Sparse Principal Component Analysis for multiblocks data and its extension to Sparse Multiple Correspondence Analysis

Anne Bernard[1,5], Hervé Abdi[2], Arthur Tenenhaus[3], Christiane Guinot[4], Gilbert Saporta[5]

[1] CE.R.I.E.S, Neuilly sur Seine, France
[2] Department of Brain and Behavioral Sciences, The University of Texas at Dallas, Richardson, TX, USA
[3] SUPELEC, Gif-sur-Yvette, France
[4] Université Francois Rabelais, département d'informatique, Tours, France
[5] Laboratoire Cedric, CNAM, Paris

CE.R.I.E.S.

le cnam

June 13, 2013

Background

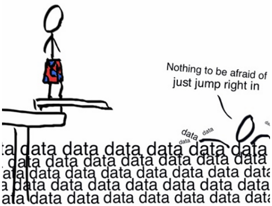
Variable selection and **dimensionality reduction** are necessary in different application domains, and more specifically in genetics.

Quantitative data analysis: PCA
Categorical data analysis: MCA

In case of high dimensional data ($n \ll p$): components (PCs) difficult to interpret.

Background

Variable selection and **dimensionality reduction** are necessary in different application domains, and more specifically in genetics.



Quantitative data analysis: PCA
Categorical data analysis: MCA

In case of high dimensional data ($n \ll p$): components (PCs) difficult to interpret.

A way for obtaining simple structures

Background

Variable selection and **dimensionality reduction** are necessary in different application domains, and more specifically in genetics.

Quantitative data analysis: PCA
Categorical data analysis: MCA

In case of high dimensional data ($n \ll p$): components (PCs) difficult to interpret.

A way for obtaining simple structures

1. **Sparse methods** like sparse PCA for quantitative data: PCs are combinations of few original variables

Background

Variable selection and **dimensionality reduction** are necessary in different application domains, and more specifically in genetics.

Quantitative data analysis: PCA
Categorical data analysis: MCA

In case of high dimensional data ($n \ll p$): components (PCs) difficult to interpret.

A way for obtaining simple structures

1. **Sparse methods** like sparse PCA for quantitative data:
PCs are combinations of few original variables
   $\rightarrow$ extension for multiblocks data structure **(GSPCA)**

Background

Variable selection and **dimensionality reduction** are necessary in different application domains, and more specifically in genetics.

Quantitative data analysis: PCA
Categorical data analysis: MCA

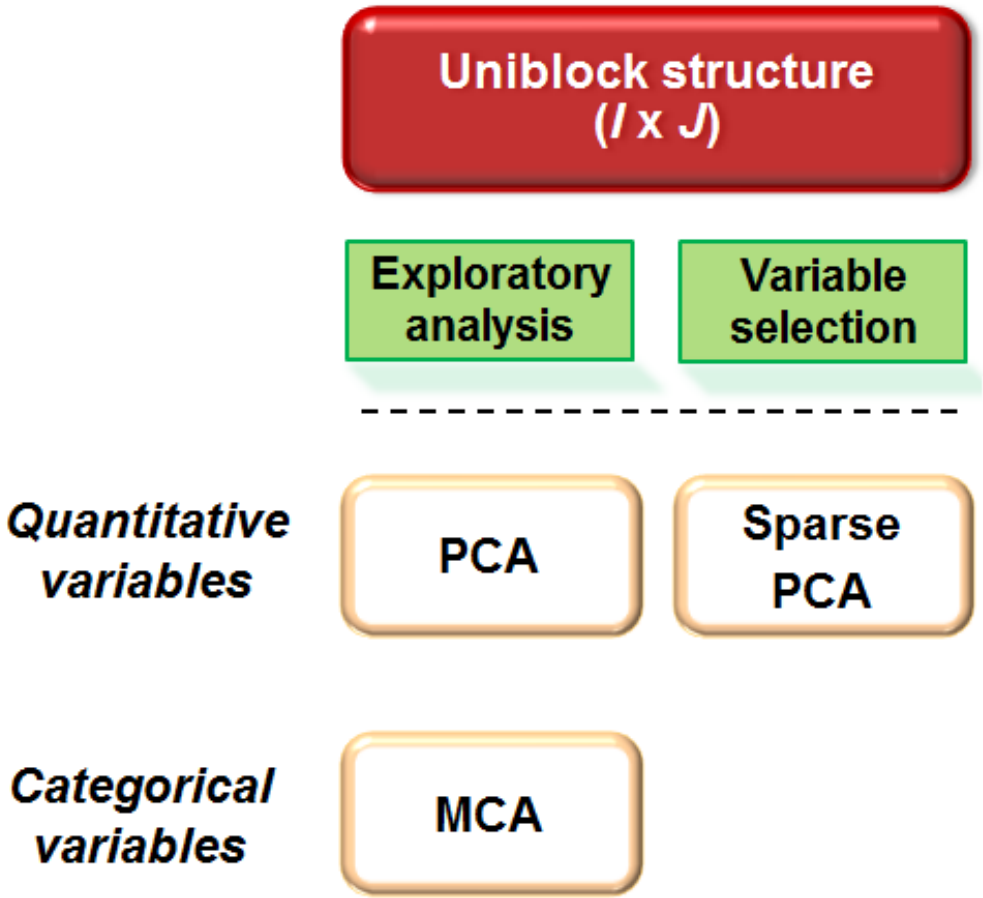
In case of high dimensional data ($n \ll p$): components (PCs) difficult to interpret.

A way for obtaining simple structures

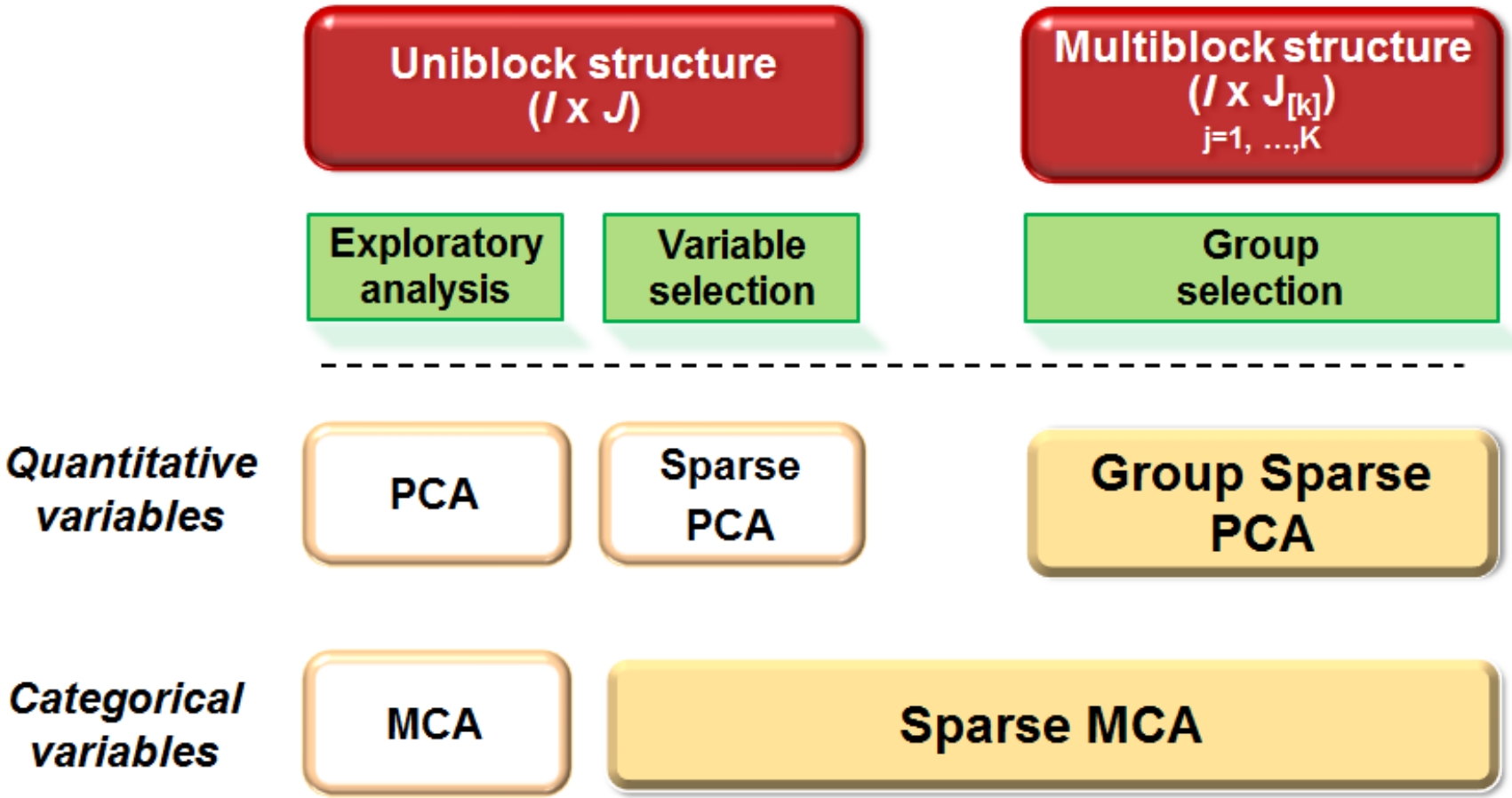
1. **Sparse methods** like sparse PCA for quantitative data:
PCs are combinations of few original variables
$\rightarrow$ extension for multiblocks data structure **(GSPCA)**

