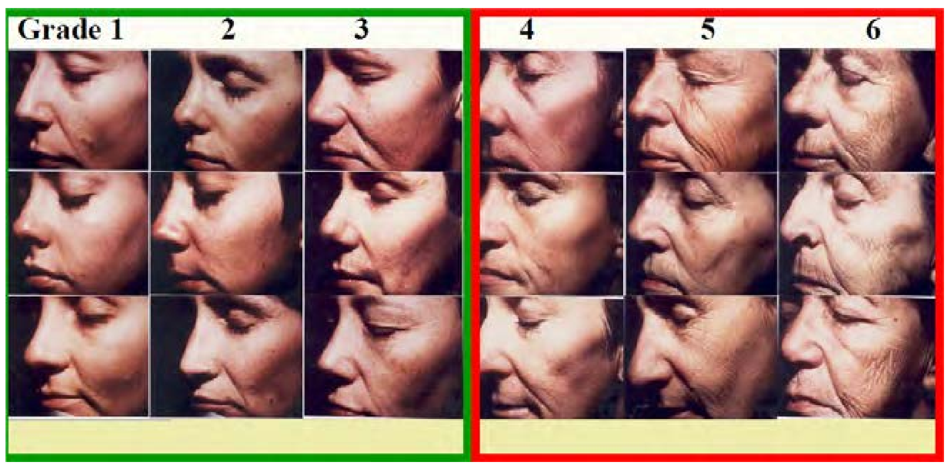
From Sparse Regression to Sparse Multiple Correspondence Analysis

Gilbert Saporta
CEDRIC, CNAM, Paris
gilbert.saporta@cnam.fr

- Joint work with Anne Bernard, Ph.D student funded by the R&D department of Chanel cosmetic company
- Industrial context and motivation:
  – Relate gene expression data to skin aging measures



  – n=500, p= 800 000 SNP’s, 15 000 genes

Outline

1. Introduction
2. Regularized 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 interpretation 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. Regularized regression

- No OLS solution when p>n
- A special case of multicollinearity
- Usual regularized regression techniques:
  - Component based: PCR, PLS
  - Ridge

2.1 Principal components regression

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

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) :

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

- Easy solution:
  - w_j proportional to $\mathrm{cov}(y; x_j)$
  - 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

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

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

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

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

- Shrinkage properties (Hastie et al. , 2009)
  - PCR discards low variance directions
  - PLS shrinks low variance directions but inflates high variance directions
  - Ridge shrinks all principal directions but shrinks more low variance directions
- Lost properties:
  - Bias, scale invariance

    need standardised data

3. Sparse regression

- Keeping all predictors is a drawback for high dimensional data: combinations of too many variables cannot be interpreted
- Sparse methods simultaneously shrink coefficients and select variables, hence better predictions

3.1 Lasso and elastic-net

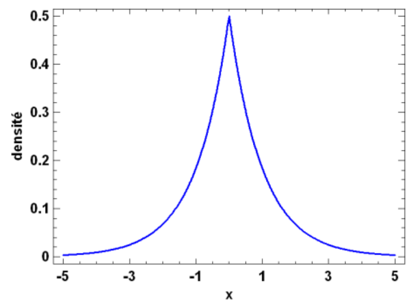
- Lasso (Tibshirani, 1996) imposes a L_1 constraint on the coefficients $\sum_{j=1}^{p} |b_j| < c$

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

- 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 an 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

- Finding the optimal parameter
  - Cross validation if optimal prediction is needed
  - BIC when the sparsity is the main concern

$$\lambda_{opt} = \arg\min_{\lambda} \left(\frac{\left\| y - X\hat{\beta}(\lambda) \right\|^2}{n\sigma^2} + \frac{\log(n)}{n} \hat{d}f(\lambda) \right)$$

a good unbiased estimate of df is the number of nonzero coefficients . (Zou et al., 2007)

- A more 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)

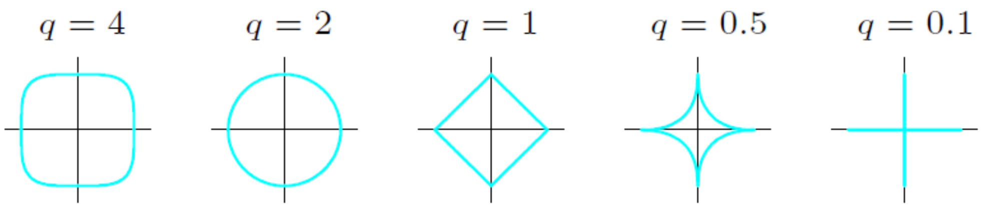


FIGURE 3.12. *Contours of constant value of* $\sum_j |\beta_j|^q$ *for given values of* q.

