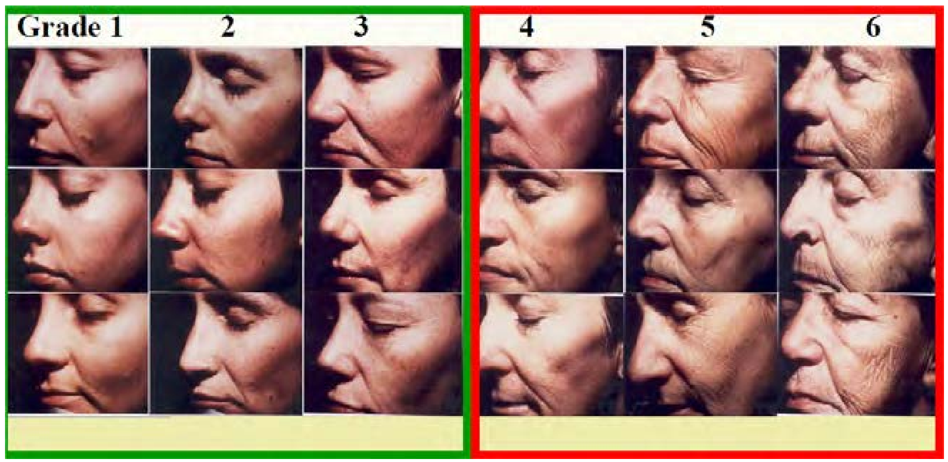
A SURVEY OF SOME SPARSE METHODS FOR HIGH DIMENSIONAL DATA

Gilbert Saporta
CEDRIC,
Conservatoire National des Arts et Métiers, Paris
gilbert.saporta@cnam.fr

- With inputs from Anne Bernard, Ph.D student
- Industrial context and motivation:
  - Relate gene expression data to skin aging measures



  - n=500, p= 800 000 SNP’s, 15 000 genes
  - Ph.D funded by a R&D department of Chanel cosmetic company

Outline

1. Introduction
2. Keeping all variables in regression
3. Sparse regression
4. Sparse PCA
5. Sparse MCA
6. Conclusion and perspectives

1.Introduction

- High dimensional data: p>>n
  - Gene expression data
  - Chemometrics
  - etc.
- Several solutions for regression problems with all variables; but intepretation is difficult
- Sparse methods: provide combinations of few variables

- This talk:
  - a survey of sparse methods for supervised (regression) and unsupervised (PCA) problems
  - New propositions in the unsupervised case when variables belong to disjoint groups or blocks:
    - Group sparse PCA
    - Sparse multiple correspondence analysis

2. Keeping all variables in regression

- No OLS solution
- A special case of multicollinearity
- Regularized regression techniques:
  - Component based: PCR, PLS
  - Ridge
- Lost properties:
  - Bias, scale invariance

→ standardised data

2.1 Principal components regression

- At most n components when p>>n
- Select q components and regress y upon them
  - Orthogonal components → sum of univariate regressions
  - Back to original variables:

$\mathbf{C}$ components matrix n,q $\mathbf{U}$ loadings matrix p,q

$$\mathbf{C} = \mathbf{XU}$$

$$\hat{\mathbf{y}} = \mathbf{C}\hat{\boldsymbol{\alpha}} = \alpha_1 \mathbf{c}_1 + \ldots + \alpha_q \mathbf{c}_q = \mathbf{XU}\hat{\boldsymbol{\alpha}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad \hat{\boldsymbol{\beta}} = \mathbf{U}\hat{\boldsymbol{\alpha}}$$

- Principal components unrelated to the response variable y
- Component ranking
  - Not according to their eigenvalues
  - but according to $r^2(y;c_j)$
- Choice of q
  - crossvalidation
- First papers: Kendall, Hotelling (1957), Malinvaud (1964)

2.2 PLS regression

- Proposed by H. and S.Wold (1960's)
- Close to PCR: projection onto a set of orthogonal combinations of predictors
- PLS components optimised to be predictive of both X and y variables
- Tucker's criterium: $\max \text{cov}^2(y\,;Xw)$

- Trade-off between maximizing correlation between $t=Xw$ and y (OLS) and maximizing variance of t (PCA) :

$$\text{cov}^2(y\,;Xw) = r^2(y\,;Xw)\,V(Xw)\,V(y)$$

- Easy solution:
  - w_j proportional to cov(y; x_j)
  - No surprising signs..
- Further components by iteration on residuals
- Stopping rule: cross-validation

2.3 Ridge regression

- Hoerl & Kennard (1970)

$$\hat{\boldsymbol{\beta}}_R = \left(\mathbf{X'X} + k\mathbf{I}\right)^{-1}\mathbf{X'y}$$

- Several interpretations
  - Tikhonov regularization

$$\min \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right\|^2 \quad \text{with } \left\|\boldsymbol{\beta}\right\|^2 \le c^2$$

$$\text{or} \quad \min\left(\left\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right\|^2 + \lambda\left\|\boldsymbol{\beta}\right\|^2\right)$$

– Bayesian regression

- Gaussian prior for $\boldsymbol{\beta}$ $N(\mathbf{0};\psi^2\mathbf{I})$
- Gaussian distribution Y/β $N(\mathbf{X\beta};\sigma^2\mathbf{I})$

⇒ Maximum a posteriori or posterior expectation :

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X'X} + \frac{\sigma^2}{\psi^2}\mathbf{I}\right)^{-1}\mathbf{X'y}$$

Gives an interpretation for k

- Choice of k :
  - cross-validation
- Effective degree of freedom:

$$df(k) = Trace\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}'\right)$$

$$= \sum_{j=1}^{P} \frac{n\lambda_j}{n\lambda_j + k}$$

- Shrinkage properties (Hastie et al. , 2009)
  - Ridge shrinks all principal directions but shrinks more low variance directions
  - PCR discards low variance direction
  - PLS shrinks low variance directions but inflates high variance directions
- Back to OLS
  - When n>p, PCR and PLS with p components, ridge with k=0 are identical to PLS
  - When p>>n, q should be <n and k>0

3. Sparse regression

- Keeping all predictors is a drawback for high dimensional data: combinations of too many variables cannot be interpreted
- Traditional algorithms of variable selection (best subsets, forward, stepwise) are a bit out of fashion
- Sparse methods simultaneously shrink coefficients and select variables which give better predictions

3.1 Lasso and elastic-net

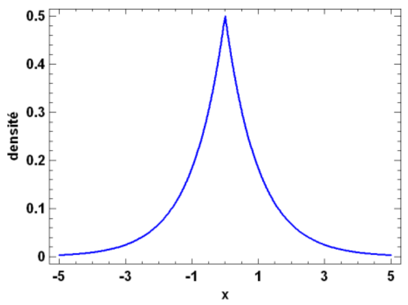
- Lasso (Tibshirani, 1996) imposes a L_1 constraint on the coefficients $\sum_{j=1}^{p} \left| b_j \right| < c$

$$\hat{\boldsymbol{\beta}}_{lasso} = \arg\min_{\boldsymbol{\beta}} \left(\left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \right)$$

