (mean-centered and scaled) variables on n cases.

- Exploratory Factor Analysis (EFA) model [Mulaik (1972)]:

$$\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{U}\mathbf{\Psi} \quad ,$$

  $\mathbf{F}$: $n \times k$ matrix of common factors ($k \ll p$);
  $\mathbf{\Lambda}$: $p \times k$ matrix of loadings with rank($\mathbf{\Lambda}$) $= k$;
  $\mathbf{U}$: $n \times p$ matrix of unique factors;
  $\mathbf{\Psi}$: $p \times p$ diagonal matrix of uniquenesses.

- Assumptions:
  $\mathbf{F}$ and $\mathbf{U}$ are mean-centered;
  $\mathbf{F}'\mathbf{F} = \mathbf{I}_k, \quad \mathbf{U}'\mathbf{U} = \mathbf{I}_p, \quad \mathbf{U}'\mathbf{F} = \mathbf{0}_{p \times k}.$

Factoring the correlation matrix

- Model correlation structure: $\mathbf{R} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2$.
- Factor extraction: Find a pair $\{\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}\}$ which gives the best fit to the sample correlation matrix $\mathbf{C} = \mathbf{Z}'\mathbf{Z}$.
- Least squares goodness-of-fit criterion:

$$\min_{\mathbf{\Lambda},\mathbf{\Psi}} ||\mathbf{C} - \mathbf{R}||_F^2 \quad .$$

Factoring the correlation matrix

- Model correlation structure: $\mathbf{R} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}^2$.
- Factor extraction: Find a pair $\{\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}\}$ which gives the best fit to the sample correlation matrix $\mathbf{C} = \mathbf{Z}'\mathbf{Z}$.
- Least squares goodness-of-fit criterion:

$$\min_{\mathbf{\Lambda},\mathbf{\Psi}} ||\mathbf{C} - \mathbf{R}||_F^2 \quad .$$

- Factor scores can be obtained as a function of $\mathbf{Z}$, $\hat{\mathbf{\Lambda}}$, and $\hat{\mathbf{\Psi}}$.
- Anderson-Rubin (1956) proposed:

$$\hat{\mathbf{F}} = \mathbf{Z}\hat{\mathbf{\Psi}}^{-2}\hat{\mathbf{\Lambda}}\left(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Psi}}^{-2}\mathbf{C}\hat{\mathbf{\Psi}}^{-2}\hat{\mathbf{\Lambda}}\right)^{-\frac{1}{2}} \quad .$$

Factoring the correlation matrix

- Model correlation structure: $\mathbf{R} = \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}^2$.
- Factor extraction: Find a pair $\{\hat{\mathbf{\Lambda}}, \hat{\mathbf{\Psi}}\}$ which gives the best fit to the sample correlation matrix $\mathbf{C} = \mathbf{Z}'\mathbf{Z}$.
- Least squares goodness-of-fit criterion:

$$\min_{\mathbf{\Lambda},\mathbf{\Psi}} ||\mathbf{C} - \mathbf{R}||_F^2 \quad .$$

- Factor scores can be obtained as a function of $\mathbf{Z}$, $\hat{\mathbf{\Lambda}}$, and $\hat{\mathbf{\Psi}}$.
- Anderson-Rubin (1956) proposed:

$$\hat{\mathbf{F}} = \mathbf{Z}\hat{\mathbf{\Psi}}^{-2}\hat{\mathbf{\Lambda}}\left(\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Psi}}^{-2}\mathbf{C}\hat{\mathbf{\Psi}}^{-2}\hat{\mathbf{\Lambda}}\right)^{-\frac{1}{2}} \quad .$$

Factoring the data matrix

- Outliers can heavily influence $\mathbf{C}$ and the parameter estimates.
  - Robust modification of $\mathbf{C}$ [Pison et al. (2003)].
  - Robust techniques working with $\mathbf{Z}$ [Croux et al. (2003)].

Factoring the data matrix

- Outliers can heavily influence **C** and the parameter estimates.
  - Robust modification of **C** [Pison et al. (2003)].
  - Robust techniques working with **Z** [Croux et al. (2003)].
- By factorization of **Z**, estimates for all EFA parameters can be obtained simultaneously.