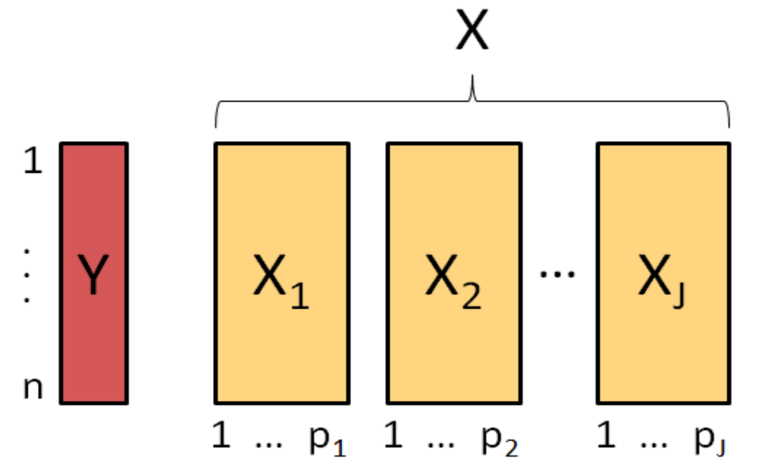
Elements of Statistical Learning (2nd Ed.) ©Hastie, Tibshirani & Friedman 2009 Chap 3

- Lasso produces a sparse model but the number of selected variables 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)

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

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)

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

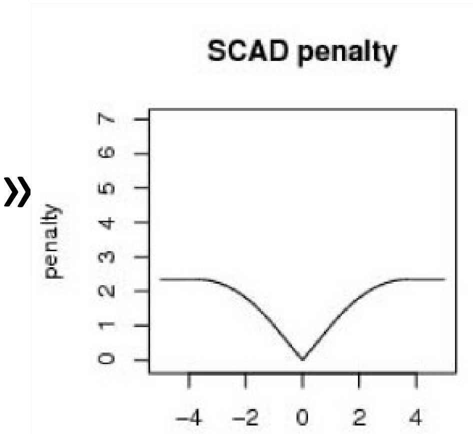
  – Two tuning parameters: grid search

3.3 Other sparse regression methods

- SCAD penalty (Fan & Li, 2001)

  « 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:

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

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

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

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 i^{th} 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.)

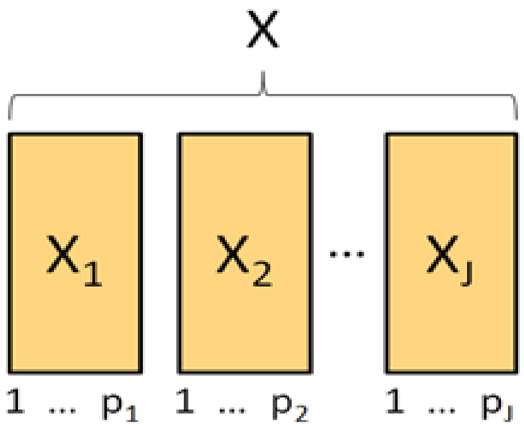
$$\mathbf{X}^{(k)} = \sum_{j=1}^{k} d_j \mathbf{u}_j \mathbf{v}_j^{'}$$

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

- 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



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

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

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.

5.Sparse MCA

In MCA:

Original table

X_J
1
p_J
.
.
.
3

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

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

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

And $\tilde{\mathbf{F}}$ 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	$\lambda_j / tot \times 100$	$\left\Vert \tilde{\mathbf{Z}}_{\mathbf{j.1,\ldots,j\text{-}1}} \right\Vert^2$
Total inertia	$\frac{1}{p}\sum_{j=1}^{p} p_j - 1$	$\sum_{j=1}^{k} \left\Vert \tilde{\mathbf{Z}}_{\mathbf{j.1,\ldots,j\text{-}1}} \right\Vert^2$

$\tilde{\mathbf{Z}}_{\mathbf{j.1,\ldots,j\text{-}1}}$ 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:

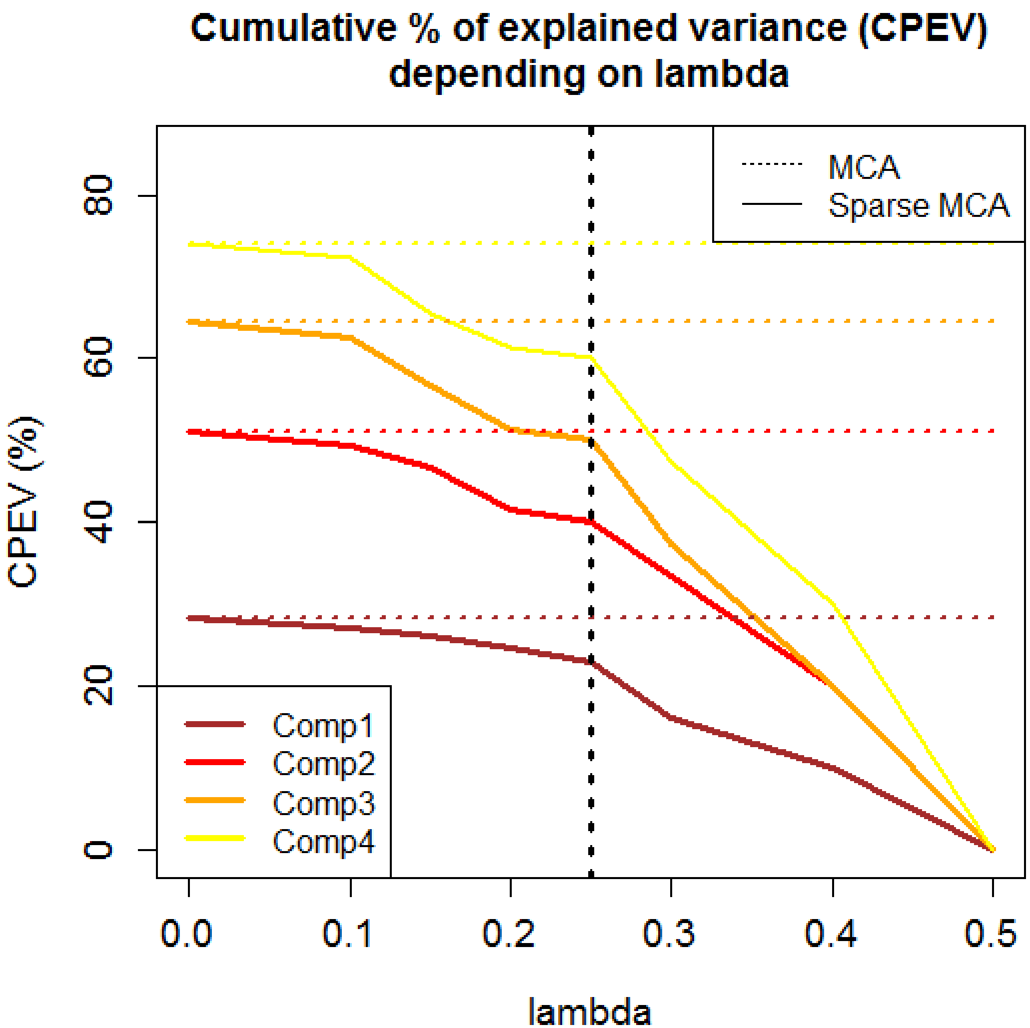
n=27 breeds of dogs
p=6 variables
q=16 (total number of columns)