- Lasso continuously shrinks the coefficients towards zero
- Convex optimisation; no explicit solution

- Constraints and log-priors
  - Like ridge regression, the Lasso is a bayesian regression but with exponential prior

$$f(\beta_j) = \frac{\lambda}{2} \exp\left(-\lambda \left|\beta_j\right|\right)$$



  - $\left|\beta_j\right|$ is proportional to the log-prior

- A general form:

$$\hat{\boldsymbol{\beta}}_{lasso} = \arg\min_{\boldsymbol{\beta}} \left(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \sum_{j=1}^{p} |\beta_j|^q \right)$$

  - q=2 ridge; q=1 Lasso; q=0 subset selection (counts the number of variables)
  - q>1 do not provide null coefficients (derivability)

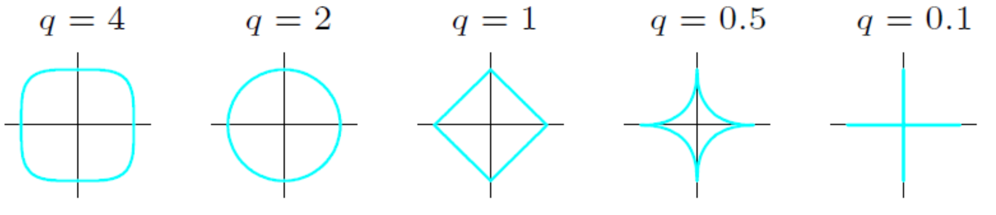


FIGURE 3.12. *Contours of constant value of* $\sum_j |\beta_j|^q$ *for given values of* q.

Elements of Statistical Learning (2nd Ed.) ©Hastie, Tibshirani & Friedman 2009 Chap 3

- Lasso produces a sparse model but the number of variables selected cannot exceed the number of units
- **Elastic net:** combine ridge penalty and lasso penalty to select more predictors than the number of observations (Zou & Hastie, 2005)

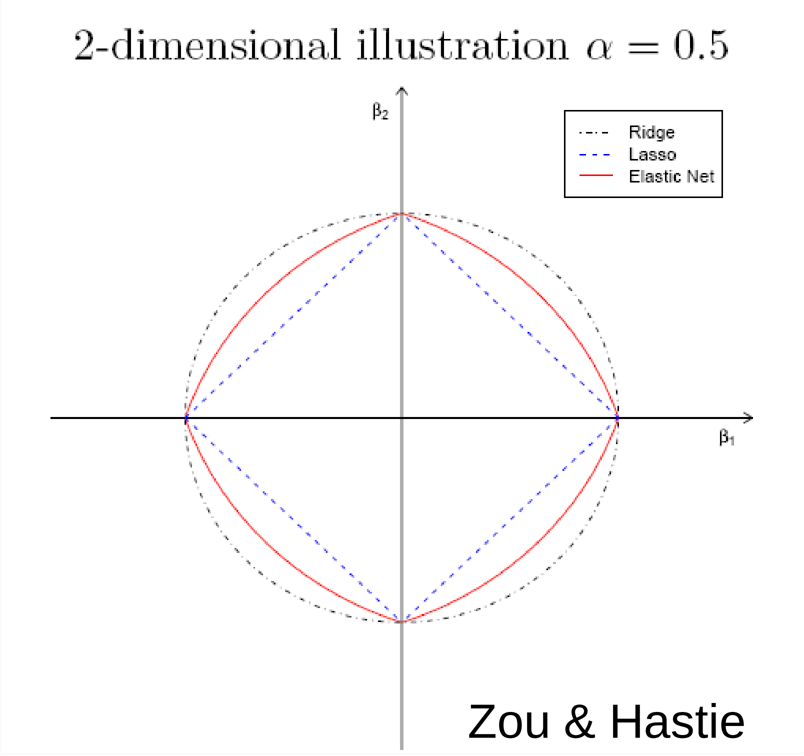
$$\hat{\boldsymbol{\beta}}_{en} = \arg\min_{\boldsymbol{\beta}} \left(\left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^2 + \lambda_2 \left\| \boldsymbol{\beta} \right\|^2 + \lambda_1 \left\| \boldsymbol{\beta} \right\|_1 \right)$$

- Equivalent formulation

$$\arg\min_{\boldsymbol{\beta}} \left(\left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|^2 + \lambda \sum_{j=1}^{p} \left(\alpha \beta_j^2 + (1-\alpha) \left| \beta_j \right| \right) \right)$$

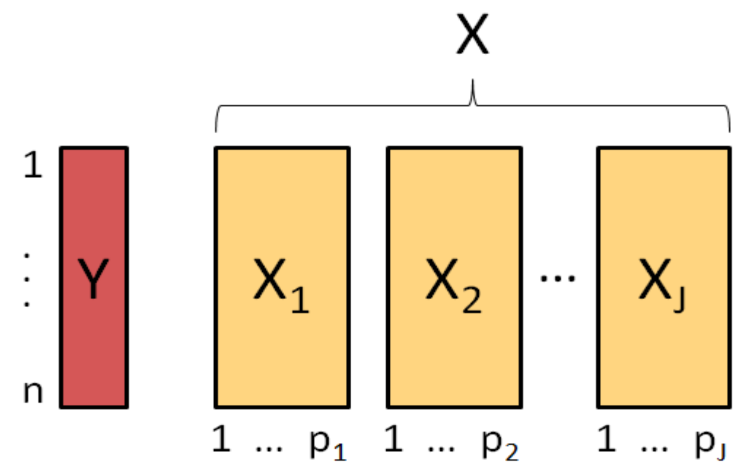
$$\textit{with} \quad \alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

  - The L_1 part selects variables, the L_2 part removes the limitation by n



3.2 Group-lasso

- X matrix divided into J sub-matrices X_j of p_j variables
- **Group Lasso**: extension of Lasso for selecting groups of variables (Yuan & Lin, 2007):



$$\hat{\boldsymbol{\beta}}_{GL} = \arg\min_{\beta} \left\| \mathbf{y} - \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda \sum_{j=1}^{J} \sqrt{p_j} \left| \boldsymbol{\beta}_j \right|$$

