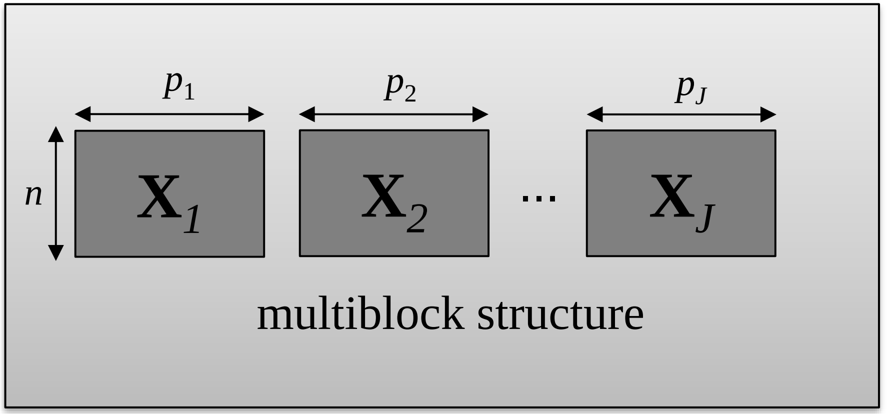
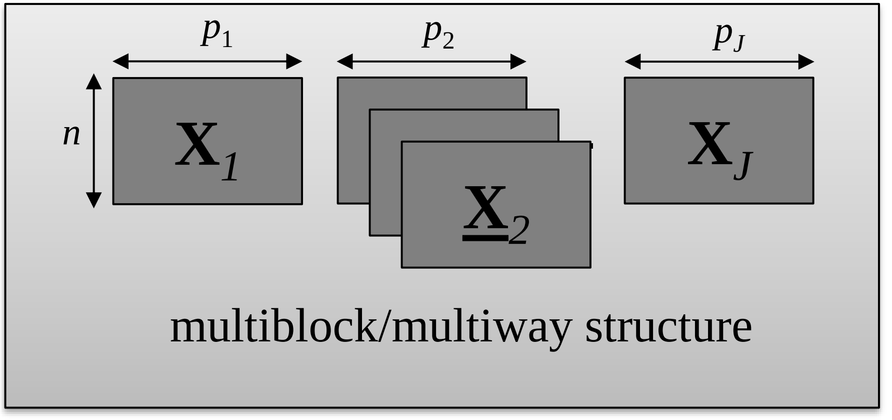
The Regularized Generalized Canonical Correlation Analysis (RGCCA) framework

2025/04/10 @ Biopuces



p_1
p_2
p_J
n
$\mathbf{X}_1$
$\mathbf{X}_2$
...
$\mathbf{X}_J$
multiblock structure



p_1
p_2
p_J
n
$\mathbf{X}_1$
$\underline{\mathbf{X}}_2$
$\mathbf{X}_J$
multiblock/multiway structure

Economic inequality and political instability
Data from Russett (1964)

Economic inequality

Agricultural inequality

GINI : Inequality of land distributions

FARM : % farmers that own half of the land (> 50)

RENT : % farmers that rent all their land

Industrial development

GNPR : Gross national product per capita ($ 1955)

LABO : % of labor force employed in agriculture

Political instability

INST : Instability of executive (45-61)

ECKS : Nb of violent internal war incidents (46-61)

DEAT : Nb of people killed as a result of civic group violence (50-62)

D-STAB : Stable democracy

D-UNST : Unstable democracy

DICT : Dictatorship

Economic inequality and political instability

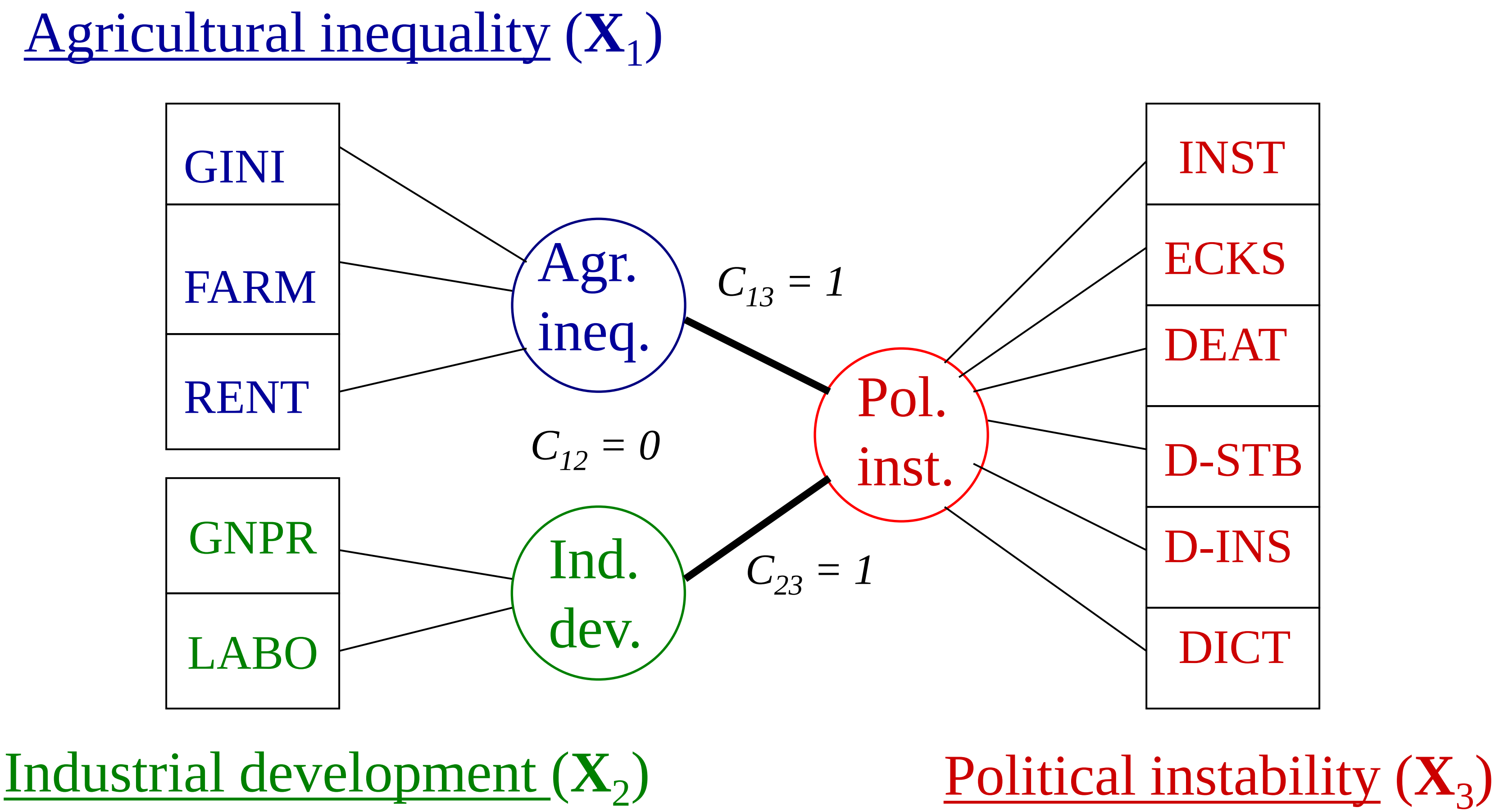
(Data from Russett, 1964)

	X_1			X_2		X_3			
	Gini	**Farm**	**Rent**	**Gnpr**	**Labo**	**Inst**	**Ecks**	**Deat**	**Demo**
Argentine	86.3	98.2	32.9	374	25	13.6	57	217	2
Australie	92.9	99.6	*	1215	14	11.3	0	0	1
Autriche	74.0	97.4	10.7	532	32	12.8	4	0	2
⋮									
France	58.3	86.1	26.0	1046	26	16.3	46	1	2
⋮									
Yougoslavie	43.7	79.8	0.0	297	67	0.0	9	0	3

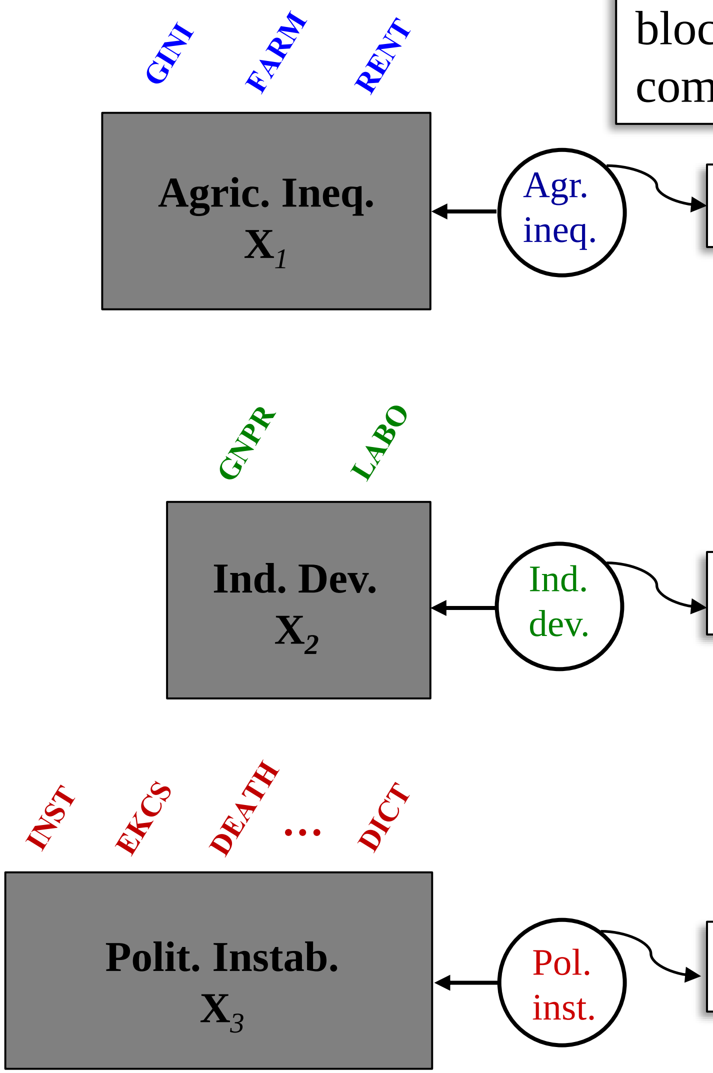
1 = Stable democracy
2 = Unstable democracy
3 = Dictatorship