X : 27 x 6 matrix of categorical variables

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

1 block
= 1 K_j matrix

Toy example: Dogs



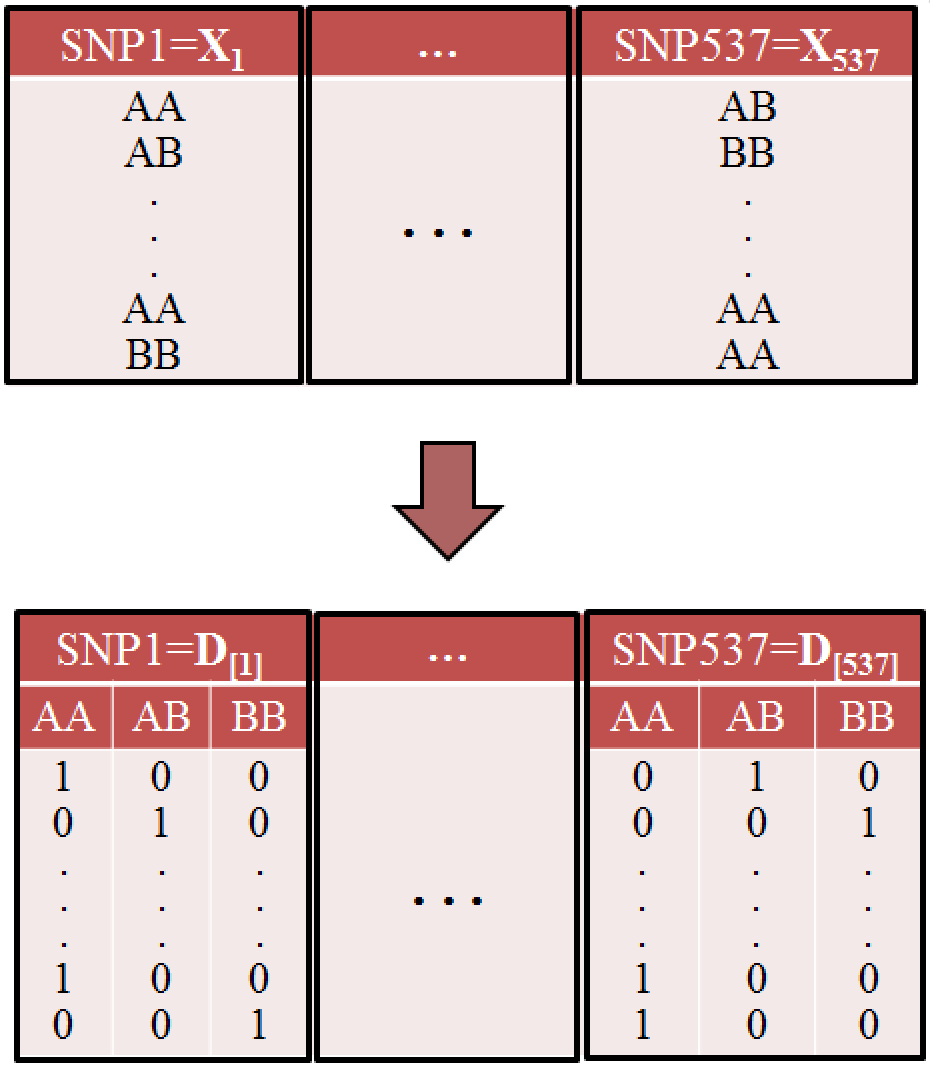
λ= 0:25 is a compromise between the number of variables selected and the % of variance lost.

Toy example: Comparison of the loadings

Variable	MCA Comp1	MCA Comp2	MCA Comp3	MCA Comp4	Sparse MCA Comp1	Sparse MCA Comp2	Sparse MCA Comp3	Sparse MCA 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

Application on genetic data

Single Nucleotide Polymorphisms



Data:

n=502 individuals

p=537 SNPs (among more than 800 000 of the original data base, 15000 genes)

q=1554 (total number of columns)