If p_j=1 for all j, group Lasso = Lasso

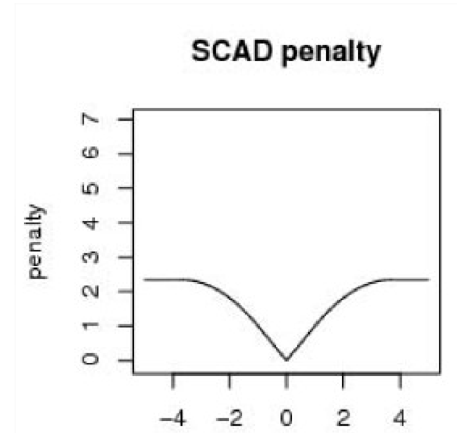
- Drawback: no sparsity within groups
- A solution: **sparse group lasso** (Simon et al. , 2012)

$$\min_{\boldsymbol{\beta}}\left(\left\|\mathbf{y}-\sum_{j=1}^{J}\mathbf{X}_j\boldsymbol{\beta}_j\right\|^2+\lambda_1\sum_{j=1}^{J}\left\|\boldsymbol{\beta}_j\right\|+\lambda_2\sum_{j=1}^{J}\sum_{i=1}^{p_j}\left|\beta_{ij}\right|\right)$$

  – Two tuning parameters

3.3 other sparse regression methods

- SCAD penalty
  - smoothly clipped absolute deviation
  - Non-convex
- Sparse PLS
  - Several extensions
    - Chun & Keles (2010)
    - Le Cao et al. (2008)



4.Sparse PCA

- In PCA, each PC is a linear combination of **all** the original variables : difficult to interpret the results
- **Challenge of SPCA:** obtain components easily interpretable (lot of zero loadings in principal factors)
- **Principle of SPCA:** modify PCA imposing lasso/elastic-net constraints to construct modified PCs with sparse loadings
- **Warning:** Sparse PCA does not provide a global selection of variables but a selection **dimension by dimension** : different from the regression context (Lasso, Elastic Net, …)

4.1 First attempts:

- **Simple PCA**
  - by Vines (2000) : integer loadings
  - Rousson, V. and Gasser, T. (2004) : loadings (+ , 0, -)
- **SCoTLASS** (Simplified Component Technique – Lasso) by Jolliffe & al. (2003) : extra L_1 constraints

$$\max \mathbf{u'Vu} \quad \text{with } \|\mathbf{u}\|^2 = \mathbf{u'u} = 1 \text{ and } \sum_{j=1}^{p} |u_j| \le t$$

SCotLass properties:

$t \geq \sqrt{p}$ usual PCA

$\mathrm{t} < 1$ no solution

$t = 1$ only one nonzero coefficient

$1 < t < \sqrt{p}$

- Non convex problem

4.2 S-PCA by Zou et al (2006)

Let the SVD of X be $\mathbf{X} = \mathbf{UDV}'$ with $\mathbf{Z} = \mathbf{UD}$ the principal components

Ridge regression:

$$\hat{\boldsymbol{\beta}}_{ridge} = \arg\min_{\beta} \|\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|^2$$

$$\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}' \; with \; \mathbf{V}'\mathbf{V} = \mathbf{I}$$

$$\hat{\boldsymbol{\beta}}_{\mathbf{i,ridge}} = \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\lambda}\,\mathbf{I}\right)^{\mathbf{-1}} \mathbf{X}'\left(\mathbf{X}\mathbf{v}_{\mathbf{i}}\right) = \mathbf{v}_{\mathbf{i}} \frac{d_{ii}^2}{d_{ii}^2 + \lambda} \quad \Longrightarrow \quad \tilde{\mathbf{v}} = \mathbf{v}_i$$

Loadings can be recovered by regressing (ridge regression) PCs on the p variables

→PCA can be written as a **regression-type optimization problem**

Sparse PCA add a new penalty to produce sparse loadings:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|^2 + \lambda_1 \|\boldsymbol{\beta}\|_1$$

Lasso penalty

$\hat{\mathbf{v}}_{\mathbf{i}} = \frac{\hat{\boldsymbol{\beta}}}{\|\hat{\boldsymbol{\beta}}\|}$ is an approximation to $\mathbf{v}_{\mathbf{i}}$, and $\mathbf{X}\hat{\mathbf{v}}_{\mathbf{i}}$ the ith approximated component

→Produces sparse loadings with zero coefficients to facilitate interpretation

Alternated algorithm between elastic net and SVD

4.3 S-PCA via regularized SVD

- Shen & Huang (2008) : starts from the SVD with a smooth penalty (L1, SCAD, etc.)

$$\mathbf{X}^{(k)} = \sum_{j=1}^{k} d_j \mathbf{u}_j \mathbf{v}_j^{'}$$

$$\min_{\mathbf{u},\mathbf{v}} \left\| \mathbf{X} - \mathbf{u}\mathbf{v}' \right\|^2 + \sum_{j=1}^{p} g_\lambda\left(\left|v_j\right|\right)$$

Example

Table 1. Definitions of Variables in Jeffers' Pitprop Data

Variable	*Definition*
x_1	Top diameter in inches
x_2	Length in inches
x_3	Moisture content, % of dry weight
x_4	Specific gravity at time of test
x_5	Oven-dry specific gravity
x_6	Number of annual rings at top
x_7	Number of annual rings at bottom
x_8	Maximum bow in inches
x_9	Distance of point of maximum bow from top in inches
x_{10}	Number of knot whorls
x_{11}	Length of clear prop from top in inches
x_{12}	Average number of knots per whorl
x_{13}	Average diameter of the knots in inches

Table 7
(Pitprops data) Loadings of the first six PCs by PCA and sPCA-rSVD-soft

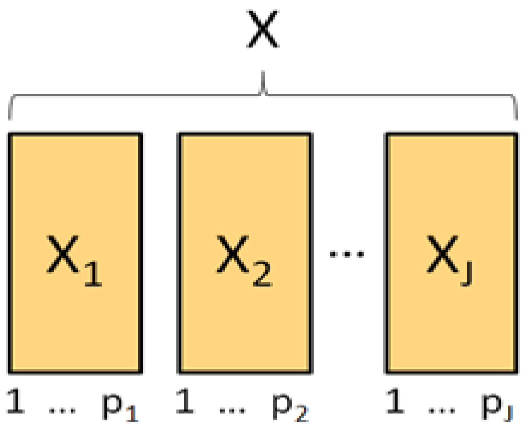
Variable	PCA						sPCA-rSVD-soft					
	PC1	PC2	PC3	PC4	PC5	PC6	PC1	PC2	PC3	PC4	PC5	PC6
x_1	−0.404	0.218	−0.207	0.091	−0.083	0.120	−0.449	0	0	−0.114	0	0
x_2	−0.406	0.186	−0.235	0.103	−0.113	0.163	−0.460	0	0	−0.102	0	0
x_3	−0.124	0.541	0.141	−0.078	0.350	−0.276	0	−0.707	0	0	0	0
x_4	−0.173	0.456	0.352	−0.055	0.356	−0.054	0	−0.707	0	0	0	0
x_5	−0.057	−0.170	0.481	−0.049	0.176	0.626	0	0	0.550	0	0	−0.744
x_6	−0.284	−0.014	0.475	0.063	−0.316	0.052	−0.199	0	0.546	−0.176	0	0
x_7	−0.400	−0.190	0.253	0.065	−0.215	0.003	−0.399	0	0.366	0	0	0
x_8	−0.294	−0.189	−0.243	−0.286	0.185	−0.055	−0.279	0	0	0.422	0	0
x_9	−0.357	0.017	−0.208	−0.097	−0.106	0.034	−0.380	0	0	0	0	0
x_{10}	−0.379	−0.248	−0.119	0.205	0.156	−0.173	−0.407	0	0	0.283	0.231	0
x_{11}	0.011	0.205	−0.070	−0.804	−0.343	0.175	0	0	0	0	−0.973	0
x_{12}	0.115	0.343	0.092	0.301	−0.600	−0.170	0	0	0	−0.785	0	0.161
x_{13}	0.113	0.309	−0.326	0.303	0.080	0.626	0	0	−0.515	−0.265	0	−0.648
Sparsity	0	0	0	0	0	0	6	11	9	6	11	10
CPEV	32.5	50.7	65.2	73.7	80.7	87.0	30.6	45.0	59.0	70.0	78.5	84.5