2. Extension for categorical variables:
Development of a new sparse method **(sparse MCA)**

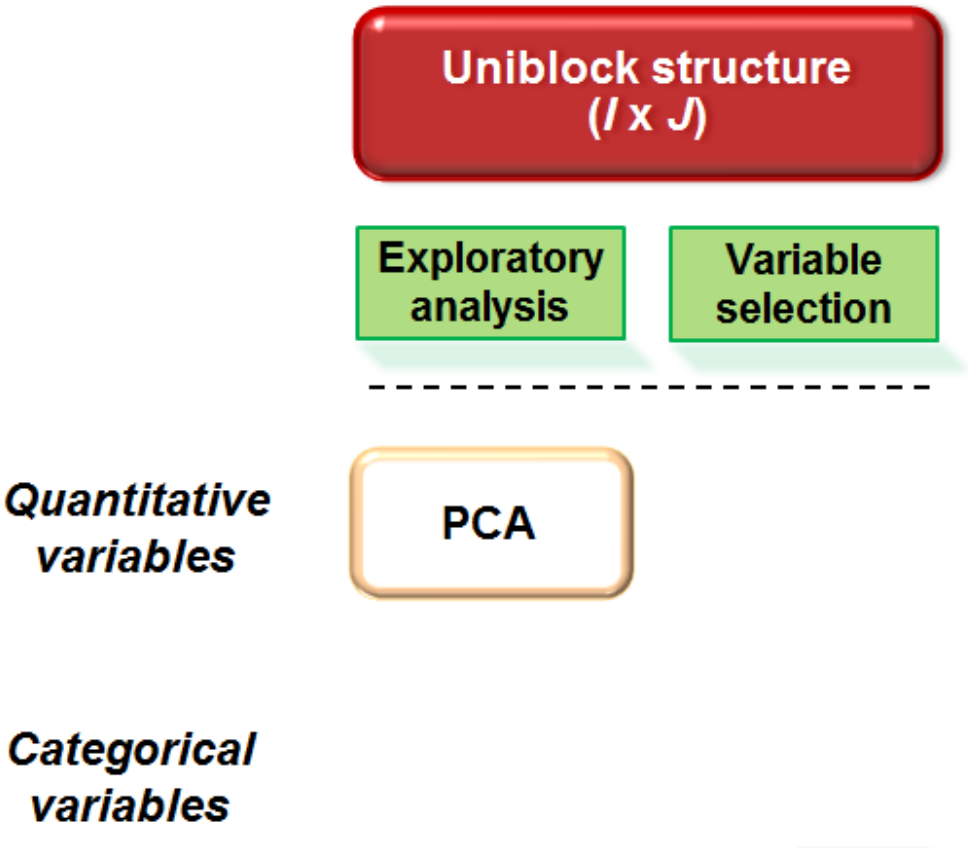
Data and methods



Data and methods

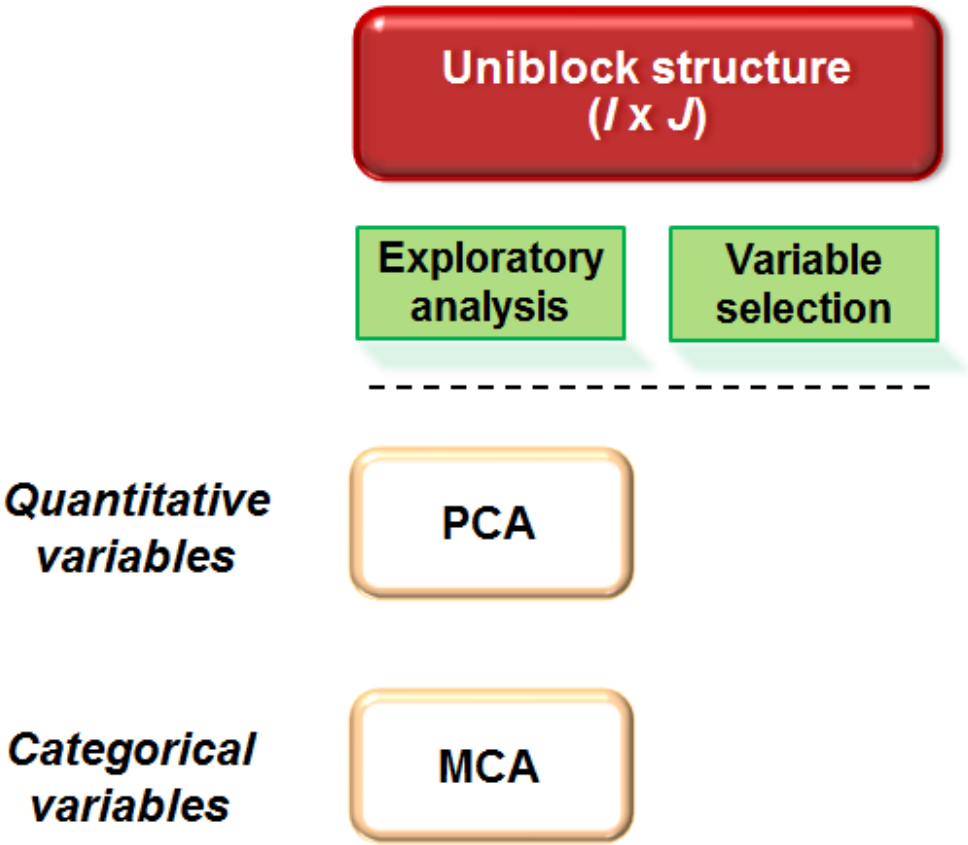


Uniblock data structure



Principal Component Analysis (PCA): Each PC is a linear combination of all the original variables, so loadings are nonzero
$\rightarrow$ PCA results difficult to interpret

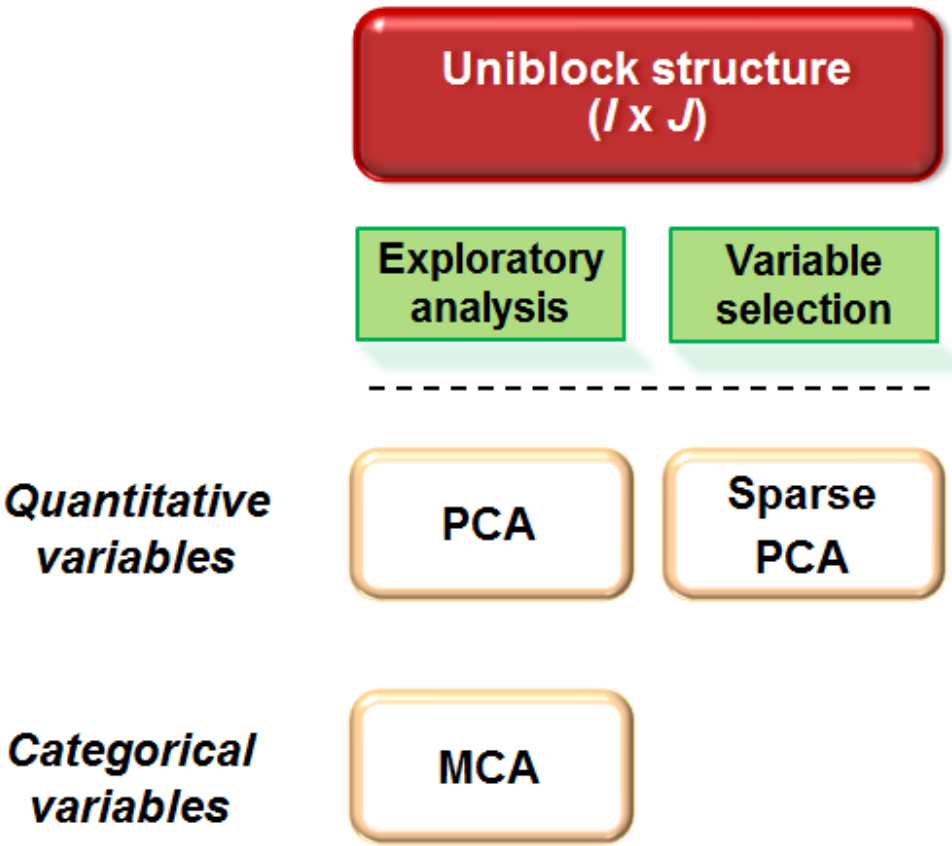
Uniblock data structure



Principal Component Analysis (PCA): Each PC is a linear combination of all the original variables, so loadings are nonzero $\rightarrow$ PCA results difficult to interpret

Multiple Correspondence Analysis (MCA): Counterpart of PCA for categorical data

Uniblock data structure



Principal Component Analysis (PCA): Each PC is a linear combination of all the original variables, so loadings are nonzero $\rightarrow$ PCA results difficult to interpret

Multiple Correspondence Analysis (MCA): Counterpart of PCA for categorical data

Sparse PCA (SPCA): Modified PCs with sparse loadings using regularized SVD

Principal Component Analysis (PCA)

X: matrix of **quantitative variables**

PCA can be computed via the SVD of **X**

Principal Component Analysis (PCA)

X: matrix of **quantitative variables**
PCA can be computed via the SVD of **X**

Singular Value Decomposition of X (SVD):

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \text{ with } \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I} \quad (1)$$

where **Z**=**UD** the PCs and **V** the corresponding **nonzero** loadings
$\rightarrow$ PCA results difficult to interpret

Principal Component Analysis (PCA)

$\mathbf{X}$: matrix of **quantitative variables**
PCA can be computed via the SVD of $\mathbf{X}$

Singular Value Decomposition of $\mathbf{X}$ (SVD):

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \text{ with } \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I} \quad (1)$$

where $\mathbf{Z}$=$\mathbf{UD}$ the PCs and $\mathbf{V}$ the corresponding **nonzero** loadings
$\rightarrow$ PCA results difficult to interpret

SVD as low rank approximation of matrices:

$$\min_{\mathbf{X}^{(1)}} \|\mathbf{X} - \mathbf{X}^{(1)}\|_2^2 = \min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 \quad (2)$$

$\mathbf{X}^{(1)} = \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$ is the best rank-one matrix approximation of $\mathbf{X}$ with:
$\tilde{\mathbf{u}} = \delta_1\mathbf{u}_1$ the first component and $\tilde{\mathbf{v}} = \mathbf{v}_1$ the first loading vector.

Sparse Principal Component Analysis (SPCA)

Challenge: facilitate interpretation of PCA results

How? PCs with a lot of zero loadings (sparse loadings)

Several approaches:

- **ScoTLass** by Jolliffe, I.T. et al. (2003),
- **SPCA** by Zou, H. et al. (2006)
- **SPCA-rSVD** by Shen, H. and Huang, J.Z. (2008)

Principle: use connection of PCA with SVD
→ low rank matrix approximation problem **with regularization penalty on the loadings**

Regularization in regression

Lasso penalty: (Tibshirani, 1996)

- a variable selection technique
- produces sparse models
- Imposes the L_1 norm on the linear regression coefficients

Regression with L_1 penalization

$$\hat{\beta}_{lasso} = \min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \sum_{j=1}^{J} |\beta_j| \qquad \text{with } P_\lambda(\beta) = \lambda \sum_{j=1}^{J} |\beta_j|$$

If λ=0: usual regression (non-zero coefficients)

If λ increases: some coefficients sets to zero

$\Rightarrow$ **selection of variables**

but the number of variables selected is bounded by the number of units

Regularization in regression

Other kinds of penalties:

- **Hard thresholding penalty:** (Antoniadis, 1997)
$$P_\lambda(|\beta|) = \lambda^2 - (|\beta| - \lambda)^2 \mathbf{I}(|\beta| < \lambda)$$
- **SCAD penalty (Smoothly Clipped Absolute Deviation penalty):** (Fan and Li, 2001)

$$P_\lambda(\beta) = \lambda \left(\mathbf{I}(|\beta| \leq \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)\lambda} \mathbf{I}(|\beta| > \lambda) \right) \text{ with } a > 2, \beta > 0$$

- **Group Lasso penalty:** (Yuan and Lin, 2007)
$$P_\lambda(\beta) = \lambda \sum_{k=1}^{K} \sqrt{J_{[k]}} \|\beta_k\|_2 \qquad (3)$$
with $J_{[k]}$ the number of variables in the group k.

