

le cnam

Sparse Correspondence Analysis

Gilbert Saporta

CEDRIC,

Conservatoire National des Arts et Métiers, Paris

gilbert.saporta@cnam.fr

A joint work with:

Ruiping Liu (Beihang University, Beijing)

Ndeye Niang (CNAM, Paris)

Huiwen Wang (Beihang University, Beijing)

Outline

1. Introduction
2. Sparse PCA
3. Sparse MCA
4. Sparse CA
5. Conclusion and perspectives

1.Introduction

- Following the Lasso in regression, several sparse versions of PCA and related methods have been proposed. (Hastie et al, 2015)
- Components are combinations of only a few variables: most loadings are zero.
- Useful in the context of high dimensional data: $p \gg n$
 - Easier interpretation (simple structure)
 - Robustness of solutions

- Correspondence analysis is a versatile method for categorical data. At the same time:
 - A PCA with weights and a special metric
 - A particular SVD
 - A canonical correlation analysis
- CA for contingency tables, MCA for more than 2 categorical variables
- Sparse CA: a particular case of sparse PCA, sparse SVD or sparse CCA

2.Sparse PCA

- In **PCA**, each PC is a linear combination of **all** the original variables : difficult to interpret the results when p is large
- **Objective of SPCA**: obtain components easily interpretable (many 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, ...)

2.1 First attempts:

- **Simple PCA**

- Hausman (1982) loadings -1,0,1
- Vines (2000) : integer loadings
- Generalized by Rousson, V. and Gasser, T. (2004) : blocks of loadings (+ , 0, -)

Example

Table 1. Definitions of Variables in Jeffers' Pitprop Data

<i>Variable</i>	<i>Definition</i>
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



PCA

.40	-.22	.21	-.09	-.08	.12
.41	-.19	.24	-.10	-.11	.16
.12	-.54	-.14	.08	.35	-.28
.17	-.46	-.35	.05	.36	-.05
.06	.17	-.48	.05	.18	.63
.28	.01	-.48	-.06	-.32	.05
.40	.19	-.25	-.06	-.22	.00
.29	.19	.24	.29	.19	-.06
.36	-.02	.21	.10	-.11	.03
.38	.25	.12	-.21	.16	-.17
.01	.21	-.07	-.80	.34	-.18
.12	.34	.09	.30	.60	.17
.11	.31	-.33	.30	-.08	-.63

SCA

.45	0	0	0	0	0
.45	0	0	0	0	0
0	0	.71	0	0	0
0	0	.71	0	0	0
0	.5	0	0	0	.71
0	.5	0	0	0	0
0	.5	0	0	0	0
.45	0	0	0	0	0
.45	0	0	0	0	0
.45	0	0	0	0	0
0	0	0	0	1	0
0	0	0	1	0	0
0	.5	0	0	0	-.71

Rousson , Gasser 2003

- **SCoTLASS** (Simplified Component Technique – Lasso) by Jolliffe & al. (2003) : extra L_1 constraints

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

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

$t < 1$ no solution

$t = 1$ only one nonzero coefficient

$$1 < t < \sqrt{p}$$

Non convex problem

2.2 S-PCA by Zou et al (2006)


Let the SVD of \mathbf{X} be $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$ with $\mathbf{Z} = \mathbf{U}\mathbf{D}$ the principal components
PCA can be written as a **ridge regression** between the principal components and the original variables

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

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

Add a L_1 penalty to produce sparse loadings:

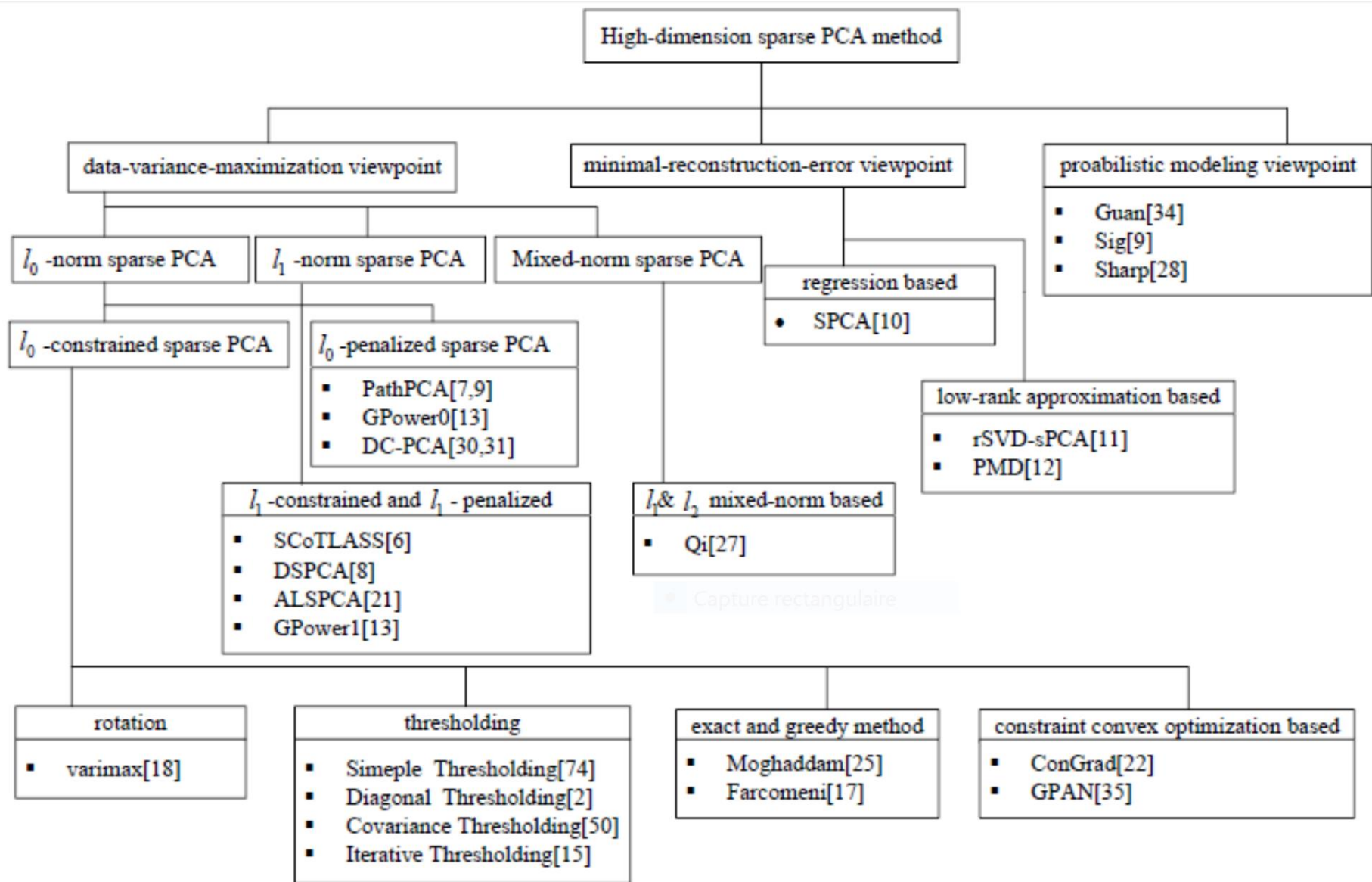
$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|^2 + \lambda_1 \|\boldsymbol{\beta}\|_1$$

Lasso
penalty 

May be solved by an alternated algorithm
between elastic net and SVD

2.3 Many other techniques

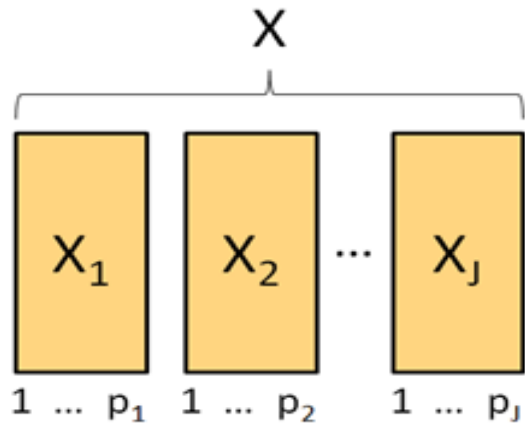
- Trendafilov, 2014: “ the existing methods to obtain sparse component loadings are so many and diverse, and their number constantly increases “
- Instead of finding a « good » value of the sparse parameter, prespecify the number of zeros (Adachi & Trendafilov 2015) ie a L_0 norm constraint



2.4 Lost properties

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

3. Sparse MCA (Bernard et al, 2012)



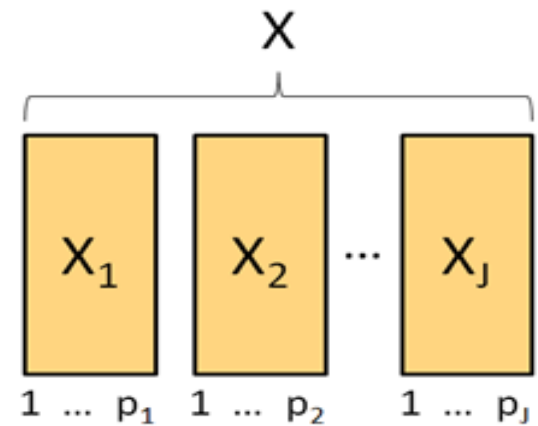
Disjunctive table of J categorical variables

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 .