Factoring the data matrix

- Outliers can heavily influence **C** and the parameter estimates.
  - Robust modification of **C** [Pison et al. (2003)].
  - Robust techniques working with **Z** [Croux et al. (2003)].
- By factorization of **Z**, estimates for all EFA parameters can be obtained simultaneously.

Methods in the EFA literature

- Young (1940), Lawley (1942), Whittle (1952), Jöreskog (1962).
- Chapter in Horst (1965).
- McDonald (1979), [also Etezadi-Amoli and McDonald (1983)].
- De Leeuw (2004), (2003).

Factoring the data matrix

- Outliers can heavily influence **C** and the parameter estimates.
  - Robust modification of **C** [Pison et al. (2003)].
  - Robust techniques working with **Z** [Croux et al. (2003)].
- By factorization of **Z**, estimates for all EFA parameters can be obtained simultaneously.

Methods in the EFA literature

- Young (1940), Lawley (1942), Whittle (1952), Jöreskog (1962).
- Chapter in Horst (1965).
- McDonald (1979), [also Etezadi-Amoli and McDonald (1983)].
- De Leeuw (2004), (2003).

Simultaneous parameter estimation by De Leeuw (2004)

- Unweighted least squares (ULS) fitting by minimizing:

$$f_{ULS} = ||\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda}' - \mathbf{U}\boldsymbol{\Psi}||_F^2$$

$$\text{s.t.} \quad \mathbf{F}'\mathbf{F} = \mathbf{I}_k, \mathbf{U}'\mathbf{U} = \mathbf{I}_p, \mathbf{U}'\mathbf{F} = \mathbf{0}_{p\times k}, \text{ and } \boldsymbol{\Psi} \text{ diagonal.}$$

- Optimization algorithm is of an alternating least squares type.

Simultaneous parameter estimation by De Leeuw (2004)

- Unweighted least squares (ULS) fitting by minimizing:

$$f_{ULS} = ||\mathbf{Z} - \mathbf{F}\boldsymbol{\Lambda}' - \mathbf{U}\boldsymbol{\Psi}||_F^2$$

s.t. $\mathbf{F}'\mathbf{F} = \mathbf{I}_k, \mathbf{U}'\mathbf{U} = \mathbf{I}_p, \mathbf{U}'\mathbf{F} = \mathbf{0}_{p\times k}$, and $\boldsymbol{\Psi}$ diagonal.

- Optimization algorithm is of an alternating least squares type.

Simultaneous parameter estimation by De Leeuw (2004)

- Defining block matrices $\mathbf{B} := [\mathbf{F} : \mathbf{U}]$ and $\mathbf{A} := [\mathbf{\Lambda} : \mathbf{\Psi}]$ of dimensions $n \times (p+k)$ and $p \times (p+k)$, respectively.

$$f_{ULS} = ||\mathbf{Z} - \mathbf{B}\mathbf{A}'||_F^2 = ||\mathbf{Z}||_F^2 + \text{trace}(\mathbf{A}\mathbf{A}') - 2\,\text{trace}(\mathbf{B}'\mathbf{Z}\mathbf{A}) \quad .$$

- As with the standard Procrustes problem, the minimization of f_{ULS} is equivalent to the maximization of trace$(\mathbf{B}'\mathbf{Z}\mathbf{A})$.

- For this problem an analytical solution via the Singular Value Decomposition of $\mathbf{ZA}$ exists.