Sparse PCA via regularized SVD (SPCA-rSVD)

Low rank matrix approximation problem

$$\min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 \tag{4}$$

Sparse PCA via regularized SVD (SPCA-rSVD)

Low rank matrix approximation problem

$$\min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 \tag{4}$$

with

$$\begin{aligned} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 &= \operatorname{tr}\left((\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}})^T(\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}})\right) \\ &= \|\mathbf{X}\|_2^2 + \operatorname{tr}\left(\tilde{\mathbf{v}}^T\tilde{\mathbf{v}}\right) - 2\operatorname{tr}\left(\mathbf{X}^T\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\right) \\ &= \|\mathbf{X}\|_2^2 + \sum_{j=1}^{J}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{\mathbf{v}}_j\right) \end{aligned} \tag{5}$$

Sparse PCA via regularized SVD (SPCA-rSVD)

Low rank matrix approximation problem
+regularization penalty function applied on $\tilde{\mathbf{v}}$

$$\min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 + P_\lambda(\tilde{\mathbf{v}}) \tag{4}$$

with

$$\begin{aligned} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 &= \operatorname{tr}\left((\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}})^T(\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}})\right) \\ &= \|\mathbf{X}\|_2^2 + \operatorname{tr}\left(\tilde{\mathbf{v}}^T\tilde{\mathbf{v}}\right) - 2\operatorname{tr}\left(\mathbf{X}^T\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\right) \\ &= \|\mathbf{X}\|_2^2 + \sum_{j=1}^{J}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{\mathbf{v}}_j\right) \end{aligned} \tag{5}$$

Sparse PCA via regularized SVD (SPCA-rSVD)

Low rank matrix approximation problem
+regularization penalty function applied on $\tilde{\mathbf{v}}$

$$\min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 + P_\lambda(\tilde{\mathbf{v}}) \tag{4}$$

with

$$\begin{aligned}
\|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 + P_\lambda(\tilde{\mathbf{v}}) &= \operatorname{tr}\left((\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}})^T(\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}})\right) + P_\lambda(\tilde{\mathbf{v}}) \qquad (5)\\
&= \|\mathbf{X}\|_2^2 + \operatorname{tr}\left(\tilde{\mathbf{v}}^T\tilde{\mathbf{v}}\right) - 2\operatorname{tr}\left(\mathbf{X}^T\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\right) + P_\lambda(\tilde{\mathbf{v}})\\
&= \|\mathbf{X}\|_2^2 + \sum_{j=1}^{J}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{\mathbf{v}}_j\right) + \sum_{j=1}^{J} P_\lambda(\tilde{\mathbf{v}}_j)\\
&= \|\mathbf{X}\|_2^2 + \sum_{j=1}^{J}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{\mathbf{v}}_j + P_\lambda(\tilde{\mathbf{v}}_j)\right)
\end{aligned}$$

Sparse PCA via regularized SVD (SPCA-rSVD)

Iterative procedure:

1. $\tilde{\mathbf{v}}$ fixed: minimizer of (4) is $\tilde{\mathbf{u}} = \mathbf{X}\tilde{\mathbf{v}}/\|\mathbf{X}\tilde{\mathbf{v}}\|$
2. $\tilde{\mathbf{u}}$ fixed: minimizer of (4) is $\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}})$ with

$$h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \min_{\tilde{v}_j}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{v}_j + P_\lambda(\tilde{v}_j)\right) \quad (6)$$

For the Lasso penalty: $h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \text{sign}(\mathbf{X}^T\tilde{\mathbf{u}})(|\mathbf{X}^T\tilde{\mathbf{u}}| - \lambda)_+$

For the hard thresholding penalty: $h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \mathbf{I}(|\mathbf{X}^T\tilde{\mathbf{u}}| > \lambda)\mathbf{X}^T\tilde{\mathbf{u}}$

Sparse PCA via regularized SVD (SPCA-rSVD)

Iterative procedure:

1. $\tilde{\mathbf{v}}$ fixed: minimizer of (4) is $\tilde{\mathbf{u}} = \mathbf{X}\tilde{\mathbf{v}}/\|\mathbf{X}\tilde{\mathbf{v}}\|$
2. $\tilde{\mathbf{u}}$ fixed: minimizer of (4) is $\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}})$ with

$$h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \min_{\tilde{v}_j}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{v}_j + P_\lambda(\tilde{v}_j)\right) \quad (6)$$

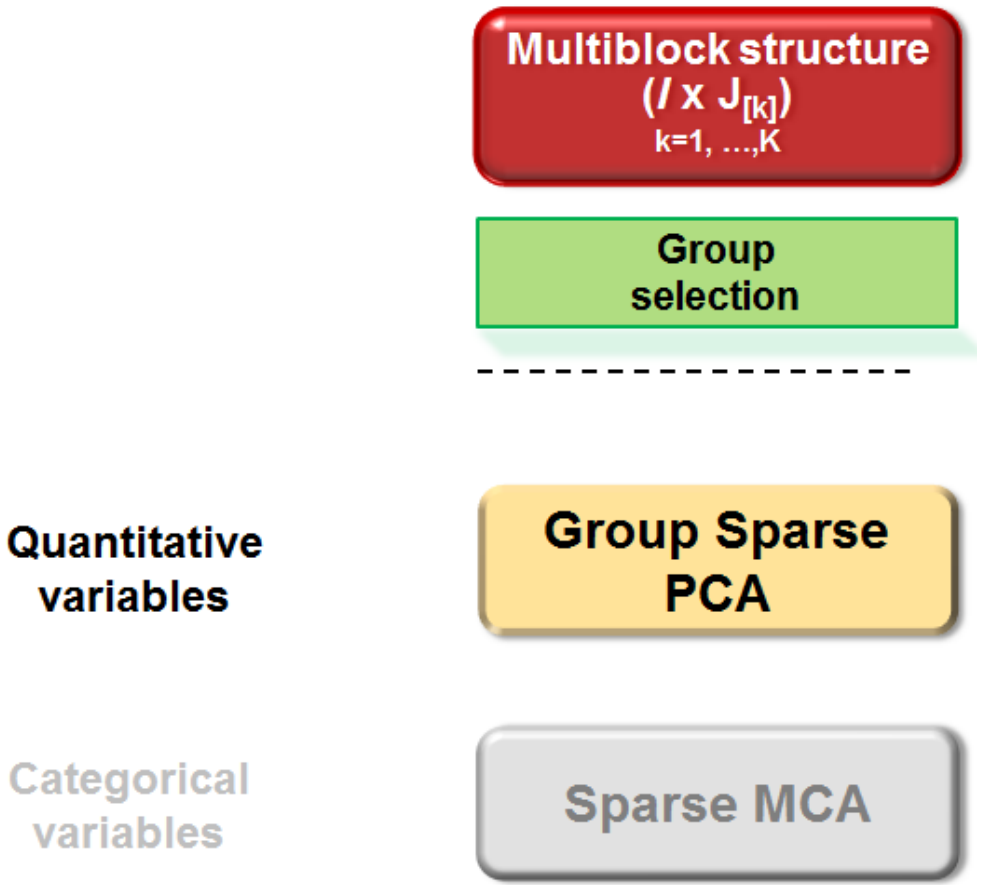
For the Lasso penalty: $h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \text{sign}(\mathbf{X}^T\tilde{\mathbf{u}})(|\mathbf{X}^T\tilde{\mathbf{u}}| - \lambda)_+$

For the hard thresholding penalty: $h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \mathbf{I}(|\mathbf{X}^T\tilde{\mathbf{u}}| > \lambda)\mathbf{X}^T\tilde{\mathbf{u}}$

min obtained by applying h_λ to the vector $\mathbf{X}^T\tilde{\mathbf{u}}$ componentwise $\rightarrow \mathbf{v}_1$ is obtained.

Subsequently sparse loadings $\mathbf{v}_i$ $(i > 1)$ obtained via rank-one approximation of residual matrices.

Multiblock data structure



Group Sparse PCA (GSPCA):

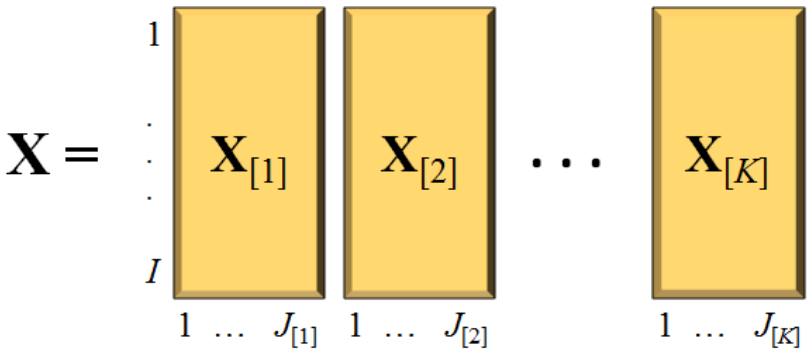
Challenge: To create sparseness in multiblock data structure

Principle: Modified PCs with sparse loadings using the *group lasso penalty*

Result: All variables of a block is selected or removed

Group Sparse PCA

Group Sparse PCA: Extension of SPCA-rSVD for blocks of quantitative variables



Challenge: select groups of continuous variables
$\rightarrow$ zero coefficients to entire blocks of variables

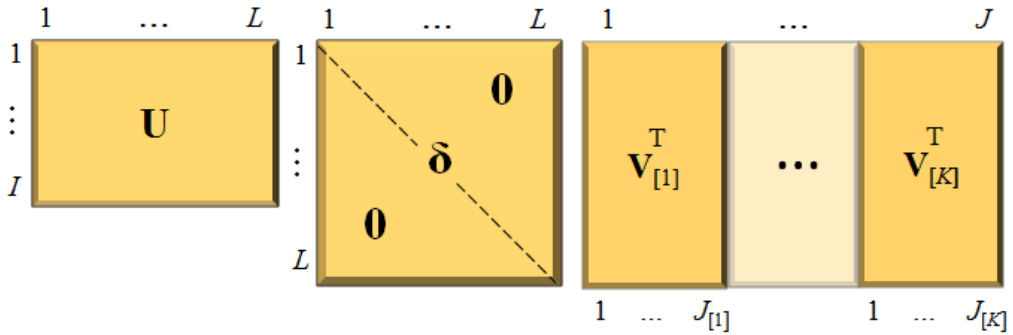
How? A compromise between SPCA-rSVD and group Lasso

Group Sparse PCA

$\mathbf{X}$: matrix of quantitative variables divided into K sub-matrices

SVD of $\mathbf{X}$:

$$\mathbf{X} = \left[\mathbf{X}_{[1]}|...|\mathbf{X}_{[k]}|...|\mathbf{X}_{[K]}\right] = \mathbf{UDV}^T$$
$$= \mathbf{UD}\left[\mathbf{V}_{[1]}^T|...|\mathbf{V}_{[k]}^T|...|\mathbf{V}_{[K]}^T\right]$$



where each line of $\mathbf{V}^T$ is a vector divided into K blocks.