An other approach: prespecified sparsity , Mori et al, 2016

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

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

by that defined in the group Lasso (Simon et al. 2013)

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

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)

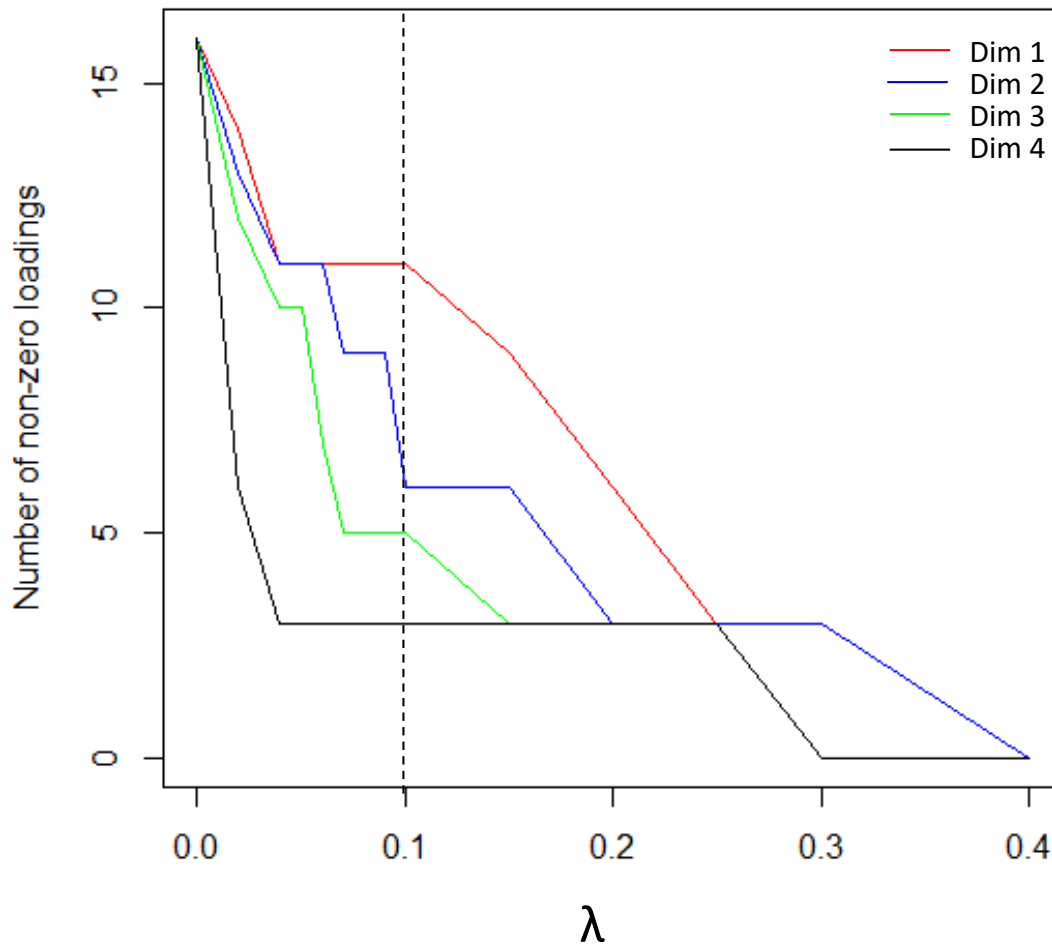
X : 27×6 matrix of categorical variables

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

1 block
= 1 K_j matrix

Toy example: Dogs

Number of non-zero loadings depending on lambda



For $\lambda=0.10$:

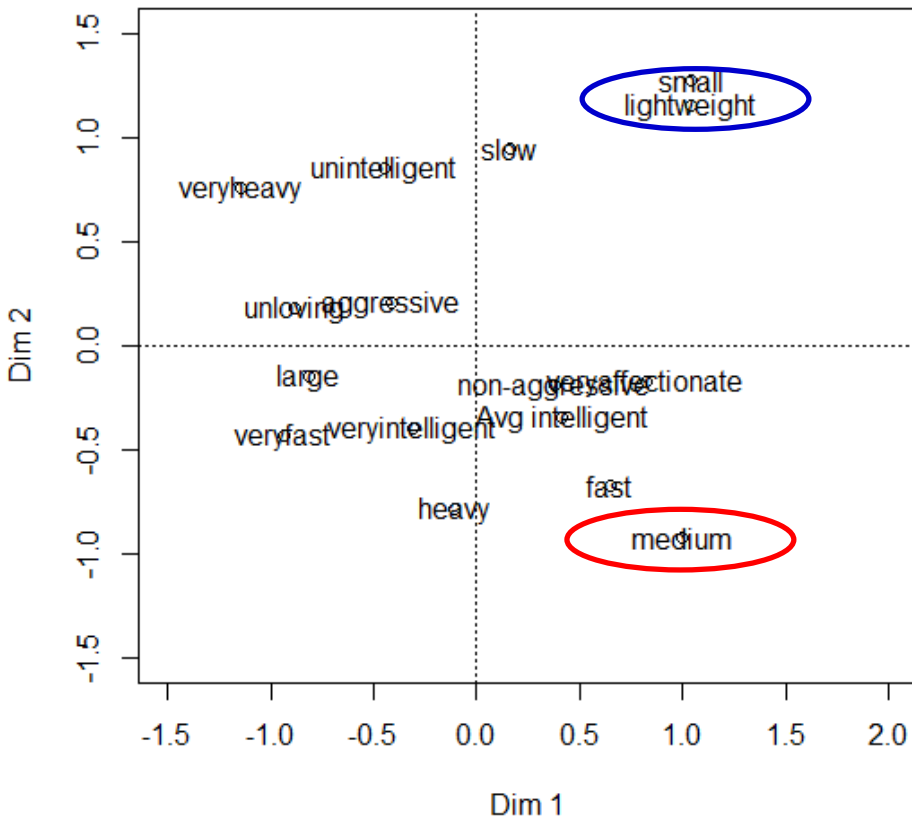
- 11 non-zero loadings on the 1st axis
- 6 non-zero loadings on the 2nd axis
- 5 non-zero loadings on the 3rd axis
- 3 non-zero loadings on the 4th axis