Simultaneous parameter estimation by De Leeuw (2004)

- Defining block matrices $\mathbf{B} := [\mathbf{F} : \mathbf{U}]$ and $\mathbf{A} := [\mathbf{\Lambda} : \mathbf{\Psi}]$ of dimensions $n \times (p+k)$ and $p \times (p+k)$, respectively.

$$f_{ULS} = ||\mathbf{Z} - \mathbf{B}\mathbf{A}'||_F^2 = ||\mathbf{Z}||_F^2 + \text{trace}(\mathbf{A}\mathbf{A}') - 2\,\text{trace}(\mathbf{B}'\mathbf{Z}\mathbf{A}) \;.$$

- As with the standard Procrustes problem, the minimization of f_{ULS} is equivalent to the maximization of trace$(\mathbf{B}'\mathbf{Z}\mathbf{A})$.
- For this problem an analytical solution via the Singular Value Decomposition of $\mathbf{ZA}$ exists.
- Updating: $\mathbf{\Lambda} = \mathbf{Z}'\mathbf{F}$ and $\mathbf{\Psi} = \text{diag}(\mathbf{U}'\mathbf{Z})$.

Simultaneous parameter estimation by De Leeuw (2004)

- Defining block matrices $\mathbf{B} := [\mathbf{F} : \mathbf{U}]$ and $\mathbf{A} := [\mathbf{\Lambda} : \mathbf{\Psi}]$ of dimensions $n \times (p+k)$ and $p \times (p+k)$, respectively.

$$f_{ULS} \quad = \quad ||\mathbf{Z} - \mathbf{B}\mathbf{A}'||_F^2 = ||\mathbf{Z}||_F^2 + \text{trace}(\mathbf{A}\mathbf{A}') - 2\,\text{trace}(\mathbf{B}'\mathbf{Z}\mathbf{A}) \quad .$$

- As with the standard Procrustes problem, the minimization of f_{ULS} is equivalent to the maximization of $\text{trace}(\mathbf{B}'\mathbf{Z}\mathbf{A})$.
- For this problem an analytical solution via the Singular Value Decomposition of $\mathbf{ZA}$ exists.
- Updating: $\mathbf{\Lambda} = \mathbf{Z}'\mathbf{F}$ and $\mathbf{\Psi} = \text{diag}(\mathbf{U}'\mathbf{Z})$.

Data matrix decomposition by Weighted Least Squares

- Weighted Least Squares (WLS) loss function:

$$f_{WLS} = ||(\mathbf{Z} - \mathbf{BA}') \odot \mathbf{W}||_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij}^2 e_{ij}^2 \ ,$$

W: $n \times p$ matrix of non-negative weights w_{ij} for residuals e_{ij};
$\odot$ denotes the Hadamard product.

- Optimization of f_{WLS} by an iterative majorization algorithm.

Data matrix decomposition by Weighted Least Squares

- Weighted Least Squares (WLS) loss function:

$$f_{WLS} = ||(\mathbf{Z} - \mathbf{BA}') \odot \mathbf{W}||_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij}^2 e_{ij}^2 \quad ,$$

  $\mathbf{W}$: $n \times p$ matrix of non-negative weights w_{ij} for residuals e_{ij};
  $\odot$ denotes the Hadamard product.

- Optimization of f_{WLS} by an iterative majorization algorithm.

Iterative Majorization by Kiers (1997), (2002)

- f_{WLS} is majorized and touched by

$$\mu(\mathbf{B},\mathbf{A}|\mathbf{B}^c,\mathbf{A}^c) = \text{constant} + w_m^2||\tilde{\mathbf{Z}} - \mathbf{BA}'||_F^2 \quad ,$$

$$\tilde{\mathbf{Z}} = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W}\odot\mathbf{W}\odot(\mathbf{Z}-\mathbf{B}^c\mathbf{A}^{c'})) \quad ,$$

w_m^2: maximum of the squared elements of $\mathbf{W}$.

- WLS fitting problem is solved by iteratively minimizing $||\tilde{\mathbf{Z}} - \mathbf{BA}'||_F^2$.

Iterative Majorization by Kiers (1997), (2002)

- f_{WLS} is majorized and touched by

$$\mu(\mathbf{B}, \mathbf{A}|\mathbf{B}^c, \mathbf{A}^c) = \text{constant} + w_m^2 ||\tilde{\mathbf{Z}} - \mathbf{B}\mathbf{A}'||_F^2 \quad ,$$

$$\tilde{\mathbf{Z}} = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W} \odot \mathbf{W} \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'})) \quad ,$$

w_m^2: maximum of the squared elements of **W**.