Three data blocks

Path diagram



The philosophy of multiblock component methods



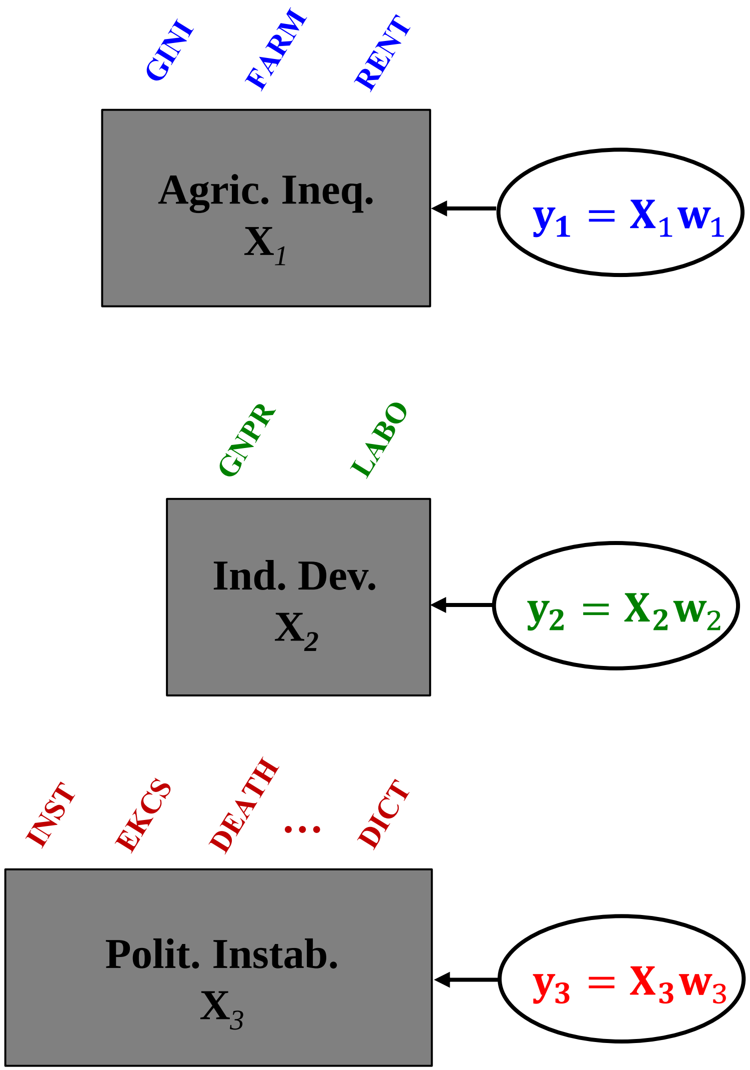
block components | block weight-vectors

$$\mathbf{y}_1 = \mathbf{X}_1\mathbf{w}_1 = w_{11}\mathbf{GINI} + w_{12}\mathbf{FARM} + w_{13}\mathbf{RENT}$$

$$\mathbf{y}_2 = \mathbf{X}_2\mathbf{w}_2 = w_{21}\mathbf{GNPR} + w_{22}\mathbf{LABO}$$

$$\mathbf{y}_3 = \mathbf{X}_3\mathbf{w}_3 = w_{31}\mathbf{INST} + w_{32}\mathbf{ECKS} \ldots + w_{3p_3}\mathbf{DICT}$$

The philosophy of multiblock component methods



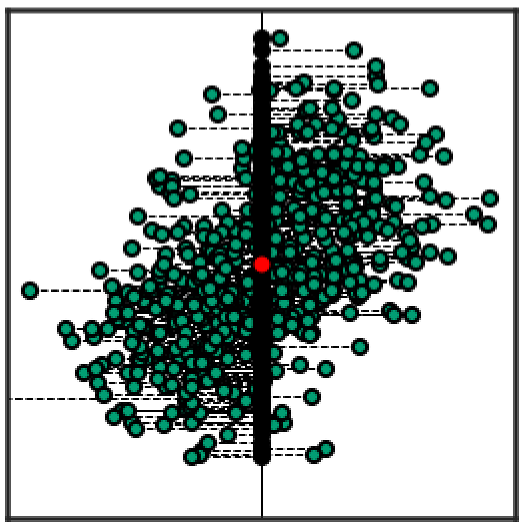
Block components should verify two properties at the same time:

1. Block components well explain their own block.
2. Block components are as correlated as possible for connected blocks.

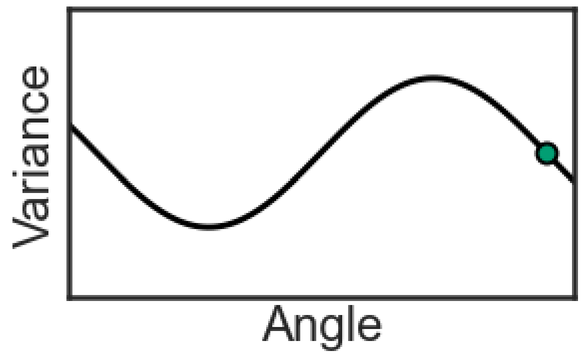
Block components well explain their own block?

Principle of Principal Component Analysis(PCA)

⇒ find direction of maximum variance



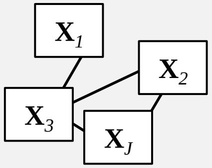
Data, mean and projection



Block components well explain their own block = find direction of high variance!

BLOCKS ARE PARTIALLY CONNECTED

$c_{jk} = 1$ if $\mathbf{X}_j \leftrightarrow \mathbf{X}_k$, 0 otherwise



SUMCOR	$\max\limits_{\mathrm{var}(\mathbf{X}_j\mathbf{w}_j)=1} \sum\limits_{j,k} c_{jk} \mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOR	$\max\limits_{\mathrm{var}(\mathbf{X}_j\mathbf{w}_j)=1} \sum\limits_{j,k} c_{jk} \mathrm{cov}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOR	$\max\limits_{\mathrm{var}(\mathbf{X}_j\mathbf{w}_j)=1} \sum\limits_{j,k} c_{jk} \lvert\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$

RGCCA for multiblock analysis

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_J} h(\mathbf{w}_1,\ldots,\mathbf{w}_J) = \sum_{j,k}^{J} c_{jk}\, g\left(\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\right)$$

$$\mathrm{s.t.}\,(1-\tau_j)\mathrm{var}(\mathbf{X}_j\mathbf{w}_j) + \tau_j\left\|\mathbf{w}_j\right\|_2^2 = 1\,, j = 1,\ldots,J$$

- $c_{jk} = 1 \text{ if } \mathbf{X}_j \leftrightarrow \mathbf{X}_k, 0 \text{ otherwise}$
- $g = \text{ any convex function} - \text{e.g.} \begin{cases} g(x) = x & \text{(Horst sheme)} \\ g(x) = x^2 & \text{(Factorial scheme)} \\ g(x) = |x| & \text{(Centroid scheme)} \end{cases}$
- $0 \le \tau_j \le 1$ continuum between correlation and covariance

Tenenhaus M, Tenenhaus A, Groenen PJF, (2017) Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods, Psychometrika, vol. 82, no. 3, 737–777

Girka F, Camenen E, Caroline P, Gloaguen A, Guillemot V, Le Brusquet L, Tenenhaus A (2023): RGCCA Package. http://cran.project.org/web/packages/RGCCA/index.html

Tenenhaus A, Philippe C, Frouin V (2015) Kernel generalized canonical correlation analysis, Computational Statistics & Data Analysis, vol. 90, pp. 114-131.

Tenenhaus A, Tenenhaus M (2011) Regularized generalized canonical correlation analysis, vol. 76, pp. 257-284, Psychometrika.

RGCCA for multiblock analysis

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_J} h(\mathbf{w}_1, \ldots, \mathbf{w}_J) = \sum_{j,k}^{J} c_{jk} g\left(\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\right)$$

$$\text{s.t. } \mathbf{w}_j^\top\left((1-\tau_j)n^{-1}\mathbf{X}_j^\top\mathbf{X}_j + \tau_j\mathbf{I}_{p_j}\right)\mathbf{w}_j = 1, j = 1, \ldots, J$$

Two key ingredients:

(i) Block relaxation

(ii) Majorization by Minorization (MM)

Tenenhaus M, Tenenhaus A, Groenen PJF, (2017) Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods, Psychometrika, vol. 82, no. 3, 737–777

Girka F, Camenen E, Caroline P, Gloaguen A, Guillemot V, Le Brusquet L, Tenenhaus A (2023): RGCCA Package. http://cran.project.org/web/packages/RGCCA/index.html

Tenenhaus A, Philippe C, Frouin V (2015) Kernel generalized canonical correlation analysis, Computational Statistics & Data Analysis, vol. 90, pp. 114-131.

Tenenhaus A, Tenenhaus M (2011) Regularized generalized canonical correlation analysis, vol. 76, pp. 257-284, Psychometrika.