Toy example: Comparison of the loadings

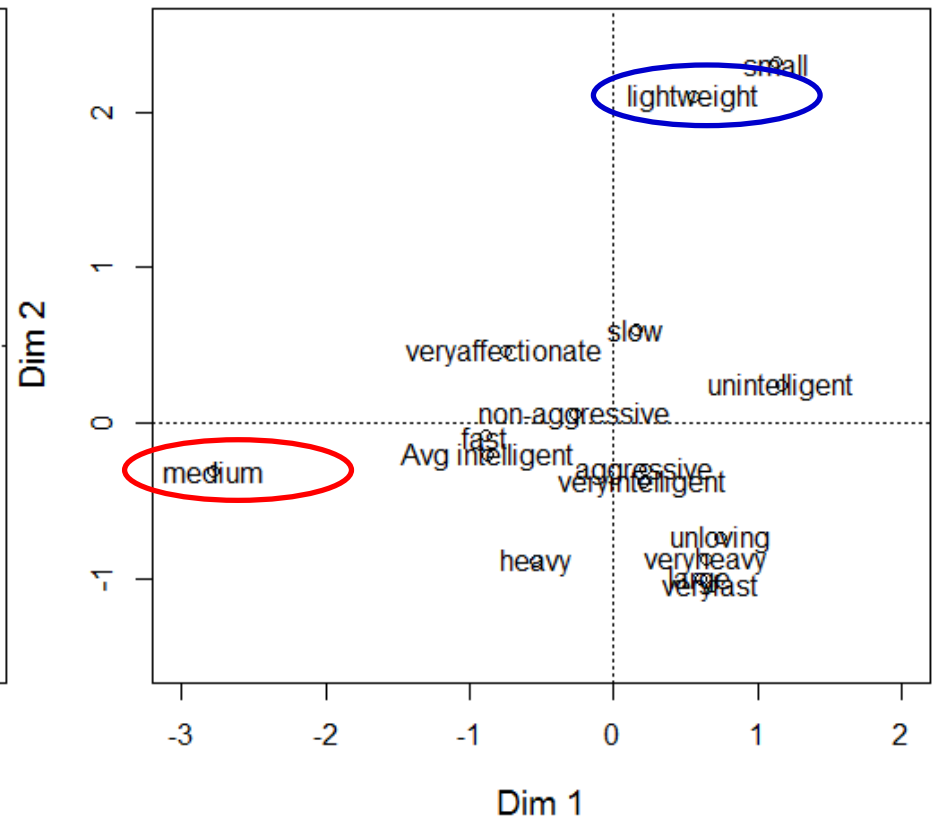
SNPs	MCA				Sparse MCA			
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 1	Dim 2	Dim 3	Dim 4
large	-0.270	0.017	-0.072	0.060	-0.399	-0.517	0.000	0.000
medium	0.222	-0.444	0.384	-0.065	0.808	0.008	0.000	0.000
small	0.453	0.402	-0.205	-0.085	-0.331	0.610	0.000	0.000
lightweight	0.437	0.332	-0.098	-0.091	0.000	0.471	0.278	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

MCA factor map



SMCA Factor Map
lambda=0.10



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

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

4. Sparse Correspondence Analysis

- For contingency tables, CA is both
 - a PCA (with weights and the chi squared metric)
 - a CCA between rows and columns indicator variables
- Sparse PCA of row profiles leads to **column sparse CA** (same for columns)
 - Useful for contingency tables with many columns like documents-terms matrix
 - But symmetry is lost
- **Doubly sparse CA**: an application of sparse CCA or of sparse SVD