- WLS fitting problem is solved by iteratively minimizing $||\tilde{\mathbf{Z}} - \mathbf{B}\mathbf{A}'||_F^2$.

Iterative Majorization by Kiers (1997), (2002)

- f_{WLS} is majorized and touched by

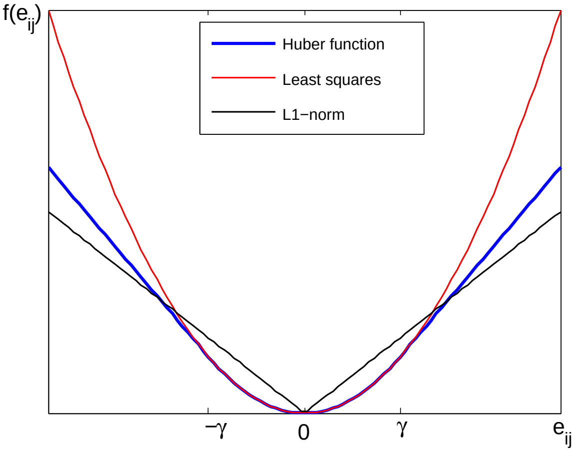
$$\mu(\mathbf{B}, \mathbf{A}|\mathbf{B}^c, \mathbf{A}^c) = \text{constant} + w_m^2||\tilde{\mathbf{Z}} - \mathbf{B}\mathbf{A}'||_F^2 \quad ,$$

$$\tilde{\mathbf{Z}} = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W} \odot \mathbf{W} \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'})) \quad ,$$

w_m^2: maximum of the squared elements of $\mathbf{W}$.

- WLS fitting problem is solved by iteratively minimizing $||\tilde{\mathbf{Z}} - \mathbf{B}\mathbf{A}'||_F^2$.

Hybrid ℓ_1-ℓ_2 goodness-of-fit criterion by Huber (1964)



- Huber's loss function:

$$f_H(\mathbf{Z} - \mathbf{BA}') = \sum_{i=1}^{n} \sum_{j=1}^{p} f_H(e_{ij}) \ ,$$

where $$f_H(e_{ij}) = \begin{cases} e_{ij}^2 & \textit{if} \quad |e_{ij}| < \gamma \ , \\ 2\gamma|e_{ij}| - \gamma^2 & \textit{if} \quad |e_{ij}| \geq \gamma \ , \end{cases}$$

and γ is a given 'tuning constant'.

- Huber criterion formulated in WLS form:

$$f_H = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij}^2 e_{ij}^2 \ ,$$

where $$w_{ij}^2 = \begin{cases} 1 & \textit{if} \quad |e_{ij}| < \gamma \ , \\ 2\gamma/|e_{ij}| - \gamma^2/e_{ij}^2 & \textit{if} \quad |e_{ij}| \geq \gamma \ . \end{cases}$$

- Huber's loss function:

$$f_H(\mathbf{Z} - \mathbf{BA}') = \sum_{i=1}^{n} \sum_{j=1}^{p} f_H(e_{ij}) \ ,$$

where $$f_H(e_{ij}) = \begin{cases} e_{ij}^2 & \textit{if} \quad |e_{ij}| < \gamma \ , \\ 2\gamma|e_{ij}| - \gamma^2 & \textit{if} \quad |e_{ij}| \geq \gamma \ , \end{cases}$$

and γ is a given 'tuning constant'.

- Huber criterion formulated in WLS form:

$$f_H = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij}^2 e_{ij}^2 \ ,$$

where $$w_{ij}^2 = \begin{cases} 1 & \textit{if} \quad |e_{ij}| < \gamma \ , \\ 2\gamma/|e_{ij}| - \gamma^2/e_{ij}^2 & \textit{if} \quad |e_{ij}| \geq \gamma \ . \end{cases}$$

- Iterative Re-weighted least squares algorithm using the iterative majorization approach.

- Huber function can be majorized by a constant plus the majorizing function $\mu(\mathbf{B}, \mathbf{A}|\mathbf{B}^c, \mathbf{A}^c)$ if...

- Iterative Re-weighted least squares algorithm using the iterative majorization approach.