Block relaxation: from $\mathbf{w}^s$ to $\mathbf{w}^{s+1}$

$\boxed{\mathbf{w}^s = \left(\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_J^s\right)}$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} \; h\left(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_J^s\right) \quad \rightarrow \mathbf{w}_1^{s+1}$$

$\rightarrow \mathbf{w}_2^{s+1}$

$\vdots$

$\rightarrow \mathbf{w}_j^{s+1}$

$\vdots$

$\rightarrow \mathbf{w}_J^{s+1}$

primal algorithm
$n \geq p_j$

dual algorithm
$n < p_j$

Primal/dual update for RGCCA

Primal update

$$\mathbf{w}_j^{s+1} = \frac{\overbrace{\left(\left(1-\tau_j\right)n^{-1}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I}_{p_j}\right)}^{p_j \times p_j}{}^{-1} \mathbf{X}_j^\top \mathbf{z}_j^s}{\left(\mathbf{z}_j^{s\top}\mathbf{X}_j\left(\left(1-\tau_j\right)n^{-1}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I}_{p_j}\right)^{-1}\mathbf{X}_j^\top\mathbf{z}_j^s\right)^{1/2}}$$

Dual update

$$\mathbf{w}_j^{s+1} = \frac{\mathbf{X}_j^\top \overbrace{\left(\left(1-\tau_j\right)n^{-1}\mathbf{X}_j\mathbf{X}_j^\top + \tau_j\mathbf{I}_n\right)}^{n \times n}{}^{-1} \mathbf{z}_j^s}{\left(\mathbf{z}_j^{s\top}\mathbf{X}_j\mathbf{X}_j^\top\left(\left(1-\tau_j\right)n^{-1}\mathbf{X}_j\mathbf{X}_j^\top + \tau_j\mathbf{I}_n\right)^{-1}\mathbf{z}_j^s\right)^{1/2}}$$

Properties of the RGCCA algorithm for multiblock data

► Monotone convergence: $h\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_J^{s+1}\right) \geq h\left(\mathbf{w}_1^{s}, \dots, \mathbf{w}_J^{s}\right)$.

In addition, assuming uniqueness of the solution of the MM step, the following properties hold:

► The sequence $\{\mathbf{w}^s\}$ is asymptotically regular: $\lim_{s \to \infty} \left\| \mathbf{w}^{s+1} - \mathbf{w}^s \right\| = 0$.

► At convergence, a stationary point is obtained.

Tenenhaus M, Tenenhaus A, Groenen P.J.F, (2017) Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods, Psychometrika, doi: 10.1007/s11336-017-9573-x
Girka F, Camenen E, Caroline P, Gloaguen A, Guillemot V, Le Brusquet L, Tenenhaus A (2023): RGCCA Package.
http://cran.project.org/web/packages/RGCCA/index.html

RGCCA as a general framework for multiblock analysis

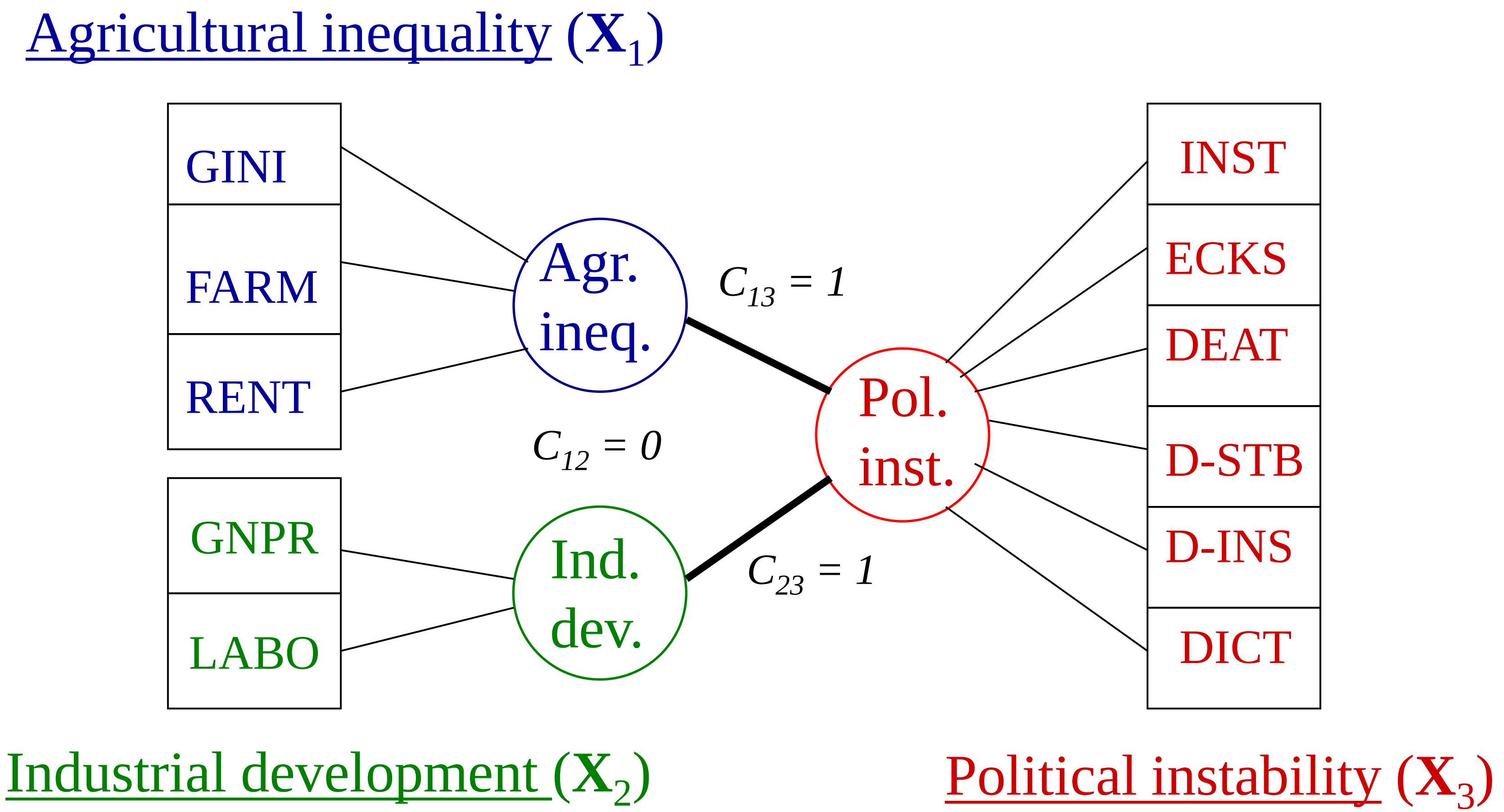
Methods	$g(x)$	τ_j or s_j	**C**	Orthogonality
Canonical correlation analysis `cca`		$\tau_1 = \tau_2 = 0$		
Inter-battery factor analysis or **PLS regression** `ifa/pls`		$\tau_1 = \tau_2 = 1$		
sparse PLS regression `spls`		$\frac{1}{\sqrt{p_1}} < s_1 \leq 1;$ $\frac{1}{\sqrt{p_2}} < s_2 \leq 1$		
Redundancy analysis `ra`	x	$\tau_1 = 1$; $\tau_2 = 0$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	Comp
Regularized redundancy analysis `rgcca`		$0 \leq \tau_1 \leq 1$; $\tau_2 = 0$		
Regularized canonical correlation analysis `rgcca`		$0 \leq \tau_1 \leq 1$; $0 \leq \tau_2 \leq 1$		

Girka F, Camenen E, Caroline P, Gloaguen A, Guillemot V, Le Brusquet L, Tenenhaus A (2023): RGCCA Package. http://cran.project.org/web/packages/RGCCA/index.html

RGCCA as a general framework for multiblock analysis

Methods	$g(x)$	τ_j	$\mathbf{C}$	Orthogonality
SUMCOR `sumcor`	x	$\tau_j = 0$	$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 1 \end{pmatrix}$	Comp
SSQCOR `ssqcor`	x^2	$\tau_j = 0$		Comp
SABSCOR `sabscor`	$\lvert x \rvert$	$\tau_j = 0$		Comp
SUMCOV-1 `sumcov-1`	x	$\tau_j = 1$		Comp
MAXBET `maxbet`	x	$\tau_j = 1$		Weight
SSQCOV-1 `ssqcov-1`	x^2	$\tau_j = 1$		Comp
MAXBET-B `maxbet-b`	x^2	$\tau_j = 1$		Weight
SABSCOV-1 `sabscov-1`	$\lvert x \rvert$	$\tau_j = 1$		Comp
SABSCOV-2 `sabscov-2`	x^2	$\tau_j = 1$		Comp
SUMCOV-2 `sumcov-2`	x	$\tau_j = 1$	$\begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}$	Comp
MAXDIFF `maxdiff`	x	$\tau_j = 1$		Weight
SSQCOV-2 `ssqcov-2`	x^2	$\tau_j = 1$		Comp
MAXDIFF-B `maxdiff-b`	x^2	$\tau_j = 1$		Weight
PLS path modeling - mode B `rgcca`	$\lvert x \rvert$	$\tau_j = 0$	$c_{jk} \neq 0$ for two connected blocks and 0 otherwise	Comp
DIABLO `sgcca`	g	$\frac{1}{\sqrt{p_j}} \leq s_j \leq 1$		Comp
Regularized Generalized Canonical Correlation Analysis `rgcca`	g	$0 \leq \tau_j \leq 1$		Comp/Weight
Sparse Generalized Canonical Correlation Analysis `sgcca`	g	$\frac{1}{\sqrt{p_j}} \leq s_j \leq 1$		Comp/Weight