4.1 Sparse SVD and sparse CCA

- A rank 1 sparse SVD or **Penalized Matrix Decomposition** (PMD) may be defined as (Witten et al, 2009)

$$\min \|\mathbf{X} - d\mathbf{u}\mathbf{v}'\|_F^2 \text{ subject to } \|\mathbf{u}\|^2 = \|\mathbf{v}\|^2 = 1, \\ P(\mathbf{u}) \leq \alpha, P(\mathbf{v}) \leq \beta, d \geq 0$$

Where P is a penalty function (usually L_1)

- Equivalent formulation:

$$\max \mathbf{u}'\mathbf{X}\mathbf{v} \text{ subject to } \|\mathbf{u}\|^2 \leq 1, \|\mathbf{v}\|^2 \leq 1, \\ P(\mathbf{u}) \leq \alpha, P(\mathbf{v}) \leq \beta$$

- Sparse canonical analysis of \mathbf{X} and \mathbf{Y}

$$\max \mathbf{u}'\mathbf{X}'\mathbf{Y}\mathbf{v} \text{ subject to } \|\mathbf{u}\|^2 \leq 1, \|\mathbf{v}\|^2 \leq 1, \\ P(\mathbf{u}) \leq \alpha, P(\mathbf{v}) \leq \beta$$

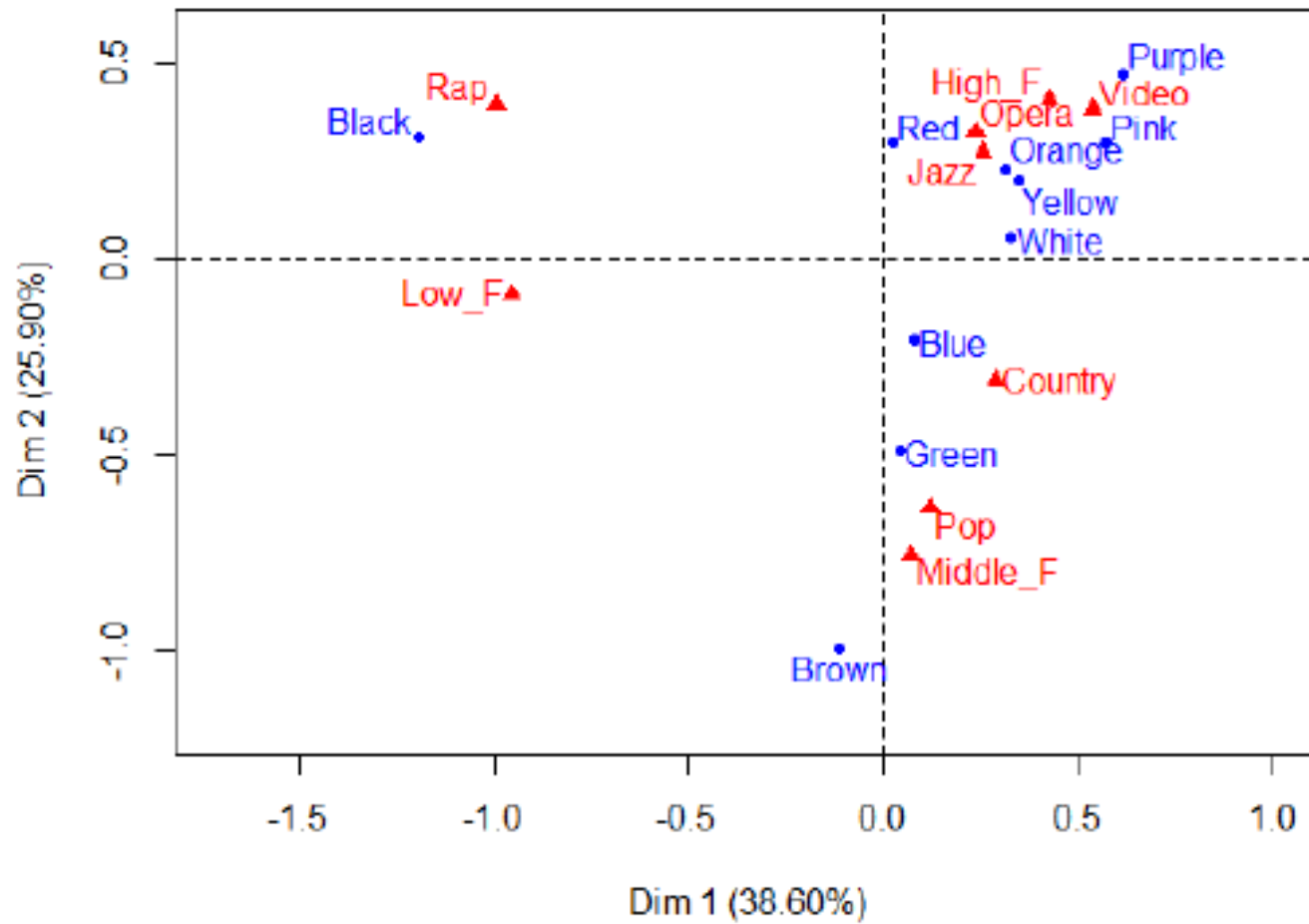
- R package PMA implements sparse PCA and sparse CCA