Group Sparse PCA

Remember SPCA-rSVD:

$$\min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 + P_\lambda(\tilde{\mathbf{v}}) \tag{7}$$

$\tilde{\mathbf{u}}$ fixed: $\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \min_{\tilde{v}_j}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{v}_j + P_\lambda(\tilde{v}_j)\right)$

Group Sparse PCA

Remember SPCA-rSVD:

$$\min_{\tilde{\mathbf{u}},\tilde{\mathbf{v}}} \|\mathbf{X} - \tilde{\mathbf{u}}\tilde{\mathbf{v}}^T\|_2^2 + P_\lambda(\tilde{\mathbf{v}}) \tag{7}$$

$\tilde{\mathbf{u}}$ fixed: $\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \min_{\tilde{v}_j}\left(\tilde{v}_j^2 - 2(\mathbf{X}^T\tilde{\mathbf{u}})_j\tilde{v}_j + P_\lambda(\tilde{v}_j)\right)$

In GSPCA:

$$\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}^T\tilde{\mathbf{u}}) = \min_{\tilde{\mathbf{v}}_{[k]}}\left(\operatorname{tr}(\tilde{\mathbf{v}}_{[k]}\tilde{\mathbf{v}}_{[k]}^T) - 2\operatorname{tr}(\tilde{\mathbf{v}}_{[k]}\tilde{\mathbf{u}}^T\mathbf{X}_{[k]}) + P_\lambda(\tilde{\mathbf{v}}_{[k]})\right)$$

Here P_λ is the **group lasso regularizer**:

$$P_\lambda(\tilde{\mathbf{v}}) = \lambda\sum_{k=1}^{K}\sqrt{J_{[k]}}\|\tilde{\mathbf{v}}_{[k]}\|_2 \tag{8}$$

The solution is

$$h_\lambda(\mathbf{X}_{[k]}^T\tilde{\mathbf{u}}) = \left(1 - \frac{\lambda}{2}\frac{\sqrt{J_{[k]}}}{\|\mathbf{X}_{[k]}^T\tilde{\mathbf{u}}\|_2}\right)_+ \mathbf{X}_{[k]}^T\tilde{\mathbf{u}} \tag{9}$$

Group Sparse PCA

Group Sparse PCA Algorithm

Group Sparse PCA

Group Sparse PCA Algorithm

1. Apply the SVD to $\mathbf{X}$ and obtain the best rank-one approximation of $\mathbf{X}$ as $\delta\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$. Set $\tilde{\mathbf{v}}_{old} = \tilde{\mathbf{v}} = [\tilde{\mathbf{v}}_{[1]}, \ldots, \tilde{\mathbf{v}}_{[K]}]$ and $\tilde{\mathbf{u}}_{old} = \delta\tilde{\mathbf{u}}$.

Group Sparse PCA

Group Sparse PCA Algorithm

1. Apply the SVD to $\mathbf{X}$ and obtain the best rank-one approximation of $\mathbf{X}$ as $\delta\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$. Set $\tilde{\mathbf{v}}_{old} = \tilde{\mathbf{v}} = [\tilde{\mathbf{v}}_{[1]}, \ldots, \tilde{\mathbf{v}}_{[K]}]$ and $\tilde{\mathbf{u}}_{old} = \delta\tilde{\mathbf{u}}$.
2. Update:
   a) $\tilde{\mathbf{v}}_{new} = [h_\lambda(\mathbf{X}_{[1]}^T\tilde{\mathbf{u}}_{old}), \ldots, h_\lambda(\mathbf{X}_{[K]}^T\tilde{\mathbf{u}}_{old})]$
   b) $\tilde{\mathbf{u}}_{new} = \mathbf{X}\tilde{\mathbf{v}}_{new}/\|\mathbf{X}\tilde{\mathbf{v}}_{new}\|$

Group Sparse PCA

Group Sparse PCA Algorithm

1. Apply the SVD to $\mathbf{X}$ and obtain the best rank-one approximation of $\mathbf{X}$ as $\delta\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$. Set $\tilde{\mathbf{v}}_{old} = \tilde{\mathbf{v}} = [\tilde{\mathbf{v}}_{[1]}, \ldots, \tilde{\mathbf{v}}_{[K]}]$ and $\tilde{\mathbf{u}}_{old} = \delta\tilde{\mathbf{u}}$.
2. Update:
   a) $\tilde{\mathbf{v}}_{new} = [h_\lambda(\mathbf{X}_{[1]}^T\tilde{\mathbf{u}}_{old}), \ldots, h_\lambda(\mathbf{X}_{[K]}^T\tilde{\mathbf{u}}_{old})]$
   b) $\tilde{\mathbf{u}}_{new} = \mathbf{X}\tilde{\mathbf{v}}_{new}/\|\mathbf{X}\tilde{\mathbf{v}}_{new}\|$
3. Repeat Step 2 replacing $\tilde{\mathbf{u}}_{old}$ and $\tilde{\mathbf{v}}_{old}$ by $\tilde{\mathbf{u}}_{new}$ and $\tilde{\mathbf{v}}_{new}$ until convergence.

Group Sparse PCA

Group Sparse PCA Algorithm

1. Apply the SVD to $\mathbf{X}$ and obtain the best rank-one approximation of $\mathbf{X}$ as $\delta\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$. Set $\tilde{\mathbf{v}}_{old} = \tilde{\mathbf{v}} = [\tilde{\mathbf{v}}_{[1]}, \ldots, \tilde{\mathbf{v}}_{[K]}]$ and $\tilde{\mathbf{u}}_{old} = \delta\tilde{\mathbf{u}}$.
2. Update:
   a) $\tilde{\mathbf{v}}_{new} = [h_\lambda(\mathbf{X}_{[1]}^T\tilde{\mathbf{u}}_{old}), \ldots, h_\lambda(\mathbf{X}_{[K]}^T\tilde{\mathbf{u}}_{old})]$
   b) $\tilde{\mathbf{u}}_{new} = \mathbf{X}\tilde{\mathbf{v}}_{new}/\|\mathbf{X}\tilde{\mathbf{v}}_{new}\|$
3. Repeat Step 2 replacing $\tilde{\mathbf{u}}_{old}$ and $\tilde{\mathbf{v}}_{old}$ by $\tilde{\mathbf{u}}_{new}$ and $\tilde{\mathbf{v}}_{new}$ until convergence.

Sparse loadings $\mathbf{v}_i$ $(i > 1)$ are obtained via rank-one approximation of residual matrices.

Group Sparse PCA

Group Sparse PCA Algorithm

1. Apply the SVD to $\mathbf{X}$ and obtain the best rank-one approximation of $\mathbf{X}$ as $\delta\tilde{\mathbf{u}}\tilde{\mathbf{v}}^T$. Set $\tilde{\mathbf{v}}_{old} = \tilde{\mathbf{v}} = [\tilde{\mathbf{v}}_{[1]}, \ldots, \tilde{\mathbf{v}}_{[K]}]$ and $\tilde{\mathbf{u}}_{old} = \delta\tilde{\mathbf{u}}$.
2. Update:
   a) $\tilde{\mathbf{v}}_{new} = [h_\lambda(\mathbf{X}_{[1]}^T\tilde{\mathbf{u}}_{old}), \ldots, h_\lambda(\mathbf{X}_{[K]}^T\tilde{\mathbf{u}}_{old})]$
   b) $\tilde{\mathbf{u}}_{new} = \mathbf{X}\tilde{\mathbf{v}}_{new}/\|\mathbf{X}\tilde{\mathbf{v}}_{new}\|$
3. Repeat Step 2 replacing $\tilde{\mathbf{u}}_{old}$ and $\tilde{\mathbf{v}}_{old}$ by $\tilde{\mathbf{u}}_{new}$ and $\tilde{\mathbf{v}}_{new}$ until convergence.

Sparse loadings $\mathbf{v}_i$ $(i > 1)$ are obtained via rank-one approximation of residual matrices.

$\rightarrow$ **Choice of** λ using cross-validation or an ad-hoc approach.

Multiblock data structure

Quantitative variables

Categorical variables

Multiblock structure
(I x $J_{[k]}$)
k=1, ...,K

Group selection

Group Sparse PCA

Sparse MCA

To select **1 column** in the original table (categorical variable **X**)

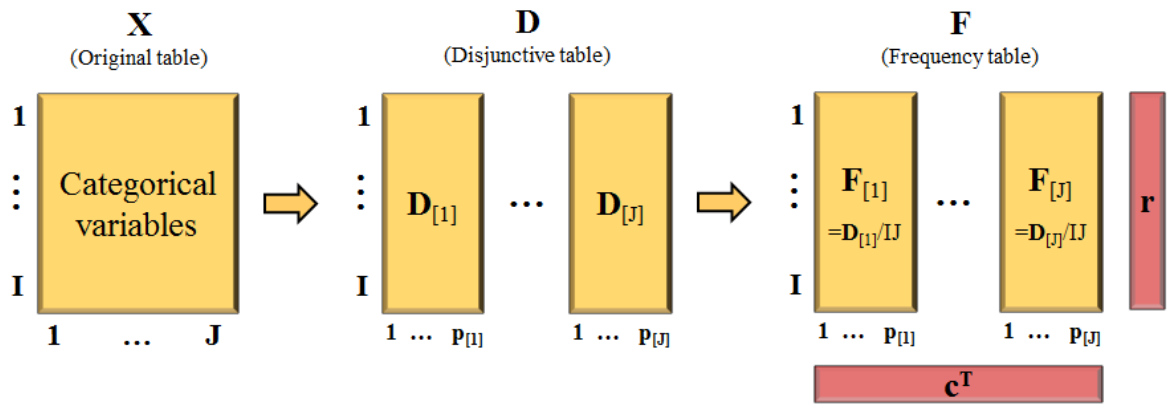
=

To select **a block of indicator variables** in the complete disjunctive table

Sparse MCA (SMCA):

Extension of the **GSPCA** for blocks of **indicator variables**

MCA



$\mathbf{r} = \mathbf{F1}$
$\mathbf{c} = \mathbf{F}^T\mathbf{1}$
$\mathbf{D_r} = diag(\mathbf{r})$
$\mathbf{D_c} = diag(\mathbf{c})$
$\mathbf{p}_{[j]}$ =Nb of modalities of variable j

Generalized SVD of F

$$\mathbf{F} = \mathbf{P\Delta Q}^T \qquad \text{with } \mathbf{P}^T\mathbf{MP} = \mathbf{Q}^T\mathbf{WQ} = \mathbf{I}$$

with $\mathbf{F} = [\mathbf{F}_{[1]}|\dots|\mathbf{F}_{[j]}|\dots|\mathbf{F}_{[J]}]$ and $\mathbf{Q} = [\mathbf{Q}_{[1]}|\dots|\mathbf{Q}_{[j]}|\dots|\mathbf{Q}_{[J]}]$

In the case of PCA: $\mathbf{M} = \mathbf{W} = \mathbf{I}$

In the case of MCA: $\mathbf{M} = \mathbf{D_r}$ $\mathbf{W} = \mathbf{D_c}$

From MCA to SMCA

GSVD as low rank approximation matrices

$$\min_{\mathbf{F}^{(1)}} \|\mathbf{F} - \mathbf{F}^{(1)}\|^2_{\mathbf{W}} = \min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} \tag{10}$$