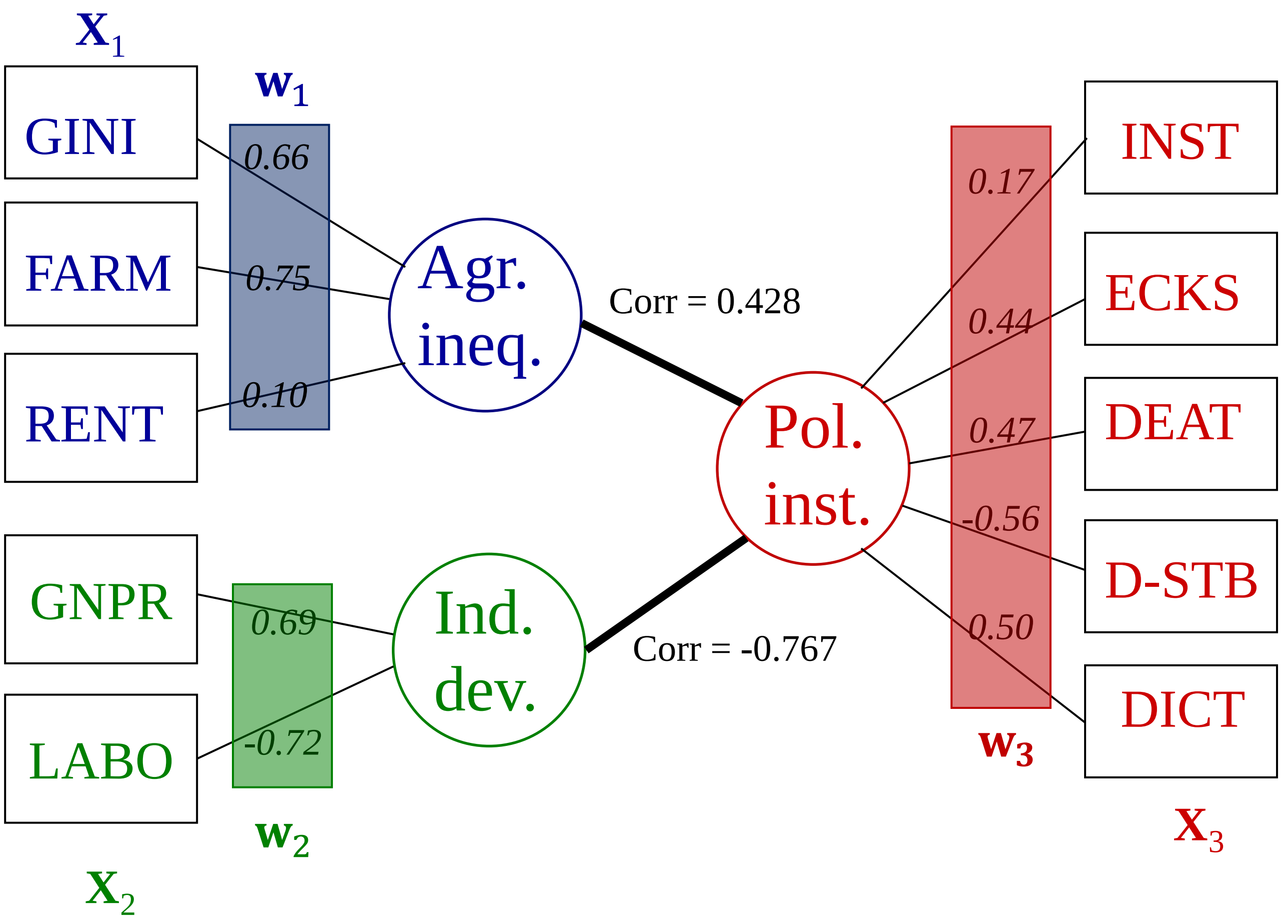
The Russett dataset



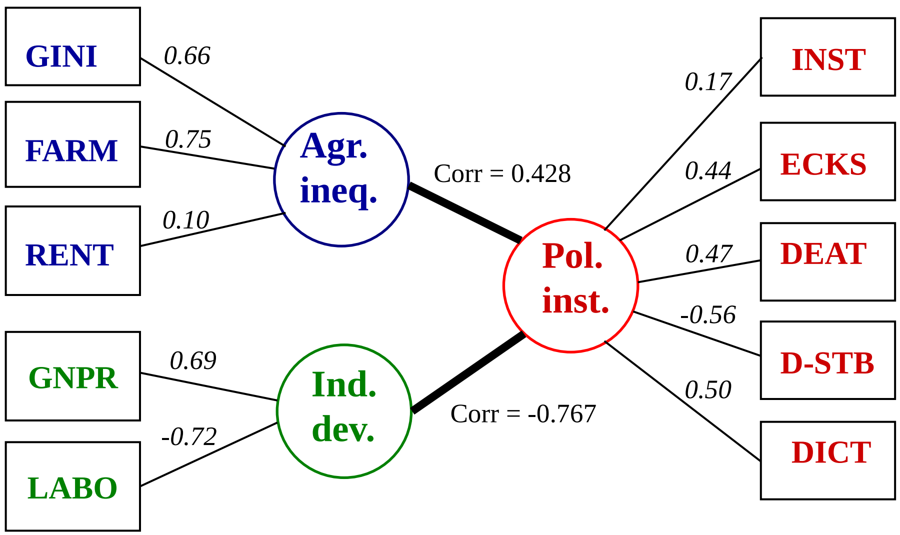
RGCCA on Russett data

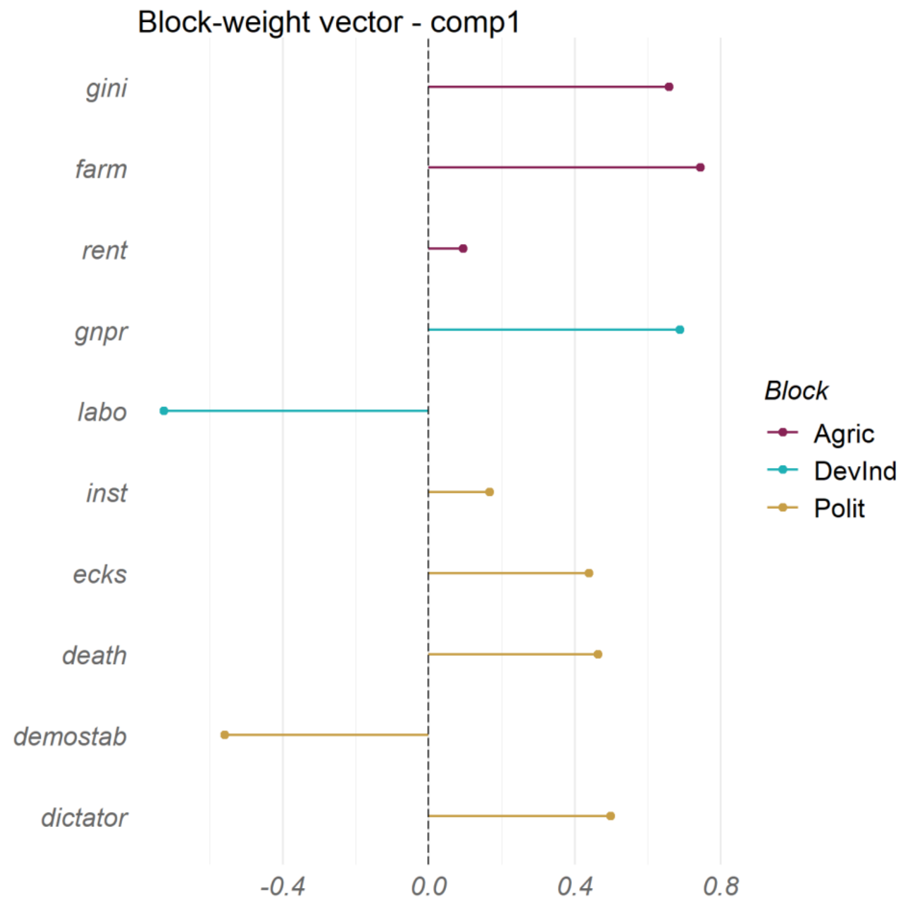
Block-weight vectors with $\tau_j = 1$ and $g(x) = x^2$

$$\max_{\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3} \operatorname{cov}^2(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_3\mathbf{w}_3) + \operatorname{cov}^2(\mathbf{X}_2\mathbf{w}_2, \mathbf{X}_3\mathbf{w}_3) \text{ s.t. } \|\mathbf{w}_j\| = 1\ , j = 1, 2, 3$$