- Huber function can be majorized by a constant plus the majorizing function $\mu(\mathbf{B}, \mathbf{A}|\mathbf{B}^c, \mathbf{A}^c)$ if...

- ...the weights w_{ij} are computed as:

$$w_{ij} = \begin{cases} 1 & \text{if} \quad |e_{ij}^c| < \gamma \ , \\ \sqrt{\gamma / |e_{ij}^c|} & \text{if} \quad |e_{ij}^c| \geq \gamma \ , \end{cases}$$

e_{ij}^c: estimates of the residuals from the previous iteration.

- Iterative Re-weighted least squares algorithm using the iterative majorization approach.
- Huber function can be majorized by a constant plus the majorizing function $\mu(\mathbf{B}, \mathbf{A}|\mathbf{B}^c, \mathbf{A}^c)$ if...
- ...the weights w_{ij} are computed as:

$$w_{ij} = \begin{cases} 1 & if \quad |e_{ij}^c| < \gamma \ , \\ \sqrt{\gamma/|e_{ij}^c|} & if \quad |e_{ij}^c| \geq \gamma \ , \end{cases}$$

e_{ij}^c: estimates of the residuals from the previous iteration.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.
3. Compute $\tilde{\mathbf{Z}}^c = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W}^c \odot \mathbf{W}^c \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'}))$.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.
3. Compute $\tilde{\mathbf{Z}}^c = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W}^c \odot \mathbf{W}^c \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'}))$.
4. Find $\mathbf{B}^{c+1}$ which maximizes trace$(\mathbf{B}'\tilde{\mathbf{Z}}\mathbf{A}^c)$.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.
3. Compute $\tilde{\mathbf{Z}}^{\mathrm{c}} = \mathbf{B}^{\mathrm{c}}\mathbf{A}^{\mathrm{c}'} + w_m^{-2}(\mathbf{W}^{\mathrm{c}} \odot \mathbf{W}^{\mathrm{c}} \odot (\mathbf{Z} - \mathbf{B}^{\mathrm{c}}\mathbf{A}^{\mathrm{c}'}))$.
4. Find $\mathbf{B}^{\mathrm{c}+1}$ which maximizes $\mathrm{trace}(\mathbf{B}'\tilde{\mathbf{Z}}\mathbf{A}^{\mathrm{c}})$.
5. Update $\boldsymbol{\Lambda}^{\mathrm{c}+1} = \tilde{\mathbf{Z}}'\mathbf{F}^{\mathrm{c}+1}$ and $\boldsymbol{\Psi}^{\mathrm{c}+1} = \mathrm{diag}(\mathbf{U}^{\mathrm{c}+1'}\tilde{\mathbf{Z}})$.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.
3. Compute $\tilde{\mathbf{Z}}^c = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W}^c \odot \mathbf{W}^c \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'}))$.
4. Find $\mathbf{B}^{c+1}$ which maximizes trace$(\mathbf{B}'\tilde{\mathbf{Z}}\mathbf{A}^c)$.
5. Update $\mathbf{\Lambda}^{c+1} = \tilde{\mathbf{Z}}'\mathbf{F}^{c+1}$ and $\mathbf{\Psi}^{c+1} = \text{diag}(\mathbf{U}^{c+1'}\tilde{\mathbf{Z}})$.
6. Compute new weights $\mathbf{W}^{c+1}$.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.
3. Compute $\tilde{\mathbf{Z}}^c = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W}^c \odot \mathbf{W}^c \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'}))$.
4. Find $\mathbf{B}^{c+1}$ which maximizes trace$(\mathbf{B}'\tilde{\mathbf{Z}}\mathbf{A}^c)$.
5. Update $\mathbf{\Lambda}^{c+1} = \tilde{\mathbf{Z}}'\mathbf{F}^{c+1}$ and $\mathbf{\Psi}^{c+1} = \text{diag}(\mathbf{U}^{c+1'}\tilde{\mathbf{Z}})$.
6. Compute new weights $\mathbf{W}^{c+1}$.
7. Compute $f_H(\mathbf{B}^{c+1}, \mathbf{A}^{c+1})$.
   If $|f_H(\mathbf{B}^c, \mathbf{A}^c) - f_H(\mathbf{B}^{c+1}, \mathbf{A}^{c+1})| > \epsilon$, set $c = c + 1$ and go to step 3; else consider the algorithm converged.