Shen & Huang

- Loss of orthogonality
  - SCotLass: orthogonal loadings but correlated components
  - S-PCA: neither loadings, nor components are orthogonal
  - Necessity of adjusting the % of explained variance

4.4 Group Sparse PCA

Data matrix X divided into J groups X_j of p_j variables, but no Y



Group Sparse PCA: compromise between SPCA and group Lasso

Goal: select groups of continuous variables (zero coefficients to entire blocks of variables)

Principle: replace the penalty function in the SPCA algorithm

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|^2 + \lambda_1 \|\boldsymbol{\beta}\|_1$$

by that defined in the group Lasso

$$\hat{\boldsymbol{\beta}}_{GL} = \arg\min_{\beta} \left\| \mathbf{Z} - \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda \sum_{j=1}^{J} \sqrt{p_j} \left| \boldsymbol{\beta}_j \right|$$

5.Sparse MCA

Original table

X_J
1
p_J
.
.
.
3

In MCA:

Selection of ***1 column*** in the original table (categorical variable X_j)

=

Selection of ***a block of p_j indicator variables*** in the complete disjunctive table

Complete disjunctive table

X_{J1}	…	X_{Jpj}
1		0
0		1
.		.
.		.
.		.
0		0

Challenge of Sparse MCA : select categorical variables, not categories

Principle: a straightforward extension of Group Sparse PCA for groups of indicator variables, with the chi-square metric . Uses s-PCA r-SVD algorithm.

Let F be the n x q disjunctive table divided by the number of units

$$\mathbf{r} = \mathbf{F}\mathbf{1}_{\mathbf{q}} \qquad \mathbf{c} = \mathbf{F}^{\mathbf{T}}\mathbf{1}_{\mathbf{n}} \qquad \mathbf{D}_{\mathbf{r}} = \mathbf{diag}(\mathbf{r}) \qquad \mathbf{D}_{\mathbf{c}} = \mathbf{diag}(\mathbf{c})$$

Let $\tilde{\mathbf{F}}$ be the matrix of standardised residuals:

$$\tilde{\mathbf{F}} = \mathbf{D}_{\mathbf{r}}^{-\frac{1}{2}}\left(\mathbf{F} - \mathbf{r}\mathbf{c}^{\mathbf{T}}\right)\mathbf{D}_{\mathbf{c}}^{-\frac{1}{2}}$$

Singular Value Decomposition $\tilde{\mathbf{F}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathbf{T}}$

Properties	MCA	Sparse MCA
Uncorrelated Components	TRUE	FALSE
Orthogonal loadings	TRUE	FALSE
Barycentric property	TRUE	TRUE
% of inertia	$\lambda_j / tot \times 100$	$\left\| \tilde{\mathbf{Z}}_{\mathbf{j.1,\ldots,j\text{-}1}} \right\|^2$
Total inertia	$\frac{1}{p}\sum_{j=1}^{p} p_j - 1$	$\sum_{j=1}^{k} \left\| \tilde{\mathbf{Z}}_{\mathbf{j.1,\ldots,j\text{-}1}} \right\|^2$

$\tilde{\mathbf{Z}}_{\mathbf{j.1,\ldots,j\text{-}1}}$ are the residuals after adjusting $\tilde{\mathbf{Z}}_{\mathbf{j}}$ for $\tilde{\mathbf{Z}}_{\mathbf{1,\ldots,j\text{-}1}}$ (regression projection)

Toy example: Dogs

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

K_1 Size			…	K_6 Aggressiveness	
S.	M.	L.		A	N
0	0	1		1	0
0	1	0		1	0
.	.	.		.	.
.	.	.	…	.	.
.	.	.		.	.
1	0	0		0	1

Data:

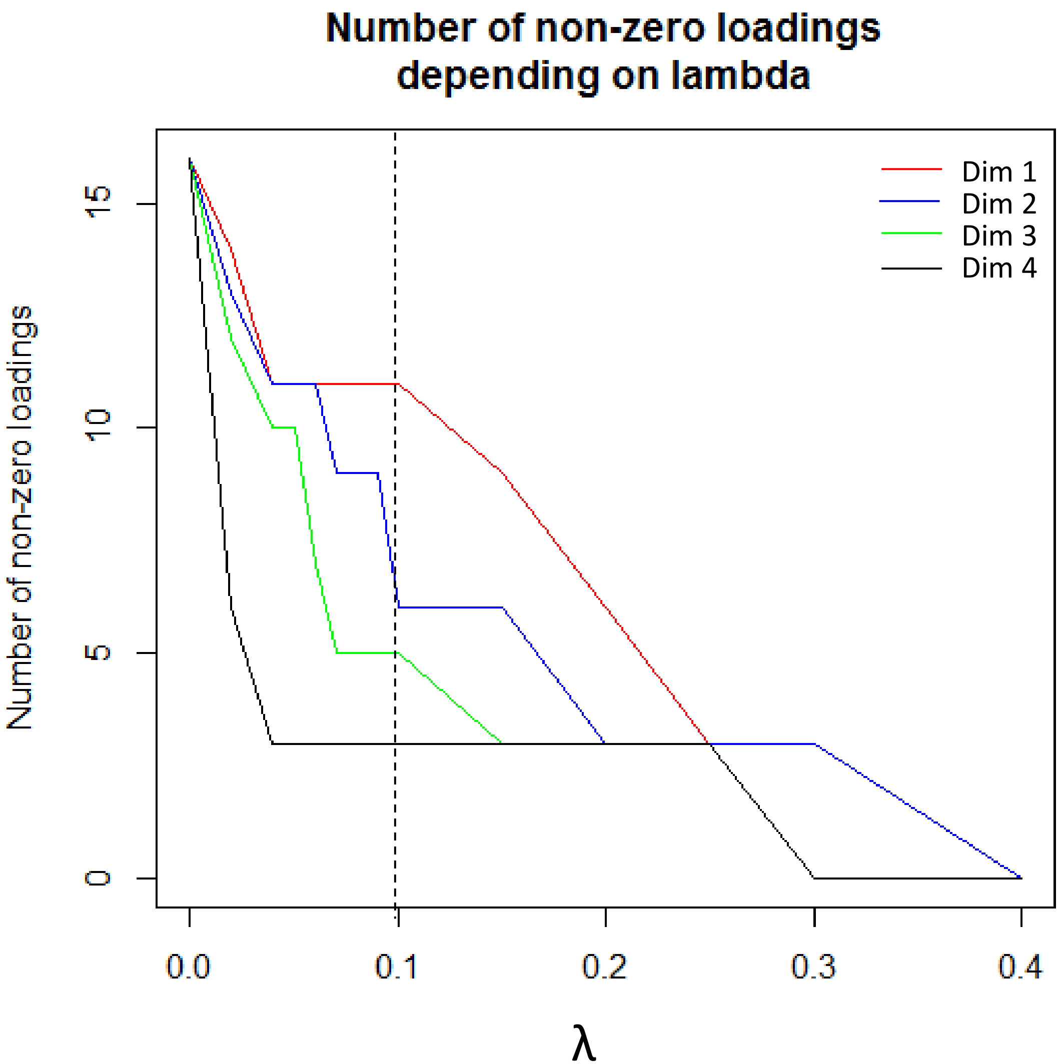
n=27 breeds of dogs
p=6 variables
q=16 (total number of columns)

X : 27 x 6 matrix of categorical variables

K : 27 x 16 complete disjunctive table → $K=(K_1, \ldots, K_6)$

1 block
= 1 K_j matrix

Toy example: Dogs



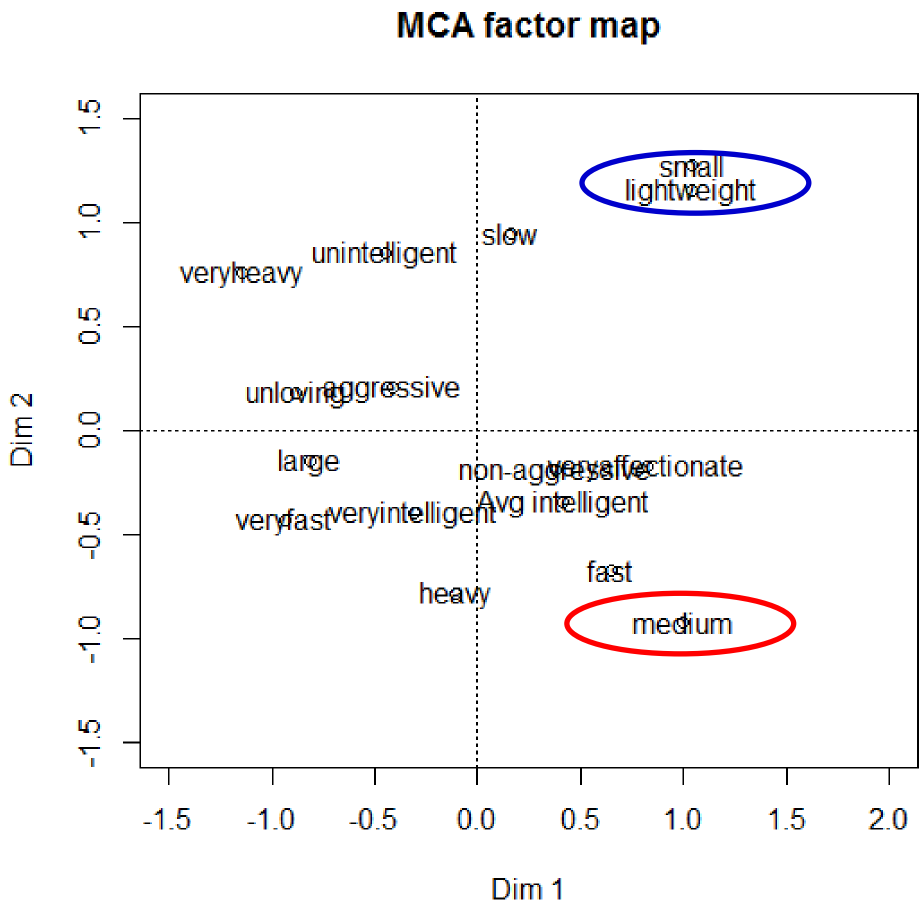
For λ=0.10:

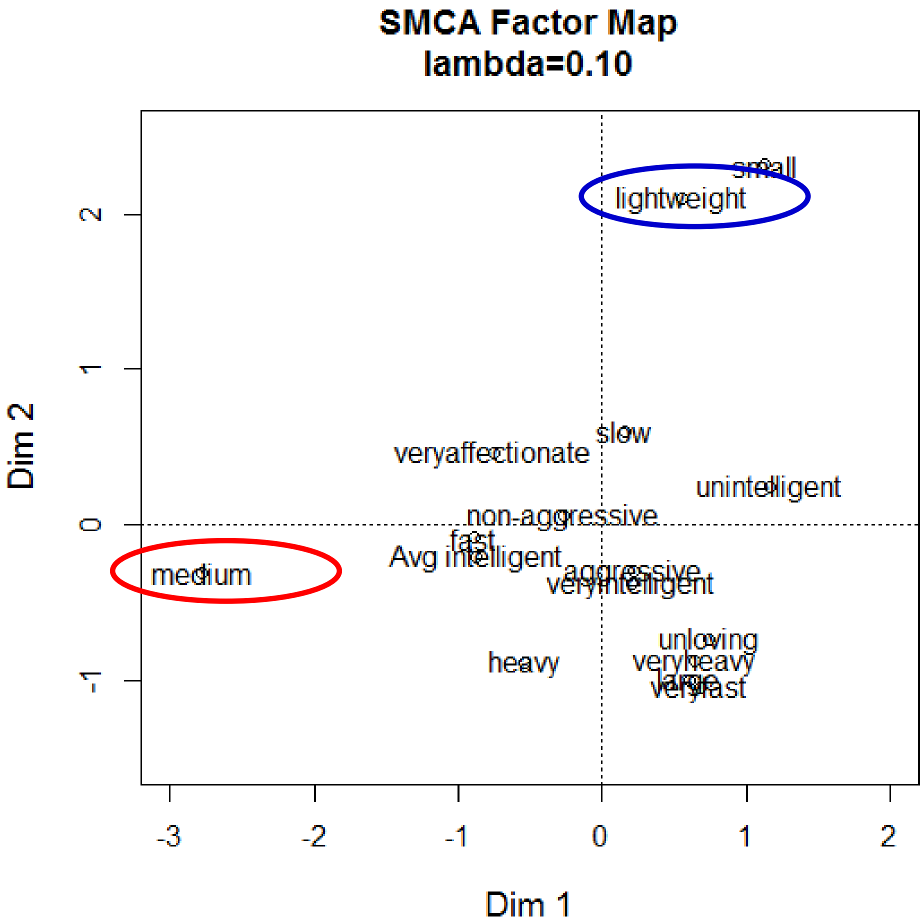
- 11 non-zero loadings on the 1st axis
- 6 non-zero loadings on the 2nd axis
- 5 non-zero loadings on the 3rd axis
- 3 non-zero loadings on the 4th axis