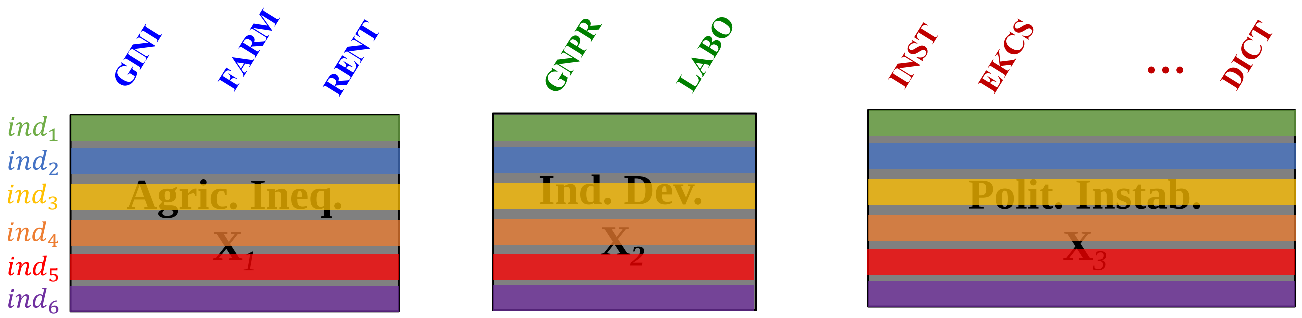


Block-weight vectors

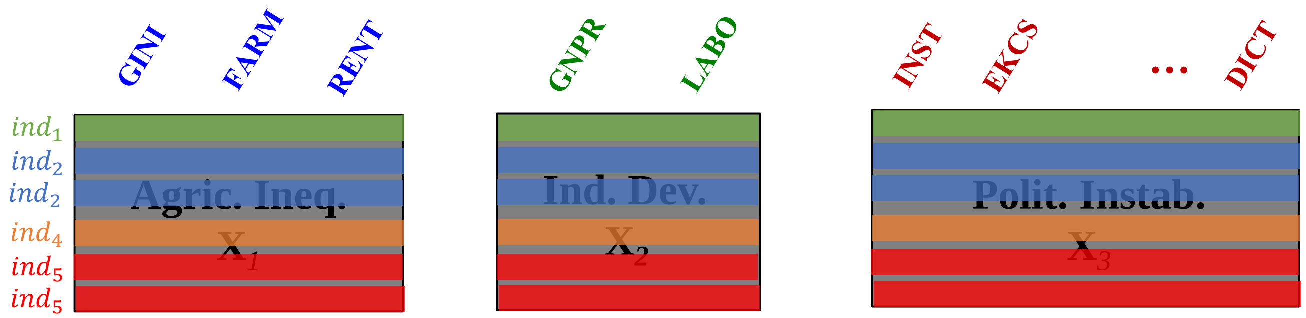




Evaluation of the model by bootstrapping

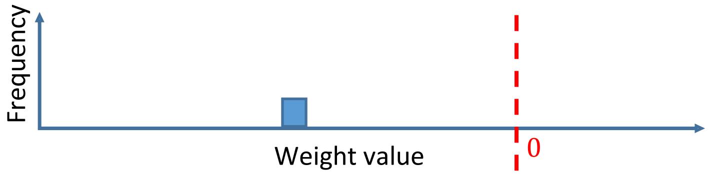


Evaluation of the model by bootstrapping



Bootstrap sample #1

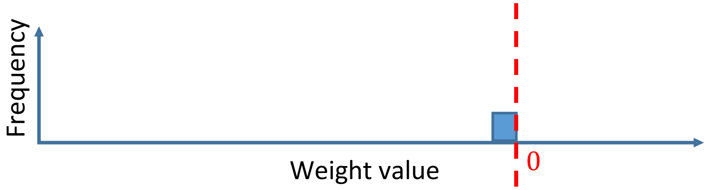
Weight $\mathbf{w}_{11}$ for GINI



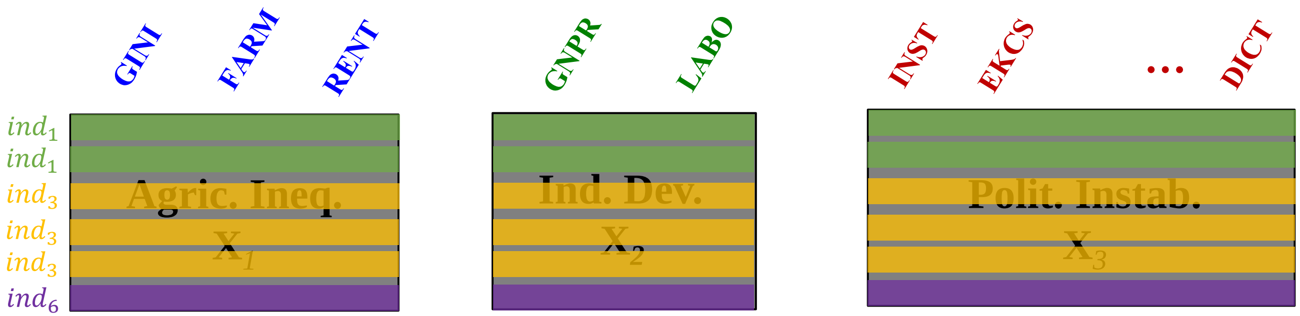
⋮

⋮

Weight $\mathbf{w}_{3p_3}$ for DICT



Evaluation of the model by bootstrapping

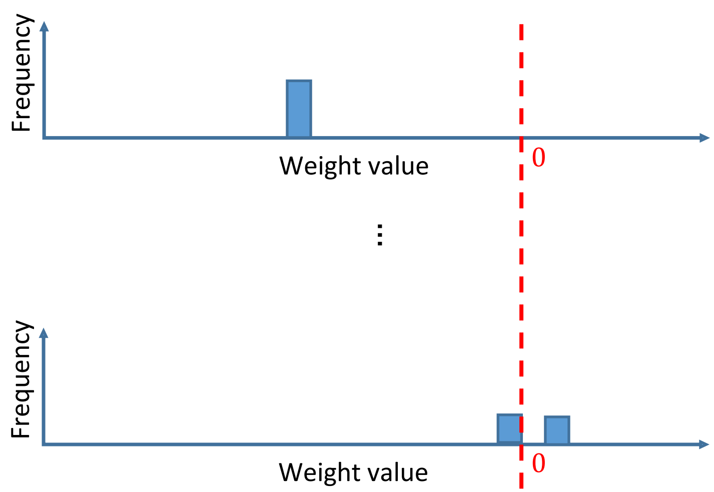


Bootstrap samle #2

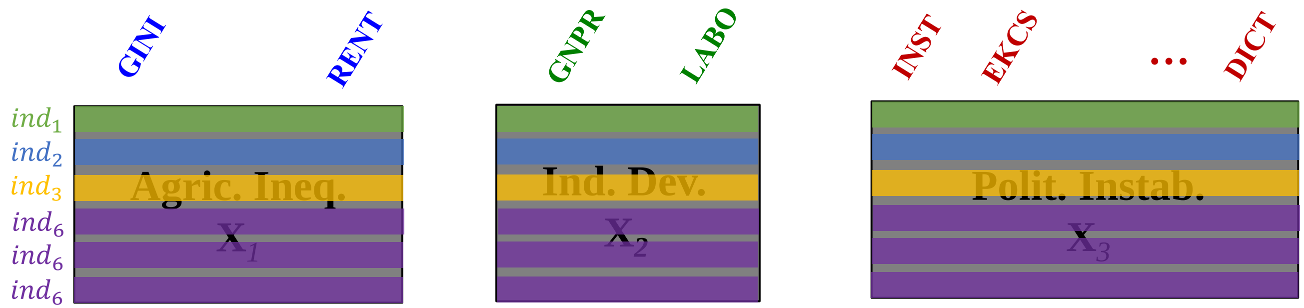
Weight $\mathbf{w}_{11}$ for GINI

⋮

Weight $\mathbf{w}_{3p_3}$ for DICT



Evaluation of the model by bootstrapping

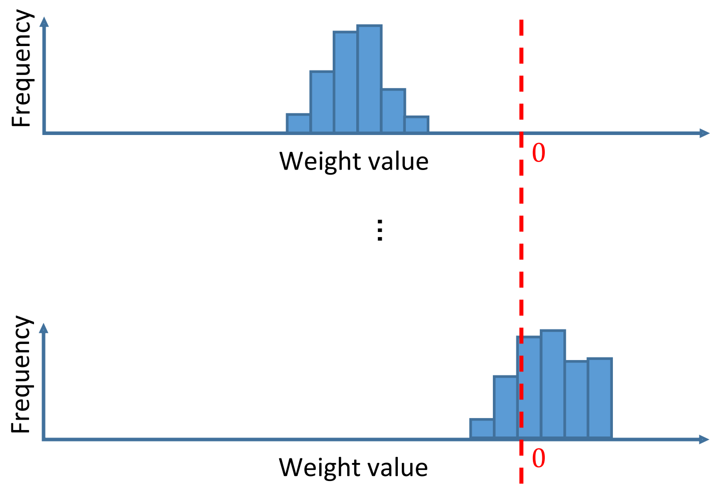


Bootstrap sample #B

Weight $\mathbf{w}_{11}$ for GINI

⋮

Weight $\mathbf{w}_{3p_3}$ for DICT

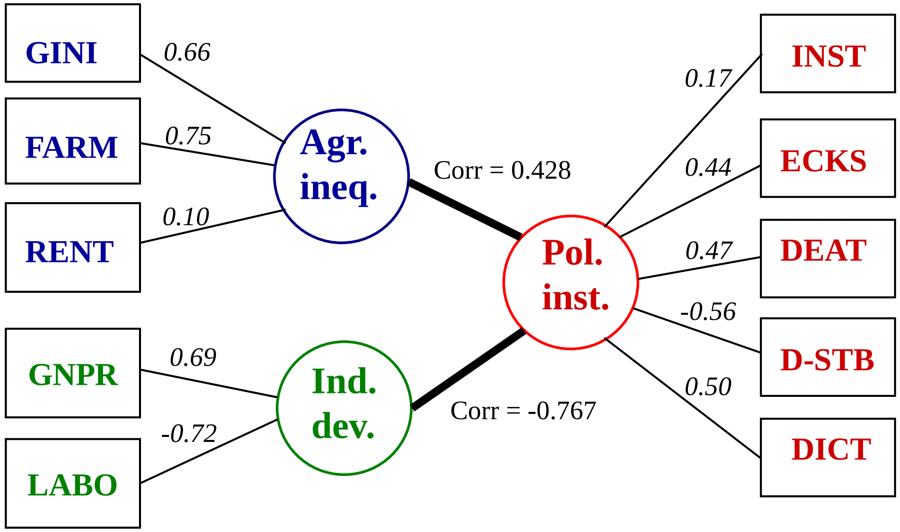


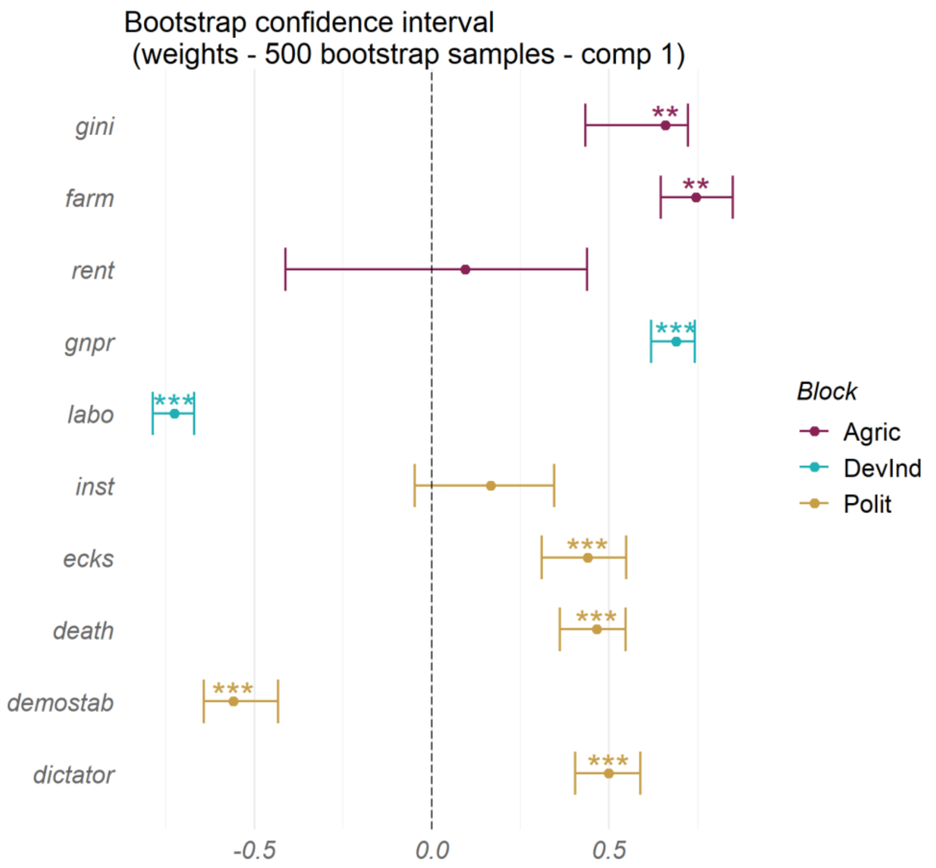
⇒ The weight is likely to be considered as significantly different from 0.