X : 502 x 537matrix of qualitative variables

K : 502 x 1554 complete disjunctive table → $K=(K_1, \ldots, K_{1554})$

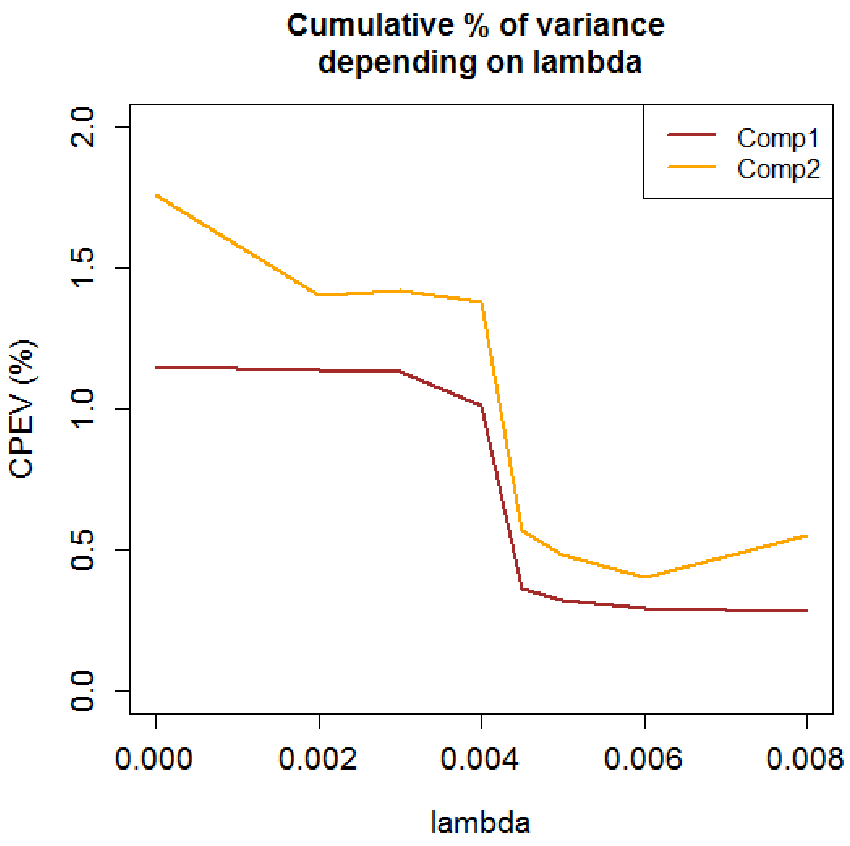
1 block

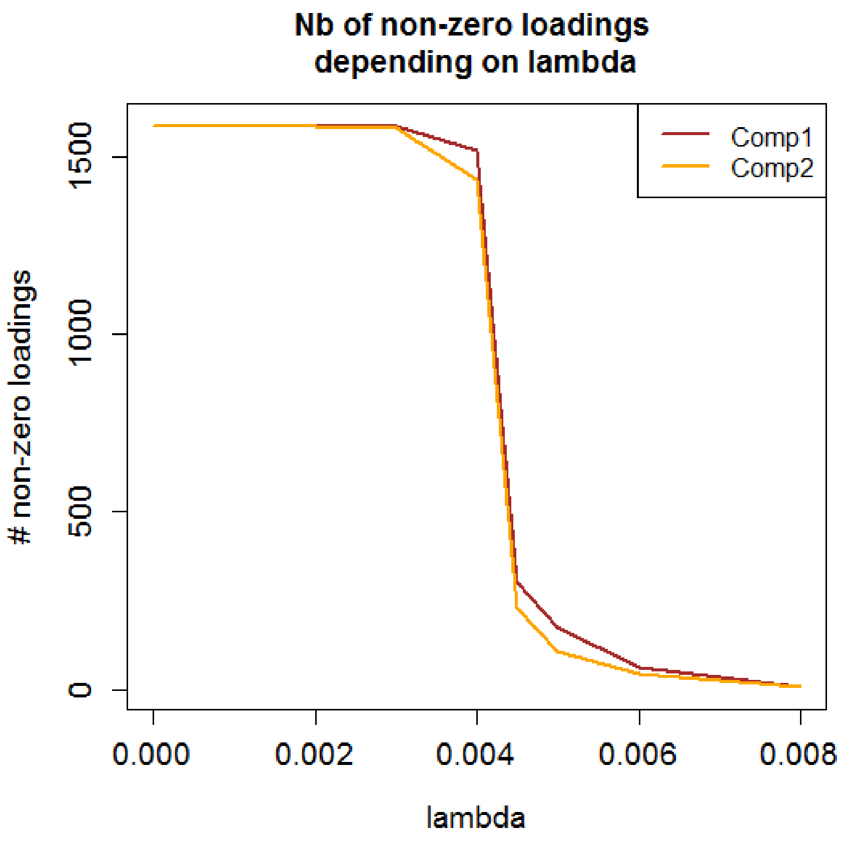
=

1 SNP = 1 K_j matrix

Application on genetic data

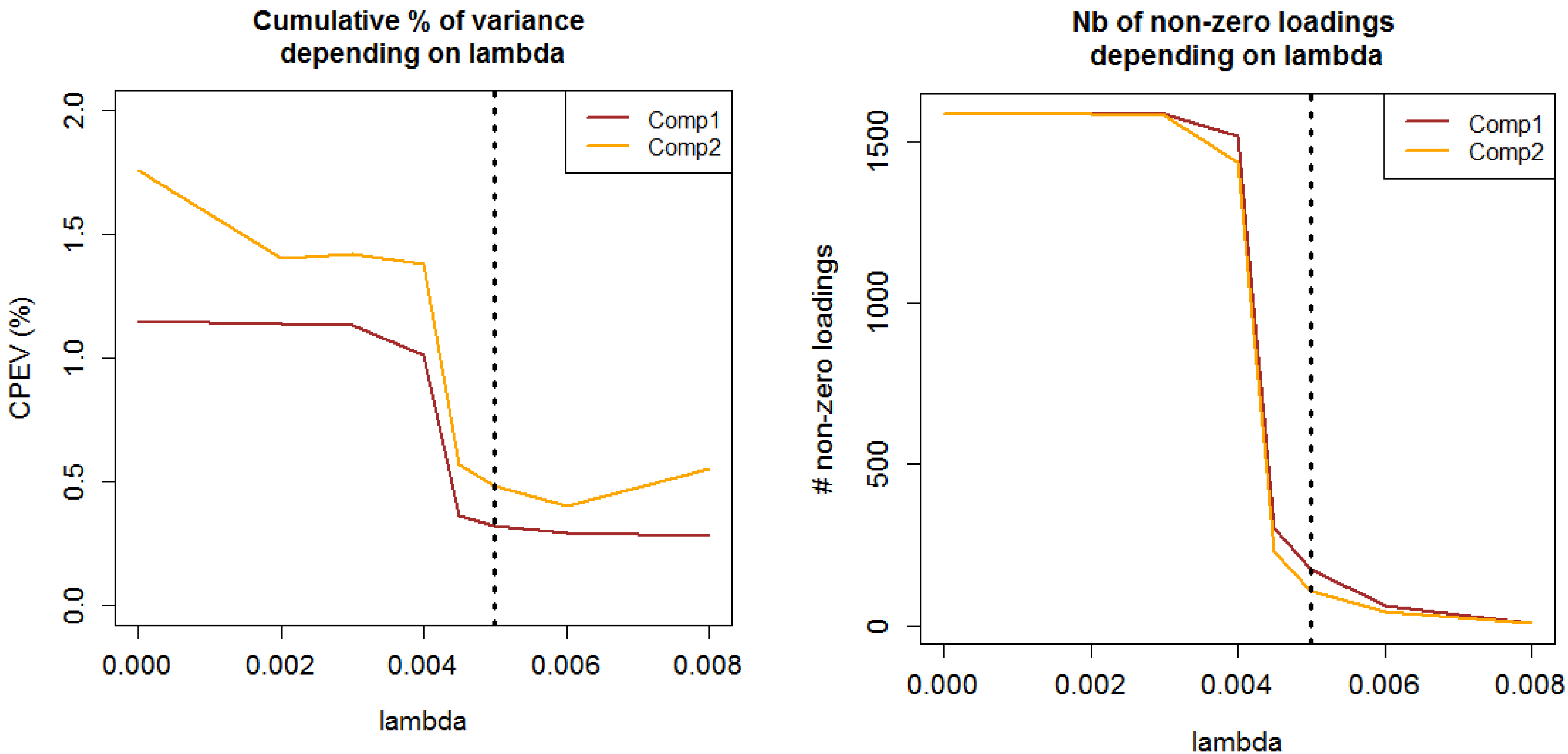
Single Nucleotide Polymorphisms





Application on genetic data

Single Nucleotide Polymorphisms



λ= 0:005: CPEV= 0:32% and 174 columns selected on Comp 1

Application on genetic 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

6.Conclusions and perspectives

- Sparse techniques provide elegant and efficient solutions to problems posed by high-dimensional data:
  - A new generation of data analysis methods with few restrictive hypothesis
- Very powerful in a context of variable selection in high dimension issues:
  - reduce noise as well as computation time.

- 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
  - Possibility of selecting superblocks (genes)
- **Research in progress**:
  - Extension of Sparse MCA to select groups and predictors within a group (sparsity within groups)
    - sparsity at both group and individual feature levels
    - compromise between Sparse MCA and sparse group lasso developed by Simon et al. (2012).

Thanks for your attention

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.
- Fan, J. and Li, R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of American Statistical Association*, **96,** 1348-1360, 2001
- 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.

PLS 2014

8th International Conference on Partial Least Squares and Related Methods

26-28 May 2014 Paris, France

http://www.pls14.org/

- **Topics:**

PLS Regression, PLS Path Modeling and their related methods with application in Management, Social Sciences, Chemometrics, Sensory Analysis, Industry and Life Sciences including genomics.

- **Keynote speakers**

Anne-Laure BOULESTEIX
LMU München - Germany
Peter BÜHLMANN
ETH Zürich - Switzerland
Mohamed HANAFI
ONIRIS, Nantes-France