From MCA to SMCA

GSVD as low rank approximation matrices

$$\min_{\mathbf{F}^{(1)}} \|\mathbf{F} - \mathbf{F}^{(1)}\|^2_{\mathbf{W}} = \min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} \quad (10)$$

where $\|\mathbf{F}\|^2_{\mathbf{W}} = \mathrm{tr}(\mathbf{M}^{\frac{1}{2}}\mathbf{F}\mathbf{W}\mathbf{F}^T\mathbf{M}^{\frac{1}{2}})$ is the norm $\mathbf{W}$-generalized

The norm is under the constraints of $\mathbf{M}$ and $\mathbf{W}$

From MCA to SMCA

GSVD as low rank approximation matrices

$$\min_{\mathbf{F}^{(1)}} \|\mathbf{F} - \mathbf{F}^{(1)}\|^2_{\mathbf{W}} = \min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} \tag{10}$$

where $\|\mathbf{F}\|^2_{\mathbf{W}} = \text{tr}(\mathbf{M}^{\frac{1}{2}}\mathbf{F}\mathbf{W}\mathbf{F}^T\mathbf{M}^{\frac{1}{2}})$ is the norm $\mathbf{W}$-generalized

The norm is under the constraints of $\mathbf{M}$ and $\mathbf{W}$

$$\|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} = \text{tr}\left(\mathbf{M}^{\frac{1}{2}}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)\mathbf{W}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)^T\mathbf{M}^{\frac{1}{2}}\right) \tag{11}$$

From MCA to SMCA

GSVD as low rank approximation matrices

$$\min_{\mathbf{F}^{(1)}} \|\mathbf{F} - \mathbf{F}^{(1)}\|^2_{\mathbf{W}} = \min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} \quad (10)$$

where $\|\mathbf{F}\|^2_{\mathbf{W}} = \mathrm{tr}(\mathbf{M}^{\frac{1}{2}}\mathbf{F}\mathbf{W}\mathbf{F}^T\mathbf{M}^{\frac{1}{2}})$ is the norm $\mathbf{W}$-generalized

The norm is under the constraints of $\mathbf{M}$ and $\mathbf{W}$

$$\|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} = \mathrm{tr}\left(\mathbf{M}^{\frac{1}{2}}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)\mathbf{W}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)^T\mathbf{M}^{\frac{1}{2}}\right) \quad (11)$$

Sparse MCA problem

$$\min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} + P_\lambda(\tilde{\mathbf{q}}) \qquad \tilde{\mathbf{p}}^T\mathbf{M}\tilde{\mathbf{p}} = \tilde{\mathbf{q}}^t\mathbf{W}\tilde{\mathbf{q}} = \mathbf{1} \quad (12)$$

From MCA to SMCA

GSVD as low rank approximation matrices

$$\min_{\mathbf{F}^{(1)}} \|\mathbf{F} - \mathbf{F}^{(1)}\|^2_{\mathbf{W}} = \min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} \quad (10)$$

where $\|\mathbf{F}\|^2_{\mathbf{W}} = \mathrm{tr}(\mathbf{M}^{\frac{1}{2}}\mathbf{F}\mathbf{W}\mathbf{F}^T\mathbf{M}^{\frac{1}{2}})$ is the norm $\mathbf{W}$-generalized

The norm is under the constraints of $\mathbf{M}$ and $\mathbf{W}$

$$\|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} = \mathrm{tr}\left(\mathbf{M}^{\frac{1}{2}}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)\mathbf{W}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)^T\mathbf{M}^{\frac{1}{2}}\right) \quad (11)$$

Sparse MCA problem

$$\min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T\|^2_{\mathbf{W}} + P_\lambda(\tilde{\mathbf{q}}) \qquad \tilde{\mathbf{p}}^T\mathbf{M}\tilde{\mathbf{p}} = \tilde{\mathbf{q}}^t\mathbf{W}\tilde{\mathbf{q}} = \mathbf{1} \quad (12)$$

$$\min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \left(\mathrm{tr}\left(\mathbf{M}^{\frac{1}{2}}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)\mathbf{W}(\mathbf{F} - \tilde{\mathbf{p}}\tilde{\mathbf{q}}^T)^T\mathbf{M}^{\frac{1}{2}}\right) + P_\lambda(\tilde{\mathbf{q}})\right) \quad (13)$$

From MCA to SMCA

Remember GSPCA solution

$$\min_{\tilde{\mathbf{v}}_{[k]}} \left(\operatorname{tr}(\tilde{\mathbf{v}}_{[k]} \tilde{\mathbf{v}}_{[k]}^T) - 2 \operatorname{tr}(\tilde{\mathbf{v}}_{[k]} \tilde{\mathbf{u}}^T \mathbf{X}_{[k]}) + P_\lambda(\tilde{\mathbf{v}}_{[k]}) \right)$$

$$\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}_{[k]}^T \tilde{\mathbf{u}}) = \left(1 - \frac{\lambda}{2} \frac{\sqrt{J_{[k]}}}{\|\mathbf{X}_{[k]}^T \tilde{\mathbf{u}}\|_2} \right)_+ \mathbf{X}_{[k]}^T \tilde{\mathbf{u}}$$

From MCA to SMCA

Remember GSPCA solution

$$\min_{\tilde{\mathbf{v}}_{[k]}} \left(\operatorname{tr}(\tilde{\mathbf{v}}_{[k]} \tilde{\mathbf{v}}_{[k]}^T) - 2 \operatorname{tr}(\tilde{\mathbf{v}}_{[k]} \tilde{\mathbf{u}}^T \mathbf{X}_{[k]}) + P_\lambda(\tilde{\mathbf{v}}_{[k]}) \right)$$

$$\tilde{\mathbf{v}} = h_\lambda(\mathbf{X}_{[k]}^T \tilde{\mathbf{u}}) = \left(1 - \frac{\lambda}{2} \frac{\sqrt{J_{[k]}}}{\|\mathbf{X}_{[k]}^T \tilde{\mathbf{u}}\|_2} \right)_+ \mathbf{X}_{[k]}^T \tilde{\mathbf{u}}$$

Sparse MCA solution

$$\min_{\tilde{\mathbf{q}}_{[k]}} \left(\operatorname{tr}(\tilde{\mathbf{q}}_{[k]} \mathbf{W}_{[k]} \tilde{\mathbf{q}}_{[k]}^T) - 2 \operatorname{tr}(\mathbf{M} \mathbf{F}_{[k]} \mathbf{W}_{[k]} \tilde{\mathbf{q}}_{[k]} \tilde{\mathbf{p}}^T) + P_\lambda(\tilde{\mathbf{q}}_{[k]}) \right)$$

$$\tilde{\mathbf{q}} = h_\lambda(\mathbf{F}_{[k]}^T \mathbf{M} \tilde{\mathbf{p}}) = \left(1 - \frac{\lambda}{2} \frac{\sqrt{J_{[k]}}}{\|\mathbf{F}_{[k]}^T \mathbf{M} \tilde{\mathbf{p}}\|_{\mathbf{W}}} \right)_+ \mathbf{F}_{[k]}^T \mathbf{M} \tilde{\mathbf{p}}$$

Sparse MCA

Sparse MCA Algorithm

1. Apply the GSVD to $\mathbf{F}$ and obtain the best rank-one approximation of $\mathbf{F}$ as $\delta\tilde{\mathbf{p}}\tilde{\mathbf{q}}^T$.
   Set $\tilde{\mathbf{p}}_{old} = \delta\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}_{old} = \tilde{\mathbf{q}}$.
2. Update:
   a) $\tilde{\mathbf{q}}_{new} = [h_\lambda(\mathbf{F}_{[1]}^T\mathbf{M}\tilde{\mathbf{p}}_{old}), \ldots, h_\lambda(\mathbf{F}_{[J]}^T\mathbf{M}\tilde{\mathbf{p}}_{old})]$
   b) $\tilde{\mathbf{p}}_{new} = \mathbf{F}\tilde{\mathbf{q}}_{new}/\|\mathbf{F}\tilde{\mathbf{q}}_{new}\|$
3. Repeat Step 2 replacing $\tilde{\mathbf{p}}_{old}$ and $\tilde{\mathbf{q}}_{old}$ by $\tilde{\mathbf{p}}_{new}$ and $\tilde{\mathbf{q}}_{new}$ until convergence.

Sparse loadings $\mathbf{q}_i$ $(i > 1)$ are obtained via rank-one approximation of residual matrices.

Properties of the SMCA

Properties	MCA	Sparse MCA
Barycentric properties	TRUE	TRUE
Non correlated components	TRUE	FALSE
Orthogonal loadings	TRUE	FALSE
Inertia	$\frac{\delta_j}{\text{tot}} \times 100$	Adjusted variance

Sparse MCA: exemple on breeds of dogs

X_1 Size	...	X_6 Aggressiveness
large (L)		agressive (A)
medium (M)		agressive (A)
.	...	.
.		.
.		.
small (S)		nonagressive (N)

$D_{[1]}$ Size			...	$D_{[6]}$ Aggressiveness	
S.	M.	L.		A	N
0	0	1		1	0
0	1	0		1	0
.	.	.	...	.	.
.	.	.		.	.
.	.	.		.	.
1	0	0		0	1

From Tenenhaus, M. (2007)

Data

$I = 27$ breeds of dogs

$J = 6$ variables

$Q = 16$ (total number of modalities)

$\mathbf{X}$: 27×6 matrix of categorical variables

$\mathbf{D}$: 27×16 complete disjunctive table $\rightarrow \mathbf{D}=(\mathbf{D}_{[1]}, ..., \mathbf{D}_{[6]})$