4.2 A toy example: colours of sound

Color	Video	Jazz	Country	Rap	Pop	Opera	Low F	High F	Middle F	x_{i+}	\mathbf{r}
Red	4	2	4	4	1	2	2	4	1	24	0.121
Orange	3	4	2	2	1	1	0	3	2	18	0.091
Yellow	6	4	5	2	3	1	1	3	0	25	0.126
Green	2	0	5	1	3	3	3	1	5	23	0.116
Blue	2	5	0	1	4	1	2	1	3	19	0.096
Purple	3	3	1	0	0	3	0	2	1	13	0.066
White	0	0	0	0	1	4	1	5	3	14	0.071
Black	0	2	0	11	1	3	10	1	1	29	0.146
Pink	2	1	1	0	2	4	0	2	0	12	0.061
Brown	0	1	4	1	6	0	3	0	6	21	0.106
x_{+j}	22	22	22	22	22	22	22	22	22	$N = 198$	1.000
\mathbf{c}^T	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11		

Abdi H., Béra M. (2017) Correspondence Analysis
 In: Alhajj R., Rokne J. (eds) Encyclopedia of Social Network Analysis and Mining,
 Springer, New York,

CA factor map



CA rows and columns coordinates and contributions

	F1	F2	F.ctr1	F.ctr2		G1	G2	G.ctr1	G.ctr2
Red	-0.026	-0.299	0	56	Video	-0.541	-0.386	113	86
Orange	-0.314	-0.232	31	25	Jazz	-0.257	-0.275	25	44
Yellow	-0.348	-0.202	53	27	Country	-0.291	0.309	33	55
Green	-0.044	0.490	1	144	Rap	0.991	-0.397	379	91
Blue	-0.082	0.206	2	21	Pop	-0.122	0.637	6	234
Purple	-0.619	-0.475	87	77	Opera	-0.236	-0.326	22	61
White	-0.328	-0.057	26	1	Low_F	0.954	0.089	351	5
Black	1.195	-0.315	726	75	High_F	-0.427	-0.408	70	96
Pink	-0.570	-0.300	68	28	Middle_F	-0.072	0.757	2	330
Brown	0.113	0.997	5	545					

CA rows and columns coordinates and contributions

	F1	F2	F.ctr1	F.ctr2
Red	-0.026	-0.299	0	56
Orange	-0.314	-0.232	31	25
Yellow	-0.348	-0.202	53	27
Green	-0.044	0.490	1	144
Blue	-0.082	0.206	2	21
Purple	-0.619	-0.475	87	77
White	-0.328	-0.057	26	1
Black	1.195	-0.315	726	75
Pink	-0.570	-0.300	68	28
Brown	0.113	0.997	5	545

	G1	G2	G.ctr1	G.ctr2
Video	-0.541	-0.386	113	86
Jazz	-0.257	-0.275	25	44
Country	-0.291	0.309	33	55
Rap	0.991	-0.397	379	91
Pop	-0.122	0.637	6	234
Opera	-0.236	-0.326	22	61
Low_F	0.954	0.089	351	5
High_F	-0.427	-0.408	70	96
Middle_F	-0.072	0.757	2	330

- Column sparse CA

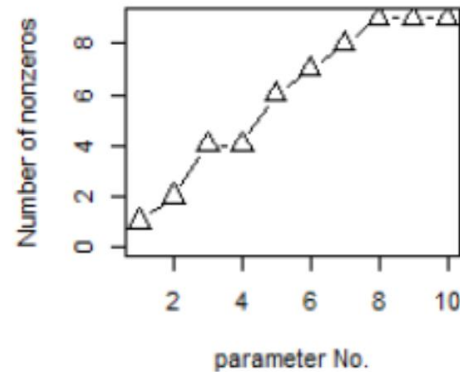
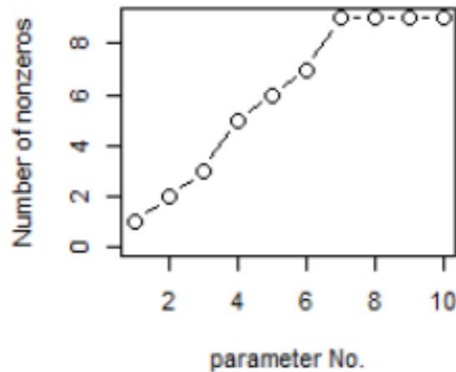
	G1	G2	G.ctr1	G.ctr2
Video	-0.263	-0.478	31	178
Jazz	0	-0.002	0	0
Country	0	0	0	0
Rap	1.097	-0.025	549	0
Pop	0	0.396	0	122
Opera	0	0	0	0
Low_F	0.954	0	415	0
High_F	-0.102	0	5	0
Middle_F	0	0.947	0	699
			1000	999