Toy example: Comparison of the loadings

SNPs	MCA				Sparse MCA			
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 1	Dim 2	Dim 3	Dim 4
large	**-0.270**	**0.017**	-0.072	0.060	**-0.399**	**-0.517**	**0.000**	**0.000**
medium	**0.222**	**-0.444**	0.384	-0.065	**0.808**	**0.008**	**0.000**	**0.000**
small	**0.453**	**0.402**	-0.205	-0.085	**-0.331**	**0.610**	**0.000**	**0.000**
lightweight	0.437	**0.332**	**-0.098**	-0.091	**0.000**	**0.471**	**0.278**	**0.000**
heavy	-0.061	**-0.265**	**-0.118**	0.154	**0.000**	**-0.369**	**0.426**	**0.000**
veryheavy	-0.428	**0.332**	**0.493**	-0.334	**0.000**	**-0.059**	**-0.860**	**0.000**
slow	0.070	0.297	0.285	-0.144	-0.002	**0.000**	**0.000**	**0.000**
fast	0.177	-0.269	0.065	-0.019	0.013	**0.000**	**0.000**	**0.000**
veryfast	-0.286	-0.068	-0.429	0.201	-0.011	**0.000**	**0.000**	**0.000**
unintelligent	-0.052	0.328	-0.087	**0.417**	-0.184	**0.000**	**0.000**	**-0.248**
avg intelligent	0.087	-0.140	0.255	**0.096**	0.197	**0.000**	**0.000**	**-0.488**
veryintelligent	-0.118	-0.134	-0.437	**-0.764**	-0.035	**0.000**	**0.000**	**0.836**
unloving	-0.264	0.123	-0.028	0.076	-0.040	**0.000**	-0.007	**0.000**
veryaffectionate	0.245	-0.115	0.026	-0.070	0.040	**0.000**	0.007	**0.000**
aggressive	-0.113	0.079	0.053	-0.034	**0.000**	**0.000**	**0.000**	**0.000**
non-agressive	0.105	-0.074	-0.049	0.032	**0.000**	**0.000**	**0.000**	**0.000**
#non-zero loadings	16	16	16	16	11	6	5	3
% inertia	28.19	22.79	13.45	9.55	21.37	20.81	12.04	5.88

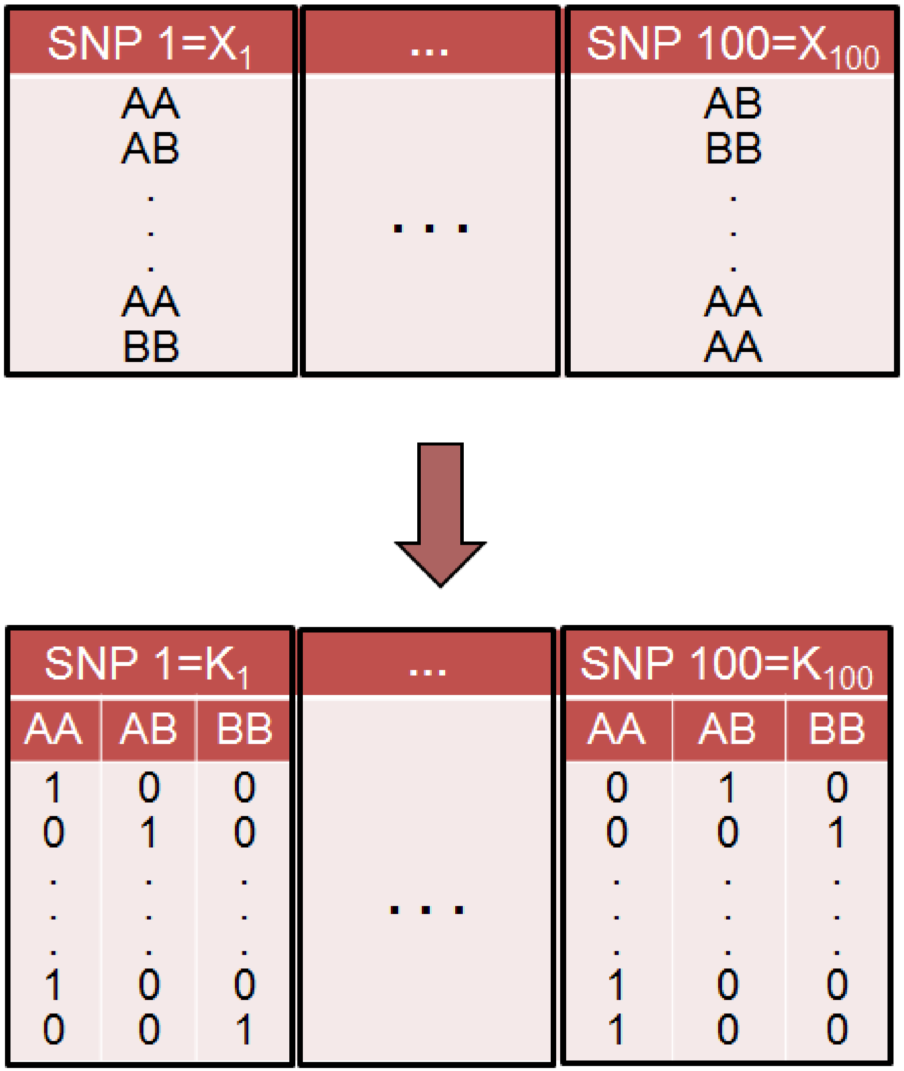
Comparison between MCA and Sparse MCA on the first plan





Application on genetic data

Single Nucleotide Polymorphisms



Data:

n=502 individuals

p=100 SNPs (among more than 800 000 of the original data base, 15000 genes)

q=281 (total number of columns)