1 block

=

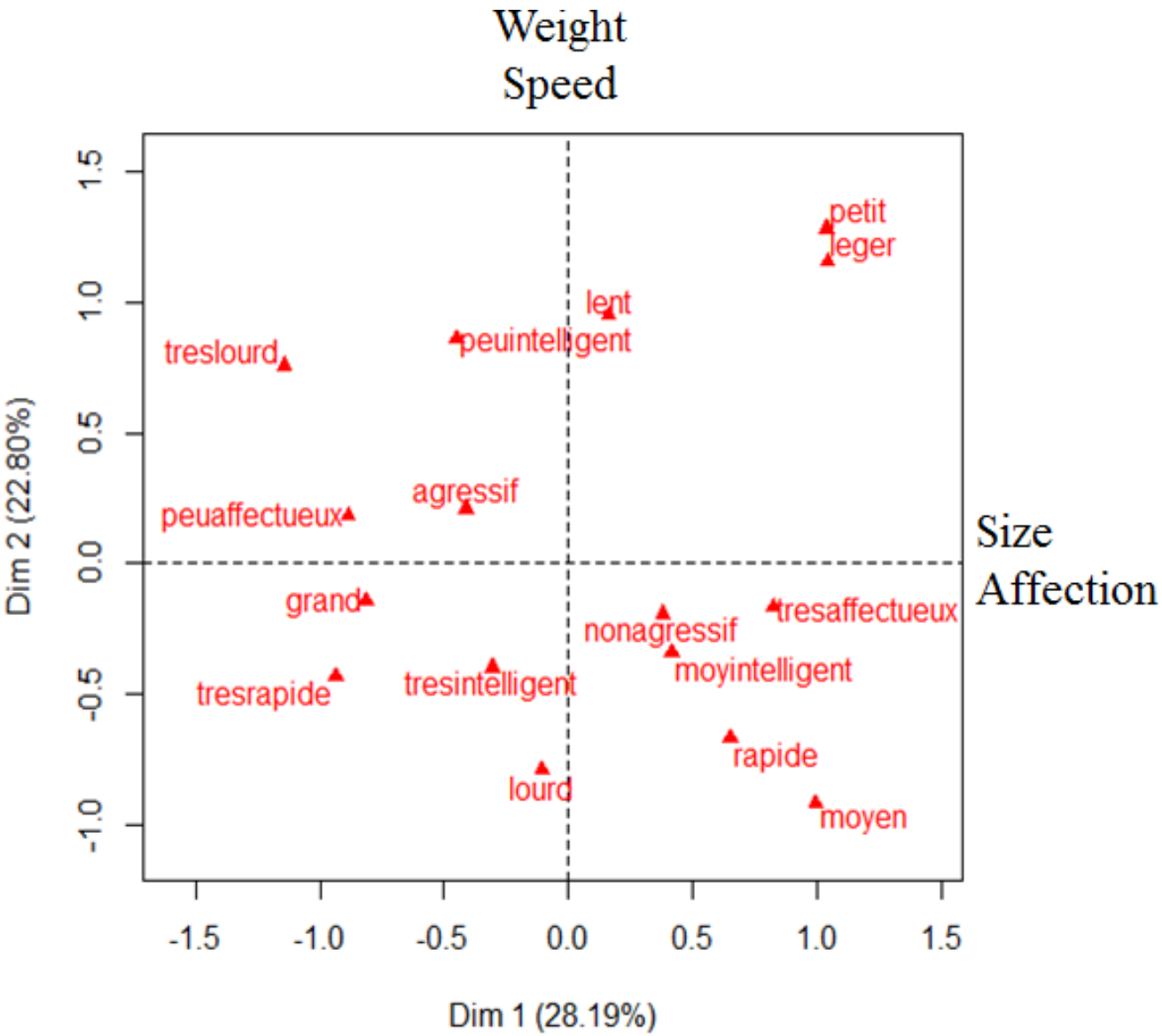
1 descriptive variable

=

1 sub-matrix $\mathbf{D}_{[j]}$

Sparse MCA: exemple on breeds of dogs

Result of the MCA

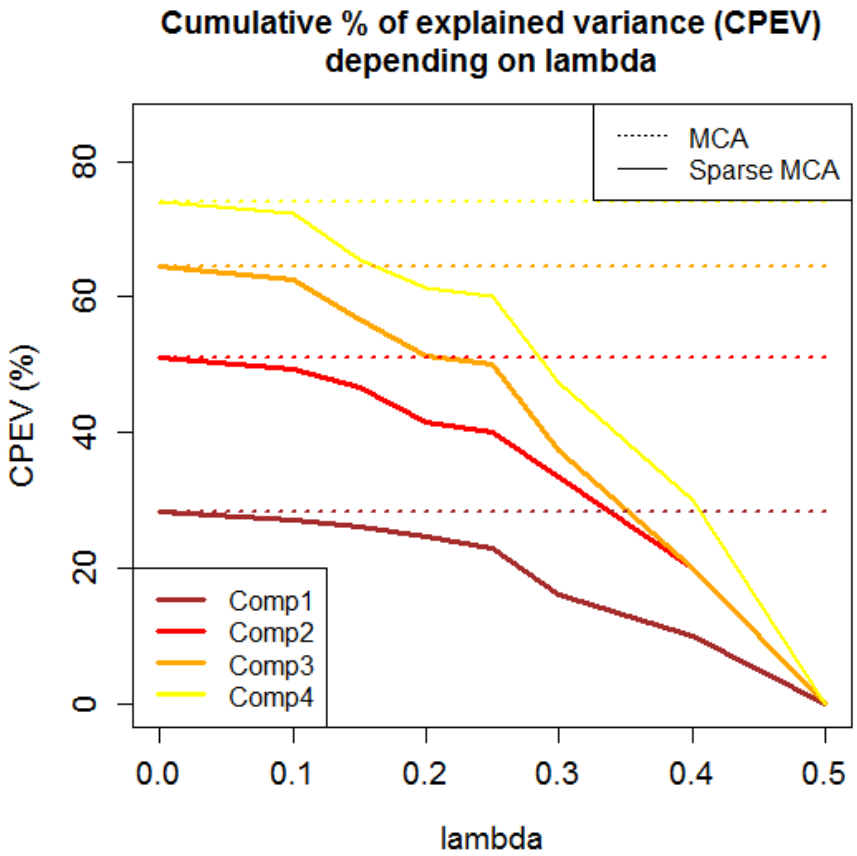


Sparse MCA: exemple on breeds of dogs

Result of the Sparse MCA

From $\lambda = 0.25$:
CPEV $\searrow$ rapidly

We choose $\lambda = 0.25$ to find a compromise between the number of variables selected and the % of variance lost.

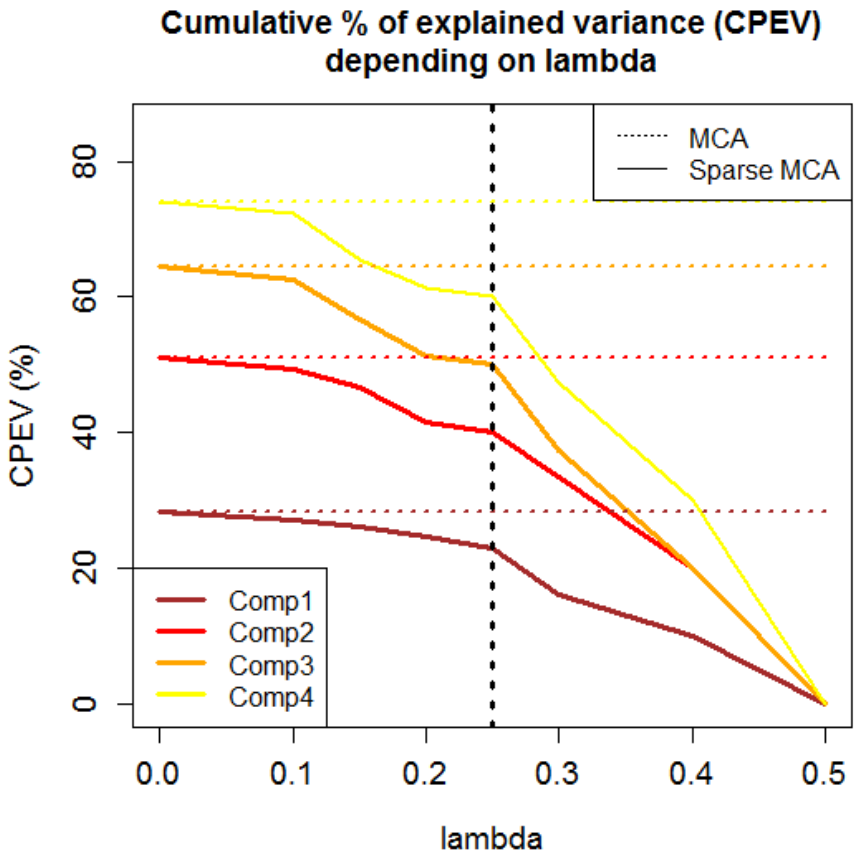


Sparse MCA: exemple on breeds of dogs

Result of the Sparse MCA

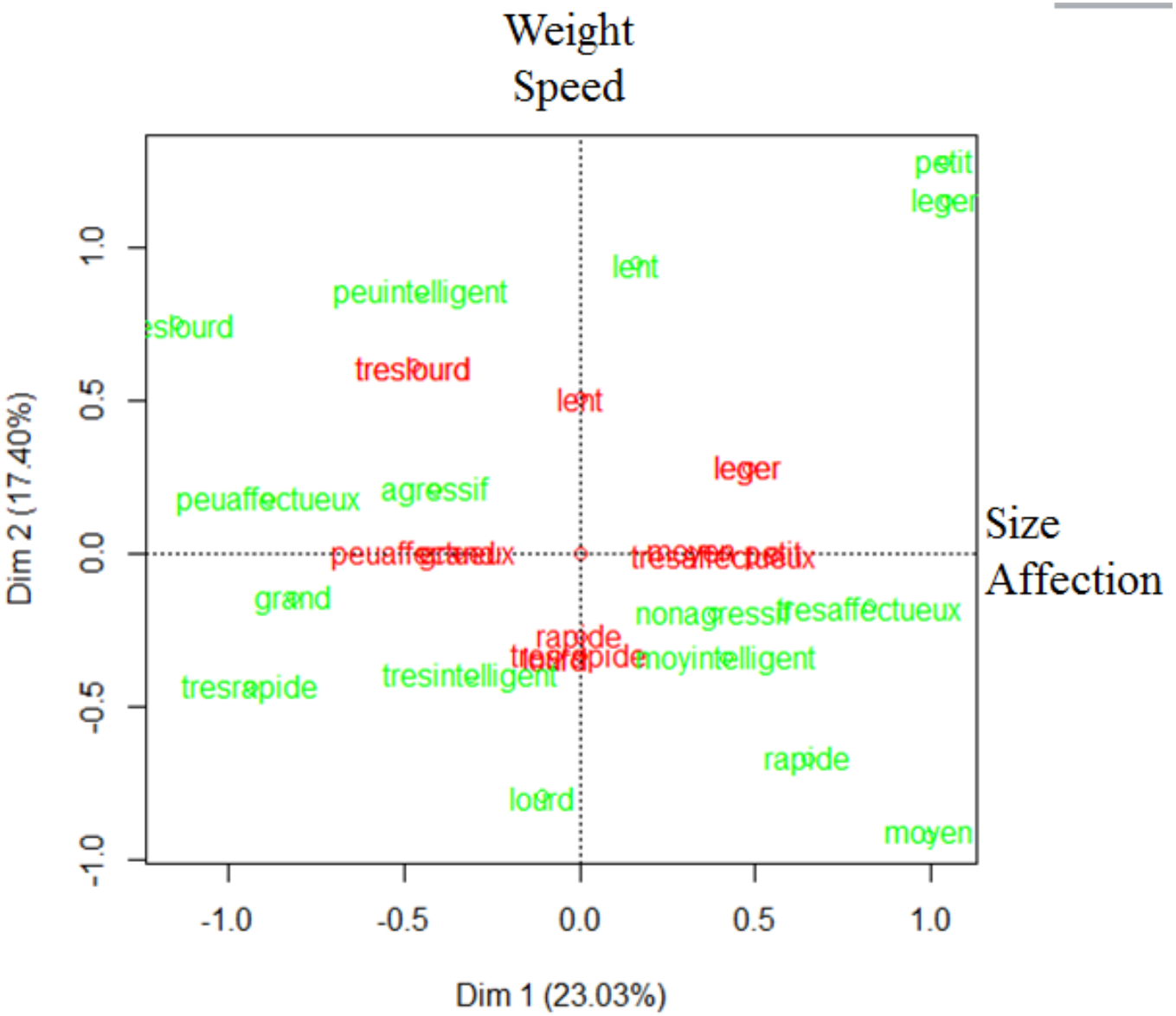
From $\lambda = 0.25$:
CPEV $\searrow$ rapidly

We choose $\lambda = 0.25$ to find a compromise between the number of variables selected and the % of variance lost.