⇒ The weight is unlikely to be considered as significantly different from 0.

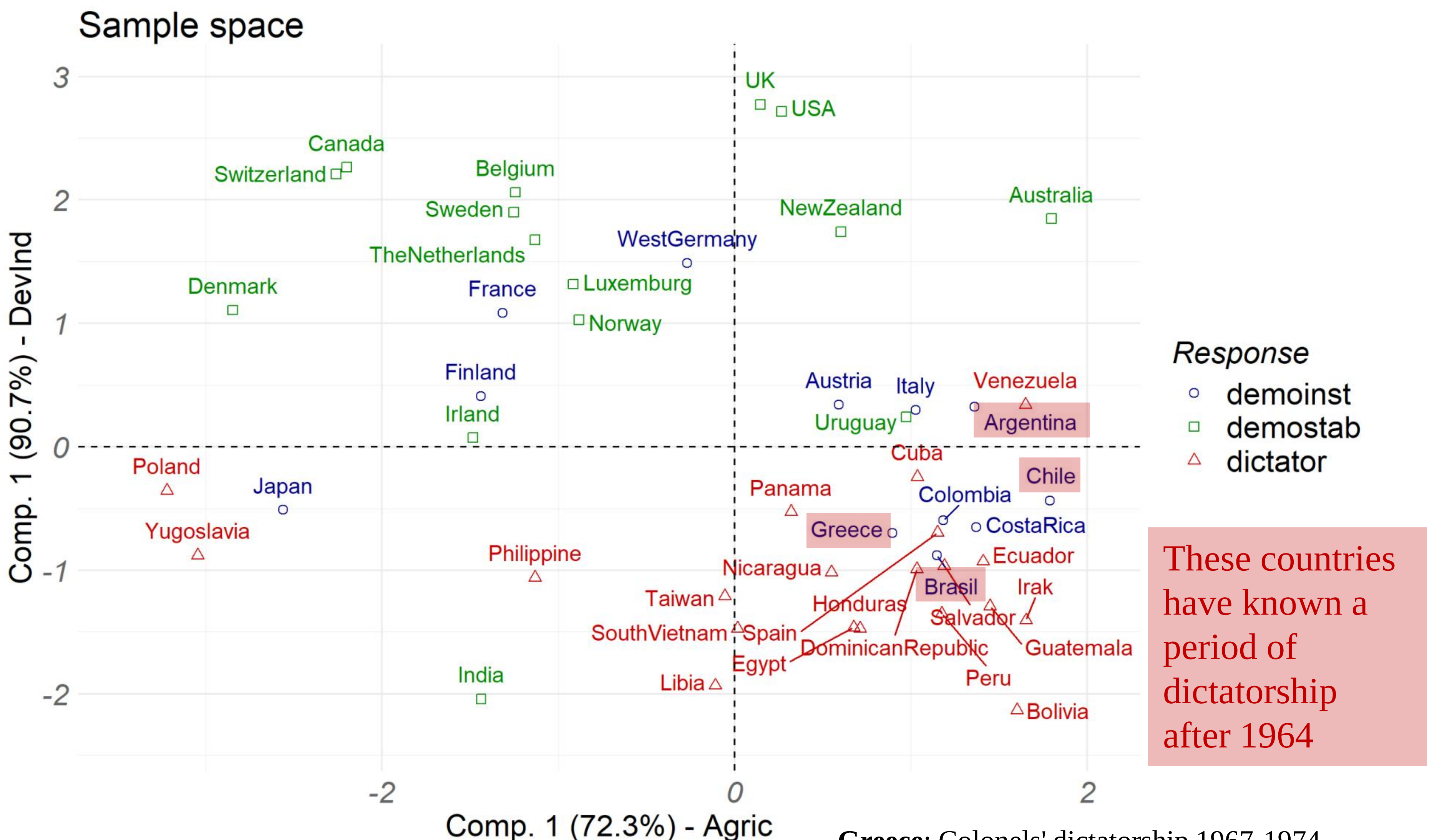
From these distributions, we can derive non-parametric confidence intervals $[q_{0.025}, q_{0.975}]$

Bootstrap confidence intervals





Data vizualization



Greece: Colonels' dictatorship 1967-1974
Chili: Pinochet's military regime 1973-1990
Argentine: Military dictartorship 1976-1983
Brasil: Branco's military dictatorship 1964-1985

Higher-level block components

$$\mathbf{w}_1^{(1)}, \dots, \mathbf{w}_J^{(1)} = \underset{\mathbf{w}_1,\dots,\mathbf{w}_J}{\operatorname{argmax}} \sum_{j,k}^{J} c_{jk}\, g\left(\operatorname{cov}\left(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k\right)\right) \text{ s.t. } \mathbf{w}_j^\top \mathbf{M}_j \mathbf{w}_j = 1\ , j = 1, \dots, J$$

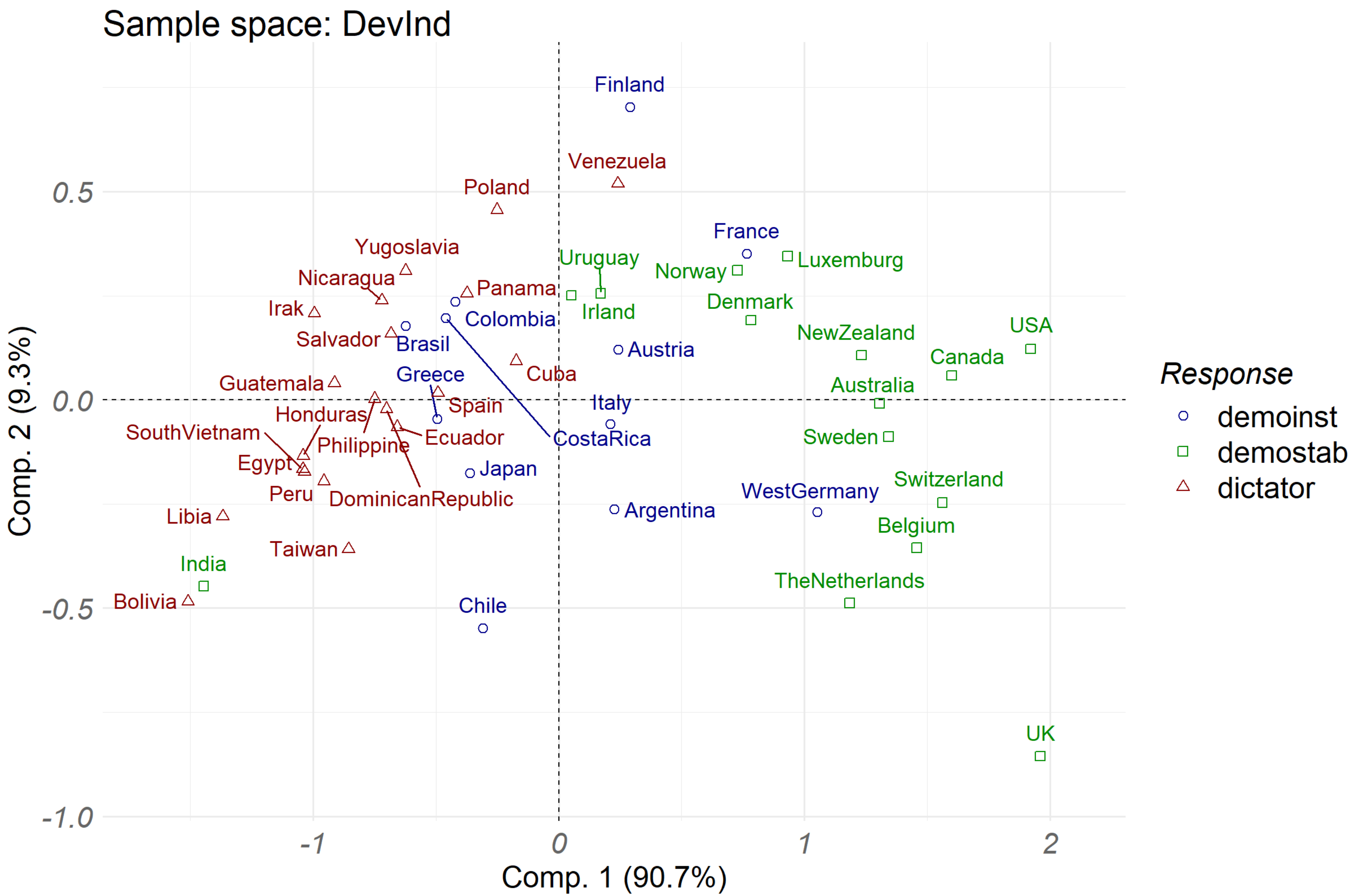
Higher level block components are obtained by considering the following optimization problem:

$$\mathbf{w}_1^{(2)}, \dots, \mathbf{w}_J^{(2)} = \underset{\mathbf{w}_1,\dots,\mathbf{w}_J}{\operatorname{argmax}} \sum_{j,k}^{J} c_{jk}\, g\left(\operatorname{cov}\left(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k\right)\right) \text{ s.t. } \begin{array}{l} \mathbf{w}_j^\top \mathbf{M}_j \mathbf{w}_j = 1\ , j = 1, \dots, J \\ \mathbf{y}_j^{(1)\top} \mathbf{X}_j \mathbf{w}_j = 0\ , j = 1, \dots, J \end{array}$$

$\Rightarrow$ Solved by deflation

Orthogonality constraints

Higher-level block components



Choice of the shrinkage constant : τ_j

(analytical formula)

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_J} h(\mathbf{w}_1,\ldots,\mathbf{w}_J) = \sum_{j,k}^{J} c_{jk}\, g\left(\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\right)$$

$$\text{s.t. } \mathbf{w}_j^\top\left((1-\tau_j)n^{-1}\mathbf{X}_j^\top\mathbf{X}_j + \tau_j\mathbf{I}_{p_j}\right)\mathbf{w}_j = 1,\ j = 1,\ldots,J$$

0

1

Favoring correlation

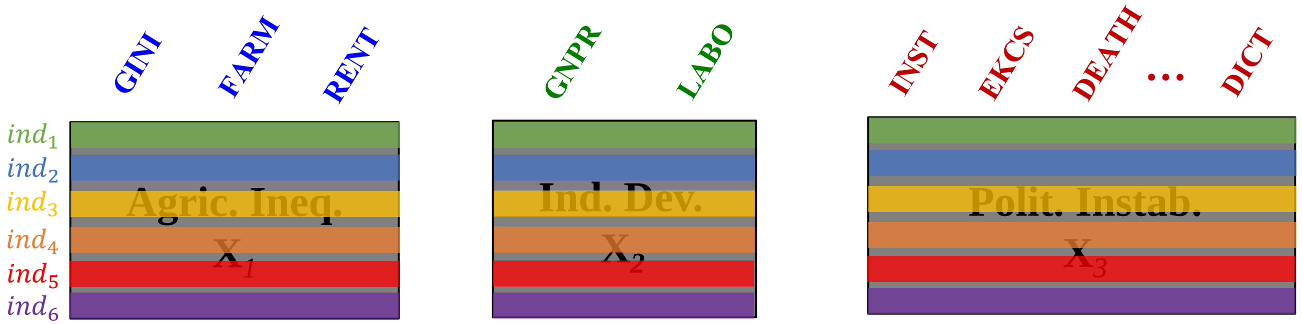
τ_j

Favoring stability

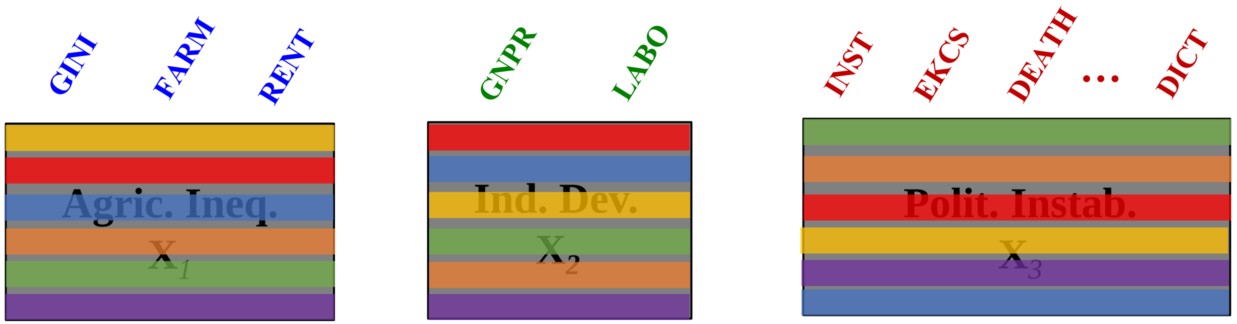
$$\tau_j^* = \underset{\tau_j}{\mathrm{argmin}}\,\mathbb{E}\left[\left\|\hat{\Sigma}_j(\tau_j) - \Sigma_j\right\|_F^2\right]$$

Schäfer & Strimmer formula can be used for an optimal determination of the shrinkage constants

Choice of the shrinkage constant : τ_j
(permutation procedure)

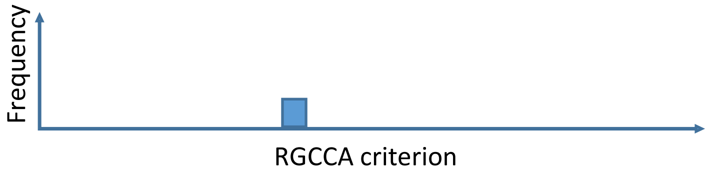


Choice of the shrinkage constant : τ_j
(permutation procedure)



Permutation #1

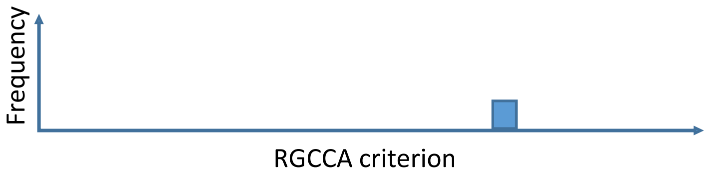
Parameter set #1



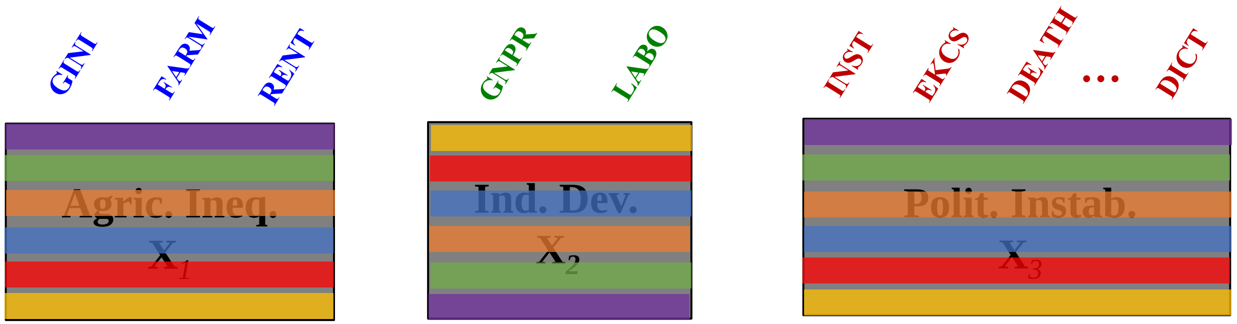
⋮

⋮

Parameter set #K

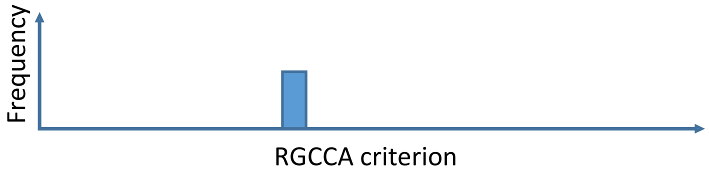


Choice of the shrinkage constant : τ_j
(permutation procedure)