X : 502 x 100 matrix of qualitative variables

K : 502 x 281 complete disjunctive table → $K=(K_1, \ldots, K_{100})$

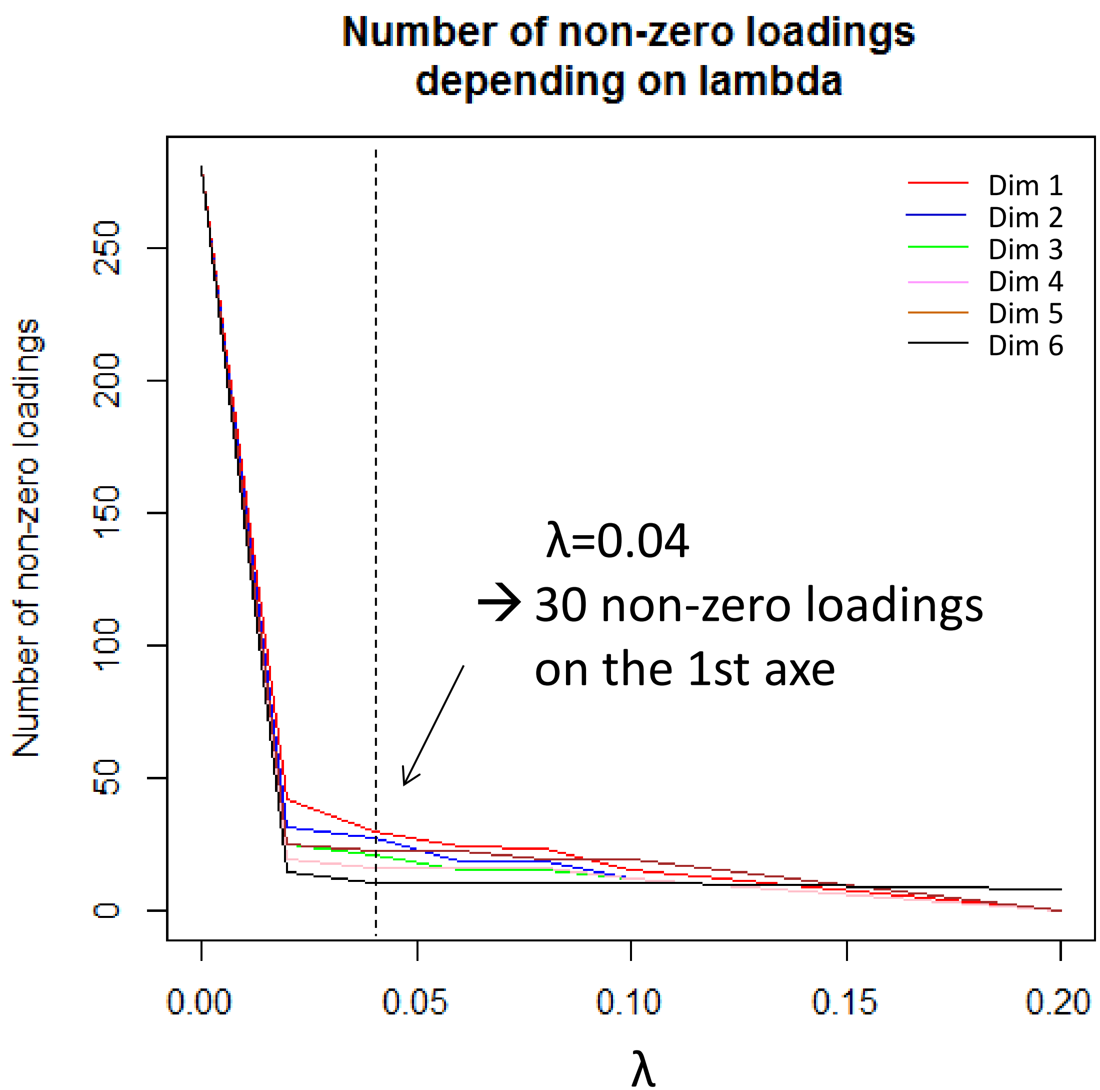
1 block

=

1 SNP = 1 K_j matrix

Application on genetic data

Single Nucleotide Polymorphisms



Application on genetic data

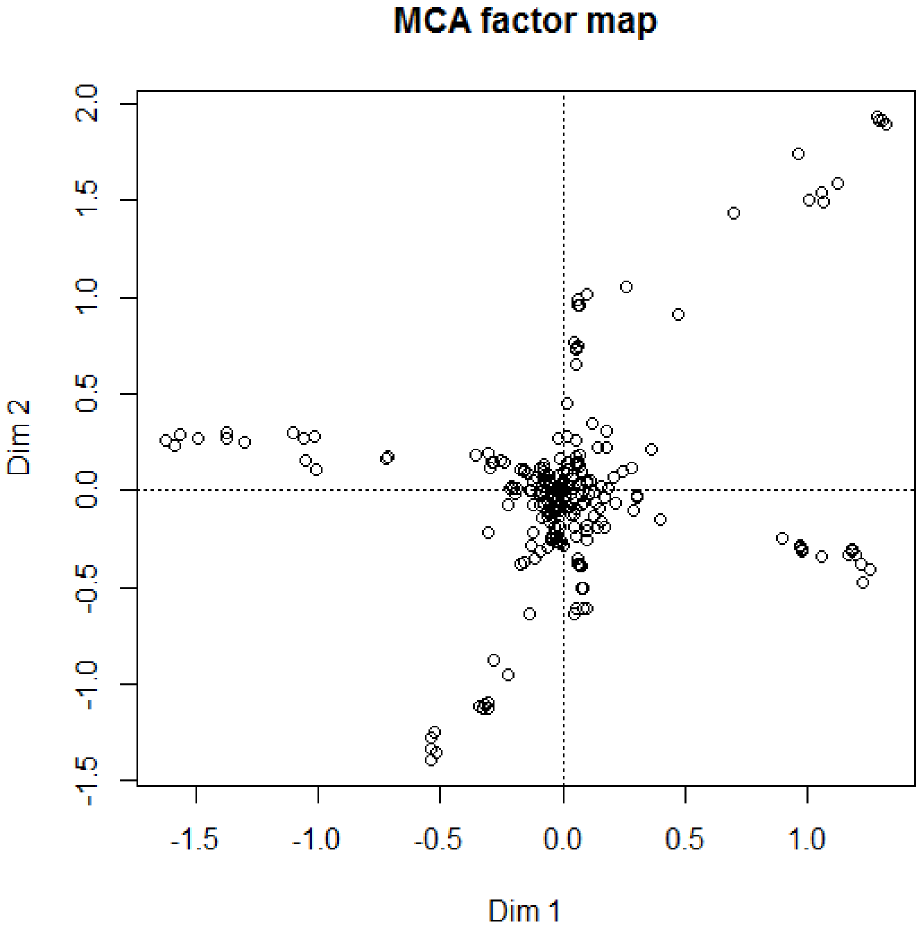
Comparison of the loadings

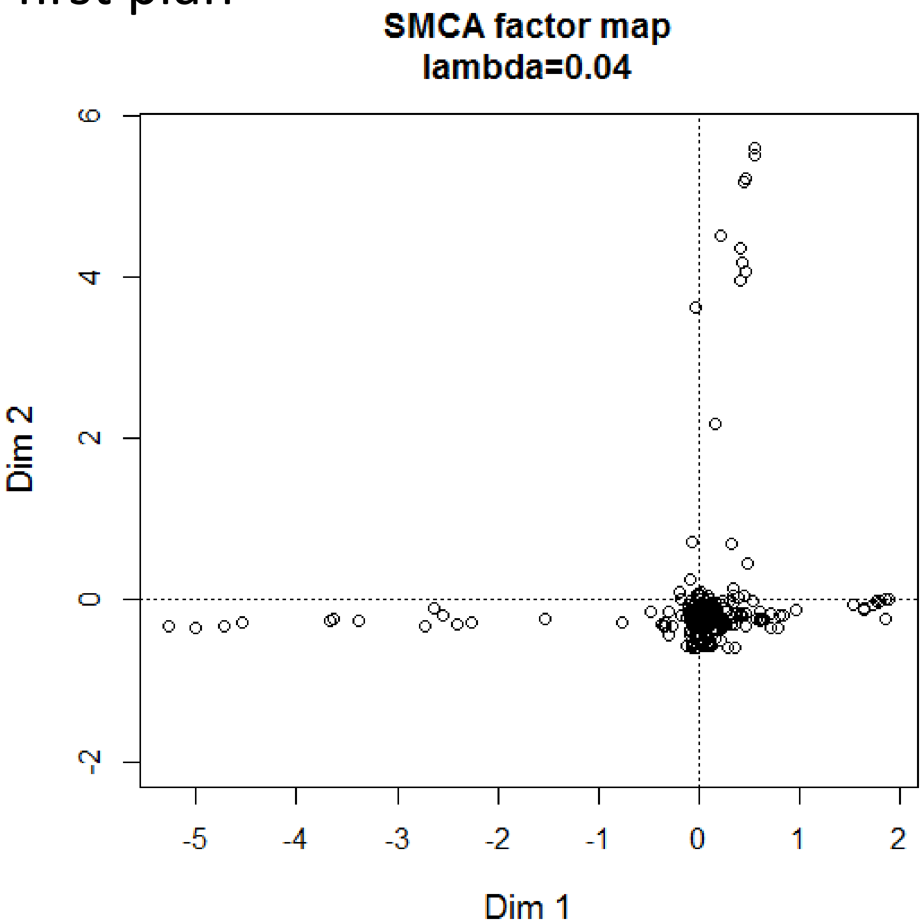
SNPs	MCA		Sparse MCA	
	Dim 1	Dim 2	Dim 1	Dim 2
rs4253711.AA	**-0.323**	-0.043	**-0.309**	**0.000**
rs4253711.AG	**0.009**	0.016	**0.057**	**0.000**
rs4253711.GG	**0.024**	-0.006	**0.086**	**0.000**
rs4253724.AA	**-0.264**	-0.025	**-0.424**	**0.000**
rs4253724.AT	**0.018**	0.014	**0.115**	**0.000**
rs4253724.TT	**0.027**	-0.008	**0.116**	**0.000**
rs26722.AG	0.054	**-0.421**	**0.000**	**-0.574**
rs26722.GG	-0.003	**0.024**	**0.000**	**0.574**
rs35406.AA	-0.002	**0.024**	**0.000**	**0.241**
rs35406.AG	0.038	**-0.388**	**0.000**	**-0.241**
⋮	⋮	⋮	⋮	⋮
#non-zero loadings	281	281	30	24
% inertia	6.86	6.73	5.03	4.95

Application on genetic data

Single Nucleotide Polymorphisms

Comparison between MCA and Sparse MCA
on the first plan





Application on genetic data

Comparison of the squared loadings

SNPs	MCA		MCA with rotation		Sparse MCA	
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
rs4253711	**0.104**	0.232	0.003	0.003	**0.106**	**0.000**
rs4253724	**0.119**	0.238	0.003	0.002	**0.206**	**0.000**
rs26722	0.001	0.003	0.003	0.003	**0.000**	**0.659**
rs35406	0.001	0.000	0.003	0.005	**0.000**	**0.115**
⋮	⋮	⋮	⋮	⋮	⋮	⋮
#of non-zero loadings	281	281	281	281	30	24
% inertia	6.86	6.73	6.73	6.46	5.03	4.95

6.Conclusions and perspectives

- Sparse techniques provide elegant and efficient solutions to problems posed by high-dimensional data
  – A new generation of data analysis with few restrictive hypothesis
- We proposed 2 new methods in a unsupervised multiblock data context: **Group Sparse PCA** for continuous variables, and **Sparse MCA** for categorical variables