Sparse MCA: exemple on breeds of dogs

Result of the Sparse MCA



Sparse MCA: exemple on breeds of dogs

Comparison of the loadings

	MCA				Sparse MCA			
Variable	Comp1	Comp2	Comp3	Comp4	Comp1	Comp2	Comp3	Comp4
large	-0.361	0.071	-0.005	0.060	-0.389	**0.000**	**0.000**	**0.000**
medium	0.280	0.287	0.300	-0.055	0.226	**0.000**	**0.000**	**0.000**
small	0.291	-0.400	-0.293	-0.041	0.390	**0.000**	**0.000**	**0.000**
lightweight	0.316	-0.389	-0.193	-0.081	0.368	-0.256	**0.000**	**0.000**
heavy	-0.047	0.390	-0.133	0.088	-0.075	0.451	**0.000**	**0.000**
very heavy	-0.294	-0.215	0.458	-0.055	-0.305	-0.479	**0.000**	**0.000**
slow	0.059	-0.383	0.296	0.133	**0.000**	-0.561	**0.000**	**0.000**
fast	0.224	0.256	0.057	-0.299	**0.000**	0.282	**0.000**	**0.000**
veryfast	-0.303	0.156	-0.391	0.168	**0.000**	0.328	**0.000**	**0.000**
unintelligent	0.173	0.157	0.356	0.236	**0.000**	**0.000**	0.693	-0.693
avg intelligent	-0.145	-0.309	-0.168	0.125	**0.000**	**0.000**	-0.327	0.327
very intelligent	-0.086	0.125	-0.330	-0.491	**0.000**	**0.000**	-0.642	0.642
unloving	-0.366	-0.084	0.030	0.087	-0.462	**0.000**	**0.000**	**0.000**
veryaffectionate	0.353	0.081	-0.029	-0.084	0.445	**0.000**	**0.000**	**0.000**
agressive	-0.170	-0.096	0.162	-0.515	**0.000**	**0.000**	**0.000**	**0.000**
non agressive	0.164	0.093	-0.156	0.497	**0.000**	**0.000**	**0.000**	**0.000**
Nb non-zero loadings	16	16	16	16	8	6	3	3
Adjusted variance (%)	28.19	22.80	13.45	9.55	23.03	17.40	10.20	9.50

Sparse MCA: application on SNPs data

Single nucleotide polymorphism (SNP):
A single base pair mutation at a specific locus
$\rightarrow$ The most frequent polymorphism

Ind 1 ...ATCCAG**A**CAG...
Ind 2 ...ATCCAG**T**CAG...

SNP1=$\mathbf{X}_1$	...	SNP537=$\mathbf{X}_{537}$
AA		AB
AB		BB
.	...	.
.		.
.		.
AA		AA
BB		AA

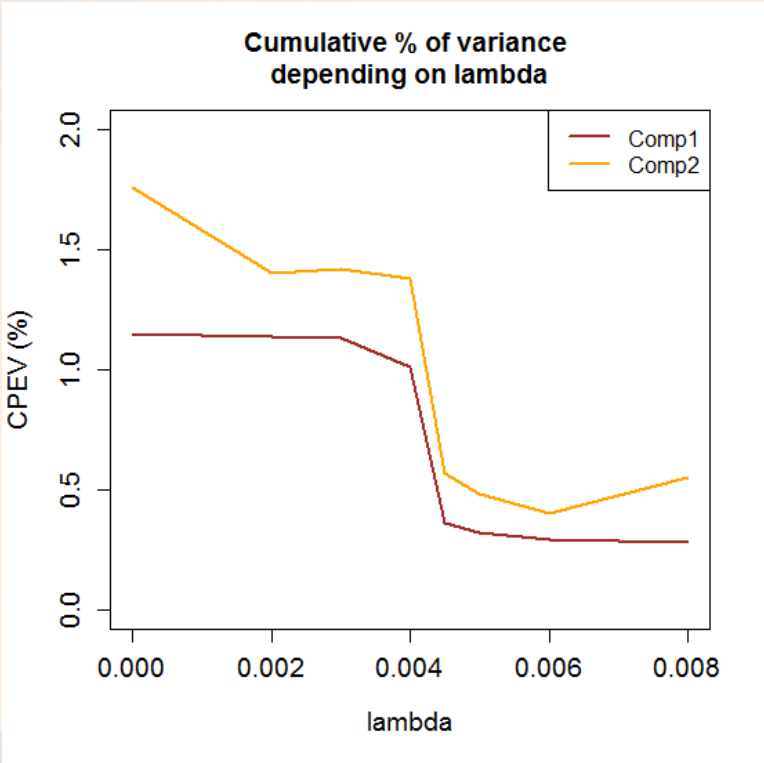
SNP1=$\mathbf{D}_{[1]}$			...	SNP537=$\mathbf{D}_{[537]}$		
AA	AB	BB		AA	AB	BB
1	0	0		0	1	0
0	1	0		0	0	1
.	.	.		.	.	.
.	.	.	...	.	.	.
.	.	.		.	.	.
1	0	0		1	0	0
0	0	1		1	0	0

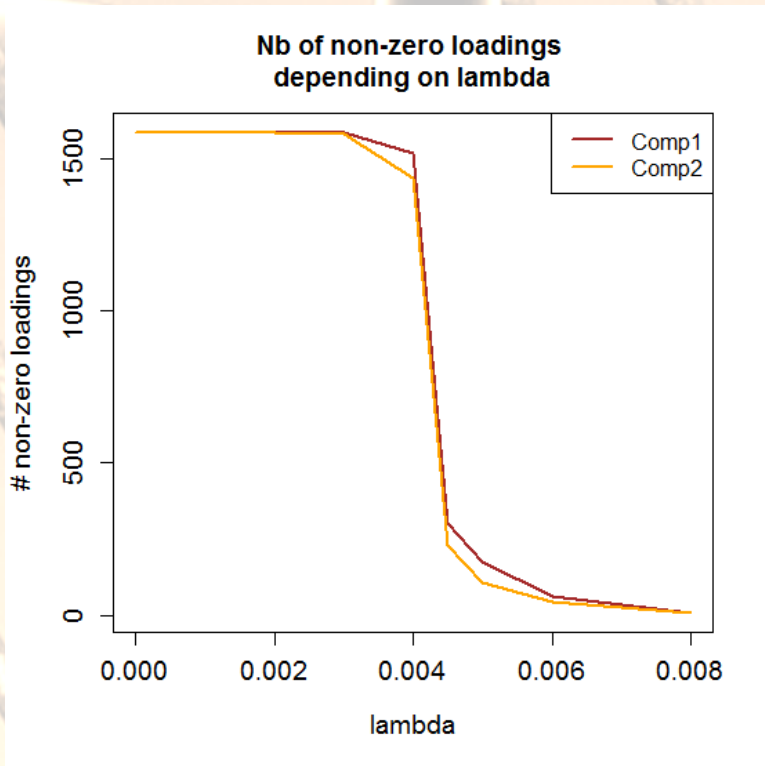
Data

$I = 502$ women
$J = 537$ SNPs
$Q = 1554$ (2 or 3 modalities by SNP)
$\mathbf{X}$: 502×537 matrix of categorical variables (SNPs)
$\mathbf{D}$: 502×1554 complete disjunctive matrix $\rightarrow \mathbf{D}=(\mathbf{D}_{[1]}, ..., \mathbf{D}_{[537]})$

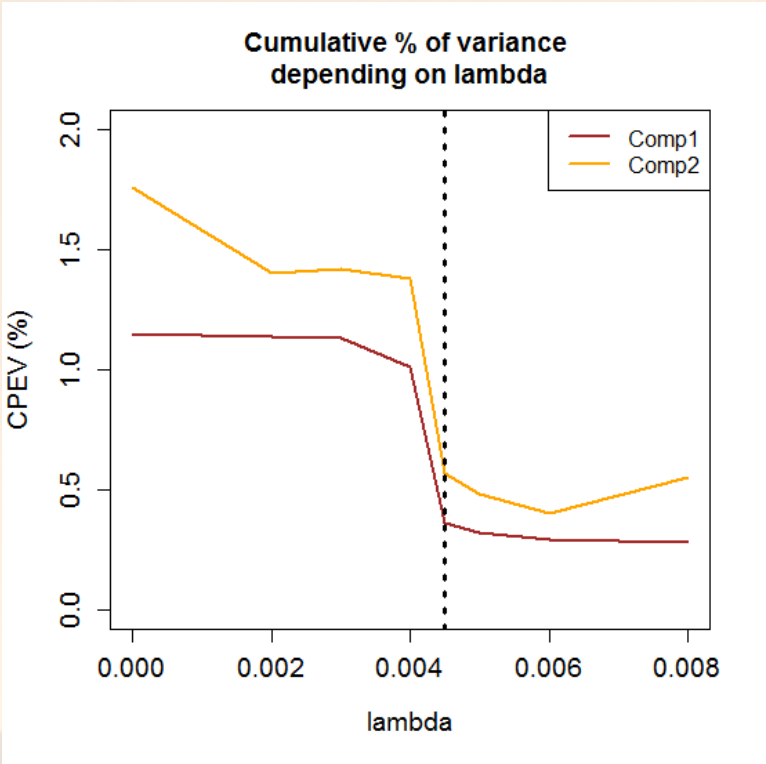
1 block
=
1 SNP=1 sub-matrix $\mathbf{D}_{[j]}$