1. Initialize $\mathbf{B}^0$ and $\mathbf{A}^0$. Set iteration counter $c = 0$. Set convergence criterion ϵ to some small value.
2. For given γ, compute $\mathbf{W}^0$ and $f_H(\mathbf{B}^0, \mathbf{A}^0)$.
3. Compute $\tilde{\mathbf{Z}}^c = \mathbf{B}^c\mathbf{A}^{c'} + w_m^{-2}(\mathbf{W}^c \odot \mathbf{W}^c \odot (\mathbf{Z} - \mathbf{B}^c\mathbf{A}^{c'}))$.
4. Find $\mathbf{B}^{c+1}$ which maximizes $\text{trace}(\mathbf{B}'\tilde{\mathbf{Z}}\mathbf{A}^c)$.
5. Update $\mathbf{\Lambda}^{c+1} = \tilde{\mathbf{Z}}'\mathbf{F}^{c+1}$ and $\mathbf{\Psi}^{c+1} = \text{diag}(\mathbf{U}^{c+1'}\tilde{\mathbf{Z}})$.
6. Compute new weights $\mathbf{W}^{c+1}$.
7. Compute $f_H(\mathbf{B}^{c+1}, \mathbf{A}^{c+1})$.
   If $|f_H(\mathbf{B}^c, \mathbf{A}^c) - f_H(\mathbf{B}^{c+1}, \mathbf{A}^{c+1})| > \epsilon$, set $c = c + 1$ and go to step 3; else consider the algorithm converged.

European health and fertility data [Croux et al. (2003)]

- For 14 European countries, the Soviet Union, and the European Union in their 1986 configuration, $p = 9$ variables related to health and fertility were measured ($n = 16$).
- Albania and Turkey are outlying in most of the variables.

European health and fertility data [Croux et al. (2003)]

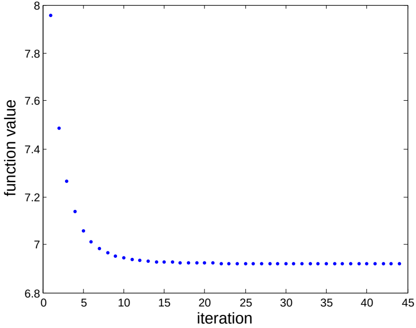
- For 14 European countries, the Soviet Union, and the European Union in their 1986 configuration, $p = 9$ variables related to health and fertility were measured ($n = 16$).
- Albania and Turkey are outlying in most of the variables.
- A two-factor-model is considered in the following.

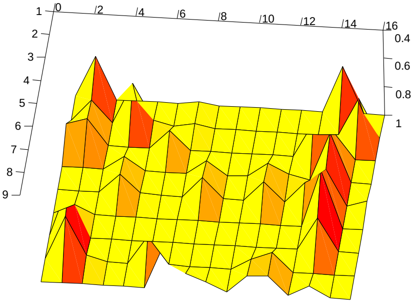
European health and fertility data [Croux et al. (2003)]

- For 14 European countries, the Soviet Union, and the European Union in their 1986 configuration, $p = 9$ variables related to health and fertility were measured ($n = 16$).
- Albania and Turkey are outlying in most of the variables.
- A two-factor-model is considered in the following.