- Both sides sparsity
 - In PMD function
sumabsu and sumabsv are equal to

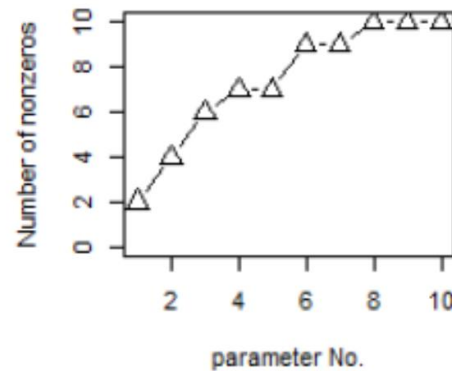
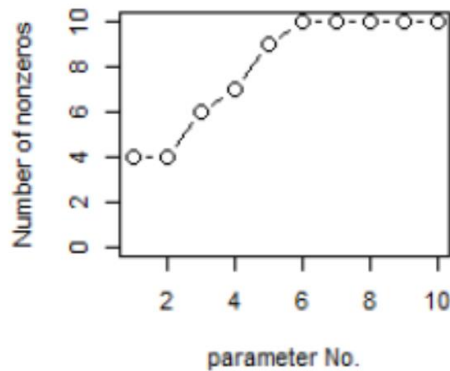
 - The smaller they are, the sparser u or v will be.