  – Both methods produce sparse loadings structures that makes easier the interpretation and the comprehension of the results

- However these methods do not yield sparsity within groups
- **Research in progress**:
  - Criteria for choosing the tuning parameter λ
  - Extension of Sparse MCA : compromise between the Sparse MCA and the sparse group lasso developed by Simon et al. (2002)
    - select groups and predictors within a group, in order to produce sparsity at both levels

Thank you !

References

- Chun, H. and Keles, S. (2010), Sparse partial least squares for simultaneous dimension reduction and variable selection, *Journal of the Royal Statistical Society - Series B*, Vol. 72, pp. 3–25.
- Hastie T., Tibshirani R., Friedman J. (2009) *The elements of statistical learning*, 2nd edition, Springer, 2009
- Jolliffe, I.T. , Trendafilov, N.T. and Uddin, M. (2003) A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics,* **12,** 531–547,
- Rousson, V. , Gasser, T. (2004), Simple component analysis. *Journal of the Royal Statistical Society*: *Series C (Applied Statistics),* **53**,539-555
- Shen, H. and Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99:1015-1034.
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2012) A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics,*

- Tenenhaus M. (1998) *La régression PLS*, Technip
- Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267-288
- Vines, S.K., (2000) Simple principal components, *Journal of the Royal Statistical Society*: *Series C (Applied Statistics)*, **49**, 441-451
- Yuan, M., Lin, Y. (2007) Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, **68**, 49-67,
- Zou, H., Hastie , T. (2005) Regularization and variable selection via the elastic net. *Journal of Computational and Graphical Statistics*, **67**, 301-320,
- Zou, H., Hastie, T. and Tibshirani, R. (2006) Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, **15**, 265-286.
- H. Zou, T. Hastie, R. Tibshirani, (2007), On the “degrees of freedom” of the lasso, *The Annals of Statistics*, 35, 5, 2173–2192.