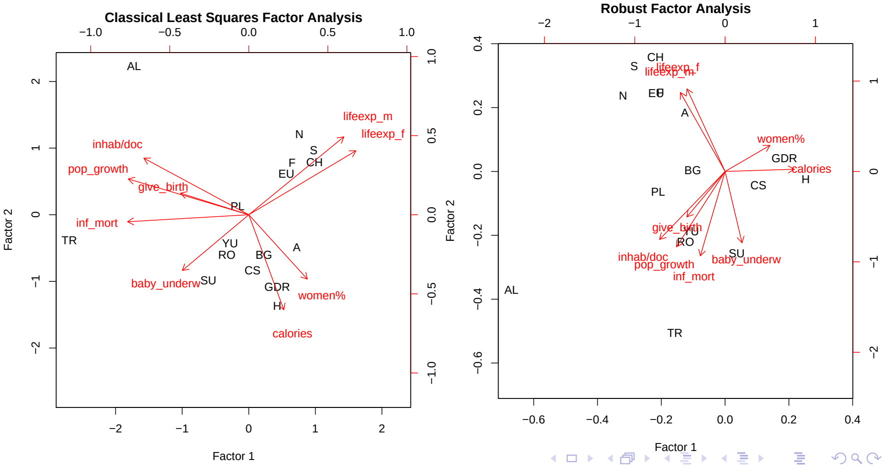
- Huber treshold [Verboon (1994)]: $\gamma = \frac{1}{3}s$, where $s = \text{median}\{|e_{ij}|\} + 4\ \text{MAD}\{|e_{ij}|\}$.

- Convergence criterion: $\epsilon = 10^{-6}$.





Biplots [Gabriel (1971), Gower and Hand (1996)]



Final remarks

- Factor indeterminacy: Factor scores are not uniquely estimable, even for fixed loadings and uniquenesses.
- Other robustifiers can be used such as e.g. the Biweight-function [Beaton and Tukey (1974), Mosteller and Tukey (1977)].

Final remarks

- Factor indeterminacy: Factor scores are not uniquely estimable, even for fixed loadings and uniquenesses.
- Other robustifiers can be used such as e.g. the Biweight-function [Beaton and Tukey (1974), Mosteller and Tukey (1977)].
- Iterative majorization algorithm by Kiers (1997) is a very general approach for minimizing weighted decomposition models.

Final remarks

- Factor indeterminacy: Factor scores are not uniquely estimable, even for fixed loadings and uniquenesses.
- Other robustifiers can be used such as e.g. the Biweight-function [Beaton and Tukey (1974), Mosteller and Tukey (1977)].
- Iterative majorization algorithm by Kiers (1997) is a very general approach for minimizing weighted decomposition models.
- Weighted iterative majorization algorithm [Groenen et al. (2003), (2005)]: Optimization becomes a WLS problem in a diagonal metric instead of an ULS problem.

Final remarks

- Factor indeterminacy: Factor scores are not uniquely estimable, even for fixed loadings and uniquenesses.
- Other robustifiers can be used such as e.g. the Biweight-function [Beaton and Tukey (1974), Mosteller and Tukey (1977)].
- Iterative majorization algorithm by Kiers (1997) is a very general approach for minimizing weighted decomposition models.
- Weighted iterative majorization algorithm [Groenen et al. (2003), (2005)]: Optimization becomes a WLS problem in a diagonal metric instead of an ULS problem.
- Proposed method will be compared with existing robust methods in Factor Analysis.

Final remarks

- Factor indeterminacy: Factor scores are not uniquely estimable, even for fixed loadings and uniquenesses.
- Other robustifiers can be used such as e.g. the Biweight-function [Beaton and Tukey (1974), Mosteller and Tukey (1977)].
- Iterative majorization algorithm by Kiers (1997) is a very general approach for minimizing weighted decomposition models.
- Weighted iterative majorization algorithm [Groenen et al. (2003), (2005)]: Optimization becomes a WLS problem in a diagonal metric instead of an ULS problem.
- Proposed method will be compared with existing robust methods in Factor Analysis.

Final remarks

- Factor indeterminacy: Factor scores are not uniquely estimable, even for fixed loadings and uniquenesses.
- Other robustifiers can be used such as e.g. the Biweight-function [Beaton and Tukey (1974), Mosteller and Tukey (1977)].
- Iterative majorization algorithm by Kiers (1997) is a very general approach for minimizing weighted decomposition models.
- Weighted iterative majorization algorithm [Groenen et al. (2003), (2005)]: Optimization becomes a WLS problem in a diagonal metric instead of an ULS problem.
- Proposed method will be compared with existing robust methods in Factor Analysis.