– Choosing the sparsity parameter

- According to the number of non zeros for each row and column loadings



columns



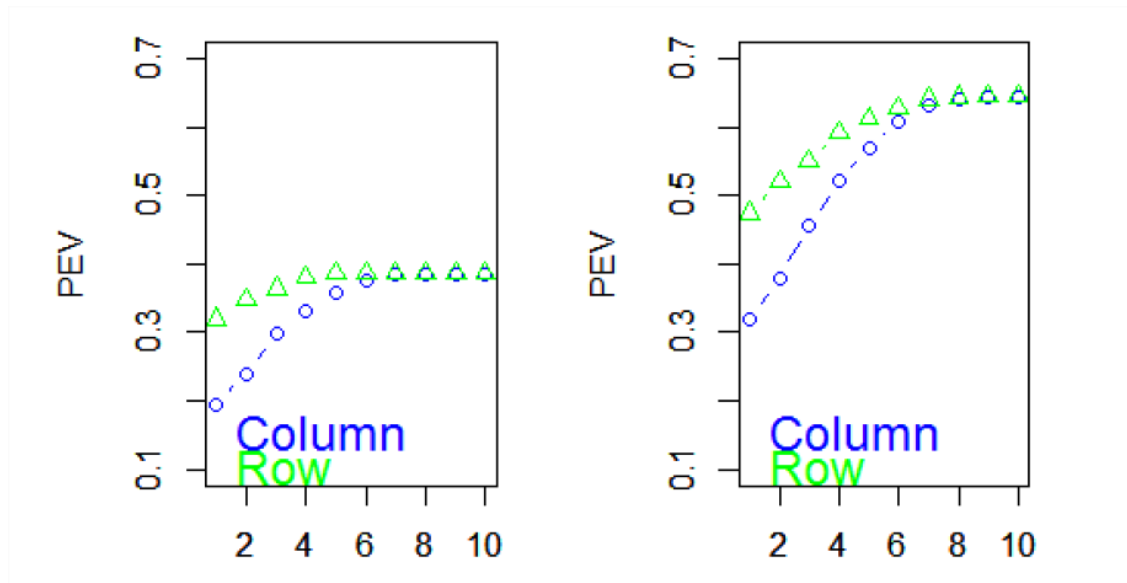
rows

First dimension

second dimension

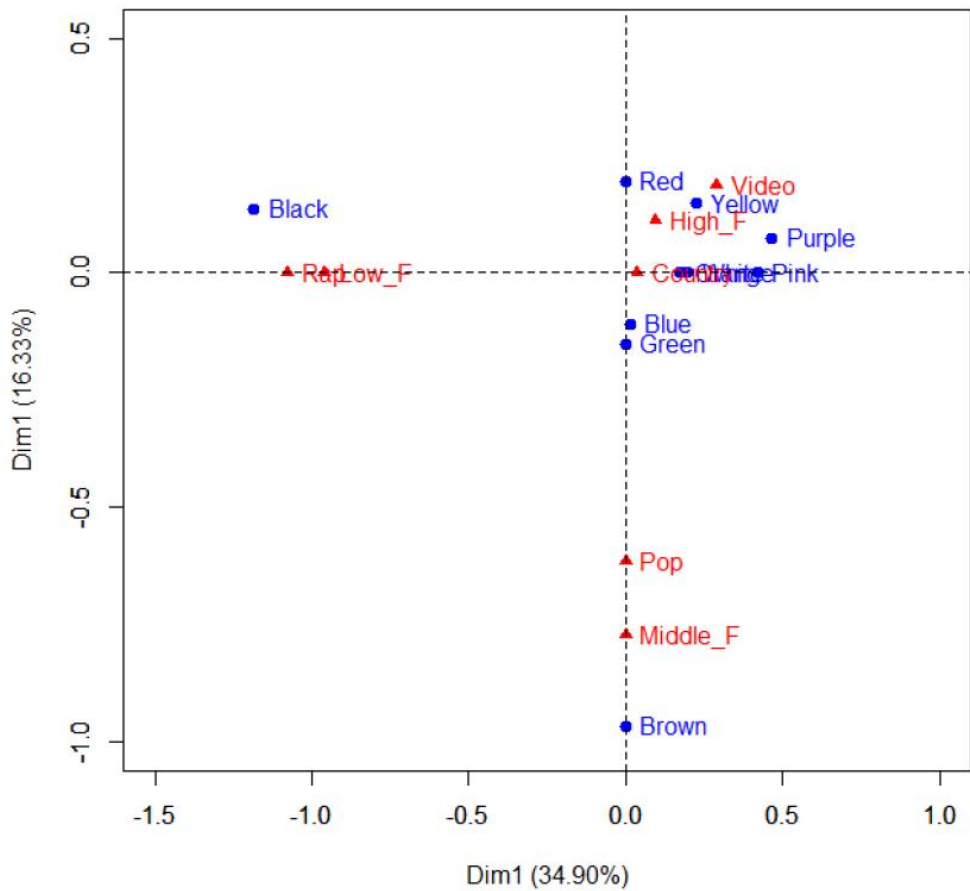
– Choosing the sparsity parameter

- According to the (corrected) cumulative percent of explained variance



1 dimension

2 dimensions



	F1	F2	F.ctr1	F.ctr2
Red	0	-0.193	0	40
Orange	-0.173	0	11	0
Yellow	-0.226	-0.148	27	24
Green	0	0.152	0	24
Blue	-0.014	0.110	0	10
Purple	-0.464	-0.074	58	3
White	-0.196	0	11	0
Black	1.184	-0.135	848	24
Pink	-0.419	0	44	0
Brown	0	0.967	0	875

	G1	G2	G.ctr1	G.ctr2
Video	-0.290	-0.186	39	34
Jazz	0	0	0	0
Country	-0.036	0	1	0
Rap	1.078	0	533	0
Pop	0	0.615	0	370
Opera	0	0	0	0
Low_F	0.961	0	424	0
High_F	-0.096	-0.111	4	12
Middle_F	0	0.772	0	584

5. Conclusions and perspectives

- Work in progress
 - Application to large textual data sets
 - Optimizing the tuning parameters
 - Packaging sparse CA in R (FactoMineR)
- Further developments: a Sparse Non Symmetric Correspondence Analysis?

References

- Adachi, K., Trendafilov, N.T. (2015) : Sparse principal component analysis subject to prespecified cardinality of loadings. *Computational Statistics* **31**, 1–25
- Bernard, A., Guinot, C., Saporta, G. (2012), Sparse principal component analysis for multiblock data and its extension to sparse multiple correspondence analysis, In: Colubi, A., et al. (eds.) *Proceedings of 20th International Conference on Computational Statistics (COMPSTAT 2012)*, 99–106
- Hastie, T., Tibshirani, R., Wainwright, M. (2015) *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman & Hall/CRC
- Jolliffe, I.T., Trendafilov, N.T. , Uddin, M. (2003) A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, **12**, 531–547
- Mori, Y., Kuroda, M., Makino, N. Y. (2016) Sparse Multiple Correspondence Analysis in Mori et al. eds *Nonlinear Principal Component Analysis and Its Applications*, chapter 5, 47-56, Springer
- Rousson, V. , Gasser, T. (2004), Simple component analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **53**, 539-555

- Shen, H. , Huang, J. (2008), Sparse principal component analysis via regularized low rank matrix approximation, *Journal of Multivariate Analysis*, **99**, 1015-1034.
- Shen, N., Li, J. (2015) A Literature Survey on High-Dimensional Sparse Principal Component Analysis *International Journal of Database Theory and Application*, **8**, 6, 57-74
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013) A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics*, **22**, 231-245
- Trendafilov, N.T. (2014) . From simple structure to sparse components: a review. *Computational Statistics*, **29**:431–454.
- Vines, S.K., (2000) Simple principal components, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **49**, 441-451
- Witten, D.M., Tibshirani, R., Hastie, T. (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation. *Biostatistics* **10**:515–534
- Zou, H., Hastie, T. , Tibshirani, R. (2006) Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, **15**, 265-286.