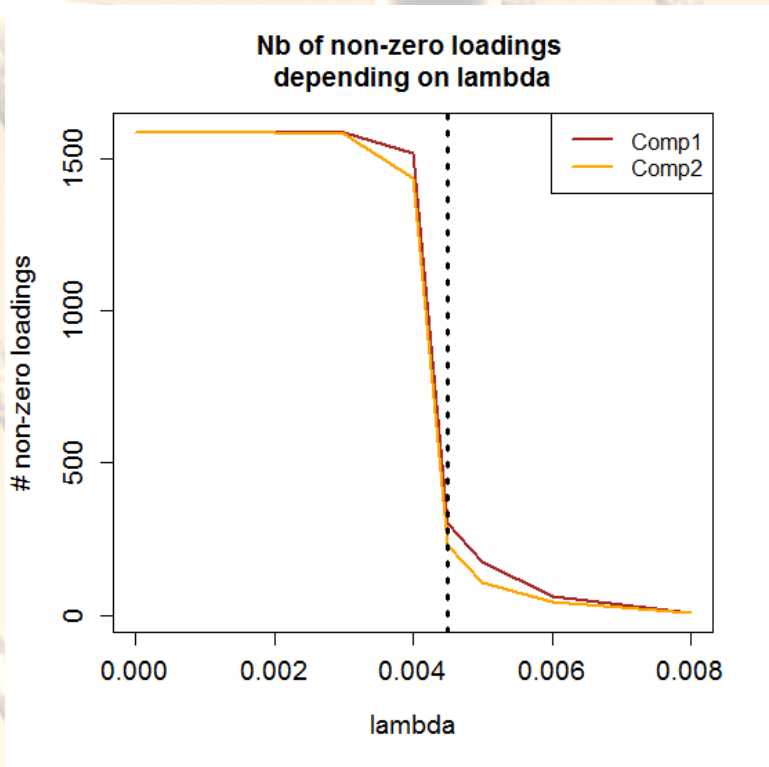
Sparse MCA: application on SNPs data





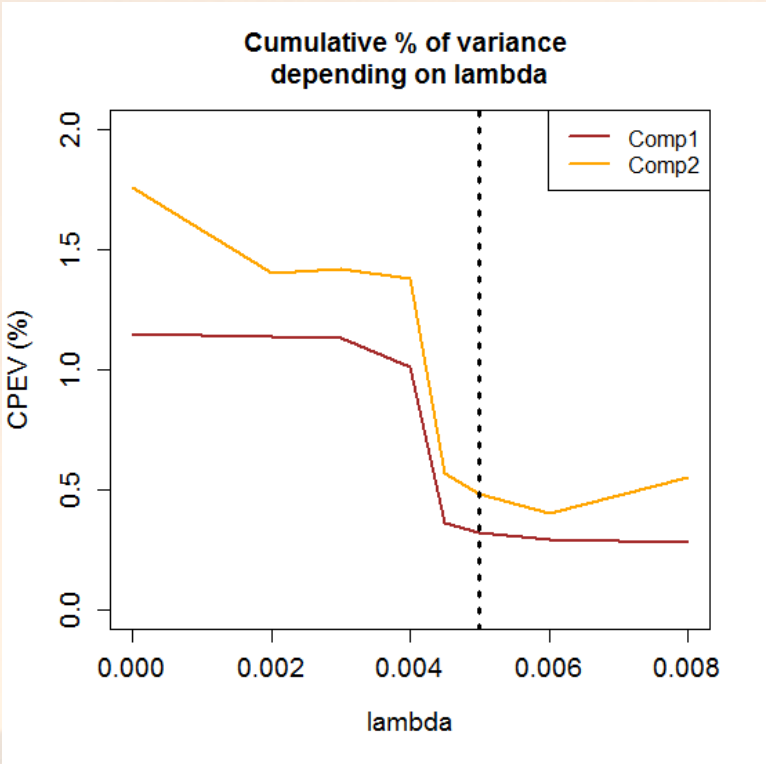
Sparse MCA: application on SNPs data

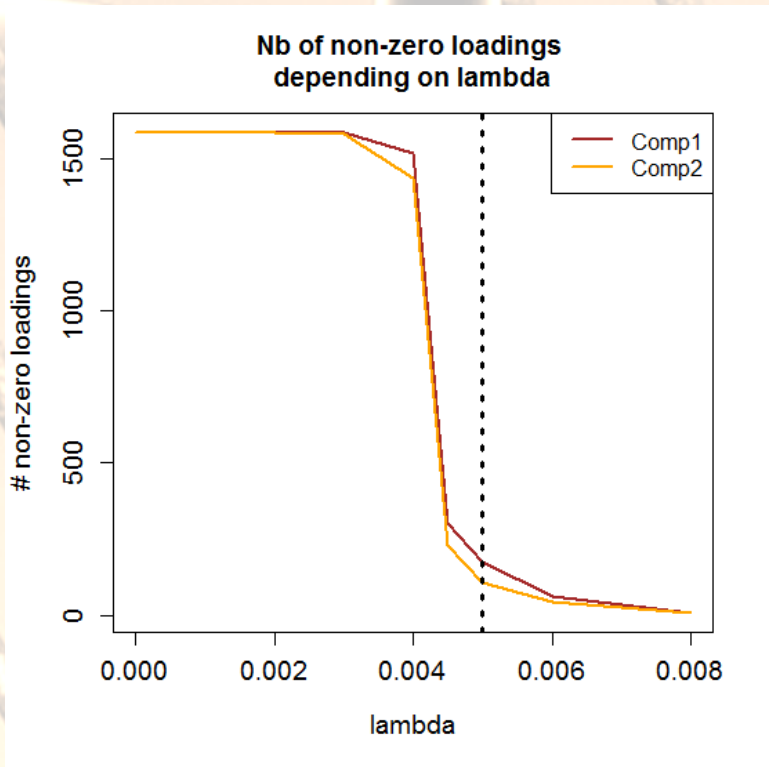




$\lambda = 0.0045$: CPEV=0.36% and 300 columns selected on Comp 1.

Sparse MCA: application on SNPs data





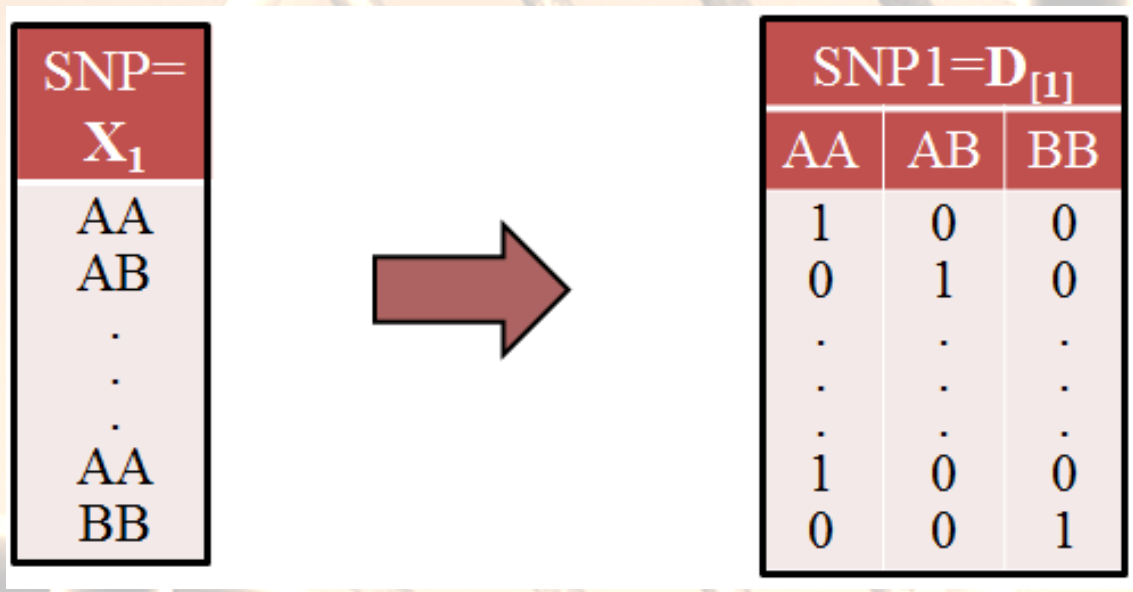
$\lambda = 0.0045$: CPEV=0.36% and 300 columns selected on Comp 1.
$\lambda = 0.005$: CPEV= 0.32% and 174 columns selected on Comp 1.
$\rightarrow$ we choose $\lambda = 0.005$.

Sparse MCA: application on SNPs data

Comparison of the loadings

SNPs	MCA		SMCA	
	Comp1	Comp2	Comp1	Comp2
SNP1.AA	-0.078	0.040	-0.092	0.102
SNP1.AG	-0.014	-0.027	-0.022	-0.053
SNP1.GG	0.150	-0.002	0.132	-0.003
SNP2.AA	-0.082	0.041	-0.118	**0.000**
SNP2.AG	-0.021	-0.025	-0.020	**0.000**
SNP2.GG	-0.081	0.040	-0.001	**0.000**
SNP3.CC	-0.004	0.050	**0.000**	**0.000**
SNP3.CG	0.016	0.021	**0.000**	**0.000**
SNP3.GG	-0.037	-0.325	**0.000**	**0.000**
SNP4.AA	0.149	-0.003	0.050	**0.000**
SNP4.AG	-0.016	-0.025	-0.002	**0.000**
SNP4.GG	-0.081	0.040	-0.100	**0.000**
...	...	...	...	...
Nb non-zero loadings	1554	1554	172	108
Variance (%)	1.14	0.63	0.32	0.16
Cumulative variance (%)	1.14	1.77	0.32	0.48

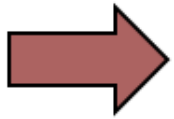
Sparse MCA: application on SNPs data



Sparse MCA: application on SNPs data

GENE N°1

SNP= X_1	SNP= X_3	SNP= X_{55}
AA	AB	AA
AB	AB	AA
.	.	.
.	.	.
.	.	.
AA	AA	AA
BB	AA	AB



Superblock: GENE N°1

SNP1=$D_{[1]}$			SNP3=$D_{[3]}$		SNP55=$D_{[55]}$	
AA	AB	BB	AA	AB	AA	AB
1	0	0	0	1	1	0
0	1	0	0	1	1	0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
1	0	0	1	0	1	0
0	0	1	1	0	0	1

Conclusions

- In a unsupervised multiblock data context: GSPCA for continuous variables, SMCA for categorical variables.
- Produce sparse loading structures (with limited loss of explained variance)
  $\rightarrow$ easier interpretation and comprehension of the results.
- Very powerful in a context of variable selection in high dimension issues
  $\rightarrow$ reduce noise as well as computation time.
- **Research in progress:** Extension of Sparse MCA to select groups and predictors within a group
  $\rightarrow$ sparsity at both group and individual feature levels
  $\rightarrow$ compromise between Sparse MCA and sparse group lasso developed by Simon et al. (2012).

References

[1] Jolliffe, I. T. et al. (2003), A modified principal component technique based on the LASSO, *Journal of Computational and Graphical Statistics*, 12, 531–547,

[2] Shen, H. et Huang, J. (2008), Sparse principal component analysis via regularized low rank matrix approximation, *Journal of Multivariate Analysis*, 99, 1015–1034,

[3] Simon, N., Friedman, J., Hastie, T. et Tibshirani, R. (2012), A Sparse-Group Lasso, *Journal of Computational and Graphical Statistics*, DOI:10.1080/10618600.2012.681250,

[4] Tenenhaus, M. (2007), *Statistique; méthodes pour décrire, expliquer et prévoir*, Dunod, 252–254,

[5] Yuan, M. et Lin, Y. (2006), Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B*, 68, 49–67,

[6] Zou, H., Hastie, T. et Tibshirani, R. (2006), Sparse Principal Component Analysis, *Journal of Computational and Graphical Statistics*, 15, 265–286.