Permutation #2

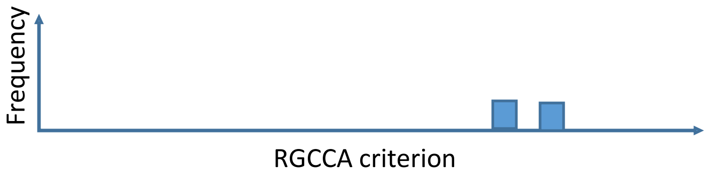
Parameter set #1



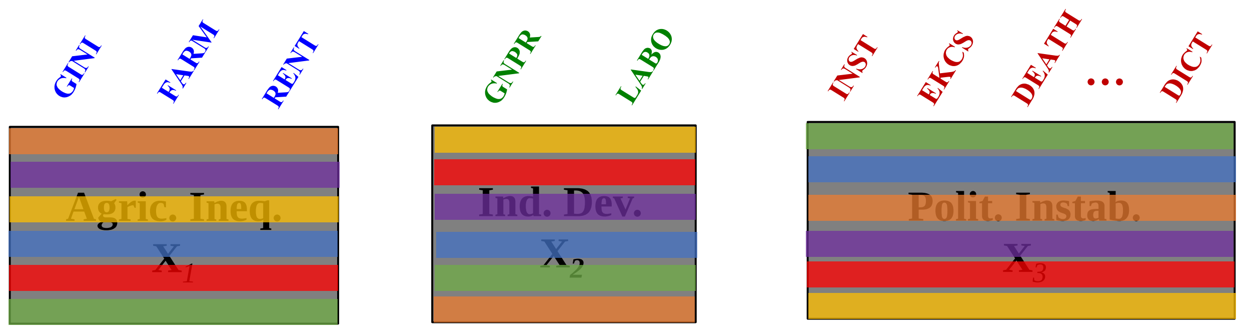
⋮

⋮

Parameter set #K

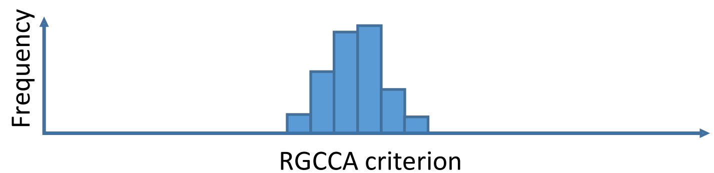


Choice of the shrinkage constant : τ_j
(permutation procedure)



Permutation #N

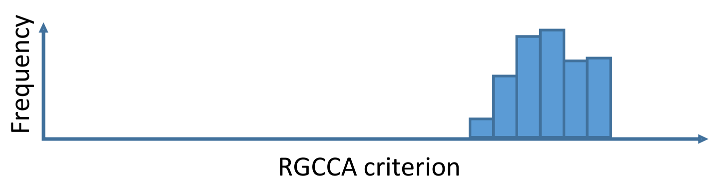
Parameter set #1



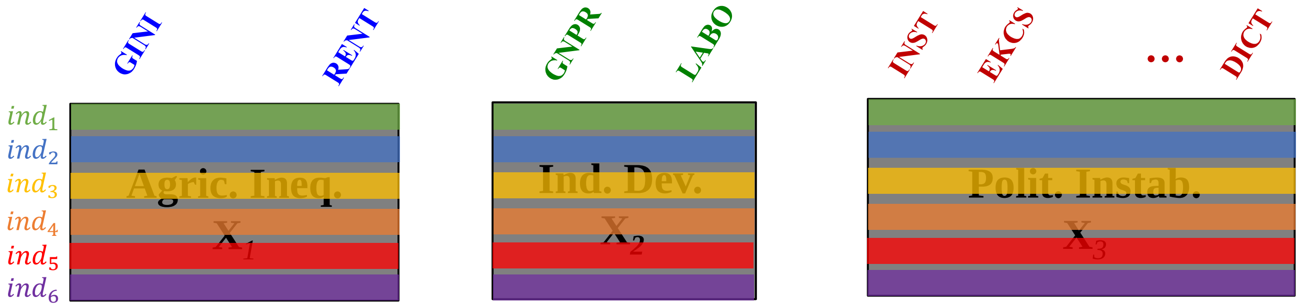
⋮

⋮

Parameter set #K

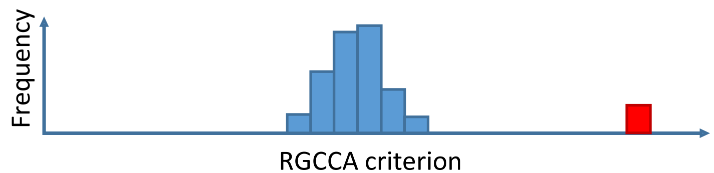


Choice of the shrinkage constant : τ_j
(permutation procedure)



No permutation

Parameter set #1

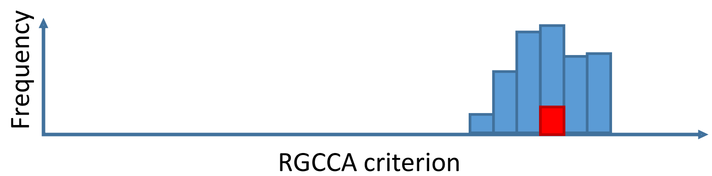


$$\Rightarrow z_1 = \frac{\text{crit} - \mu_{\text{crit\_perm}}}{\sigma_{\text{crit\_perm}}}$$

⋮

⋮

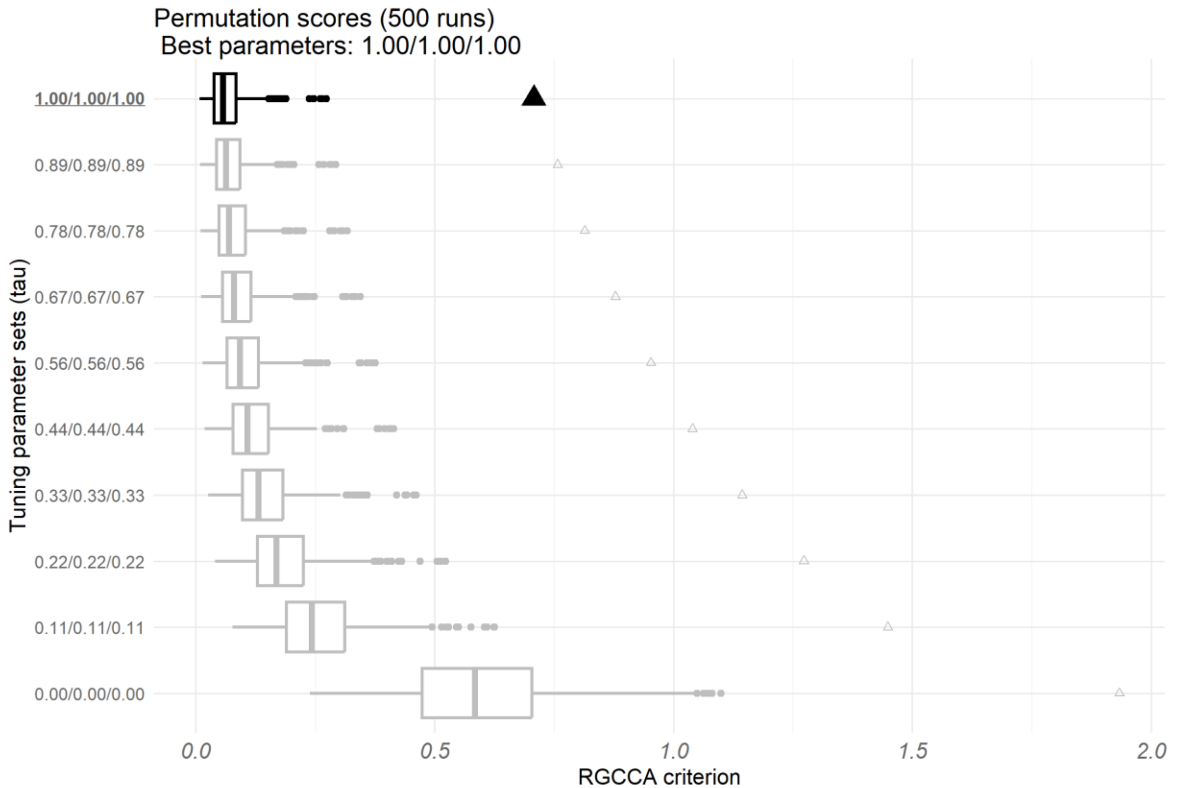
Parameter set #K



$$\Rightarrow z_K = \frac{\text{crit} - \mu_{\text{crit\_perm}}}{\sigma_{\text{crit\_perm}}}$$

The best set of parameters is associated with the highest z-value

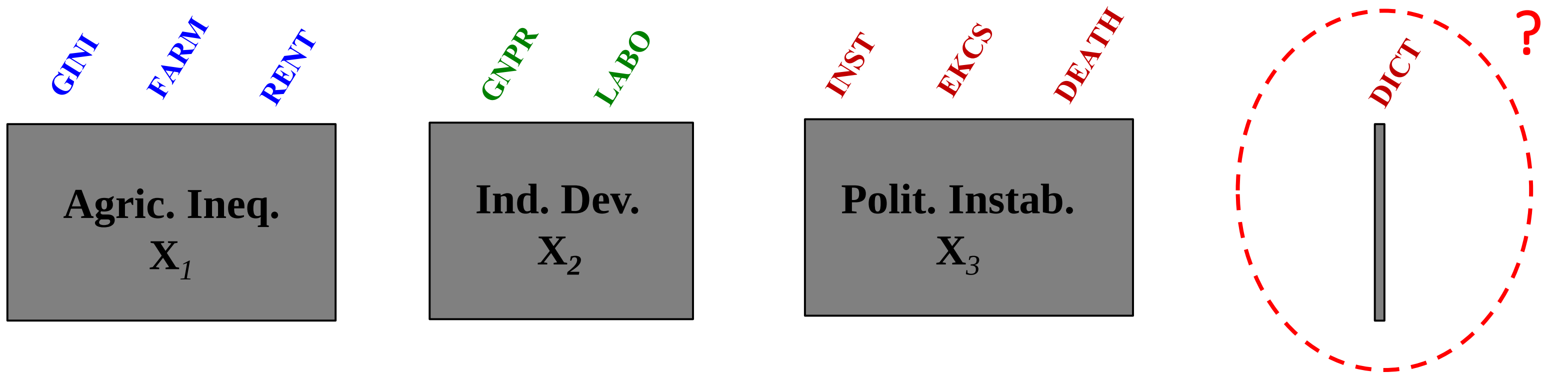
Determination of τ_j by permutation



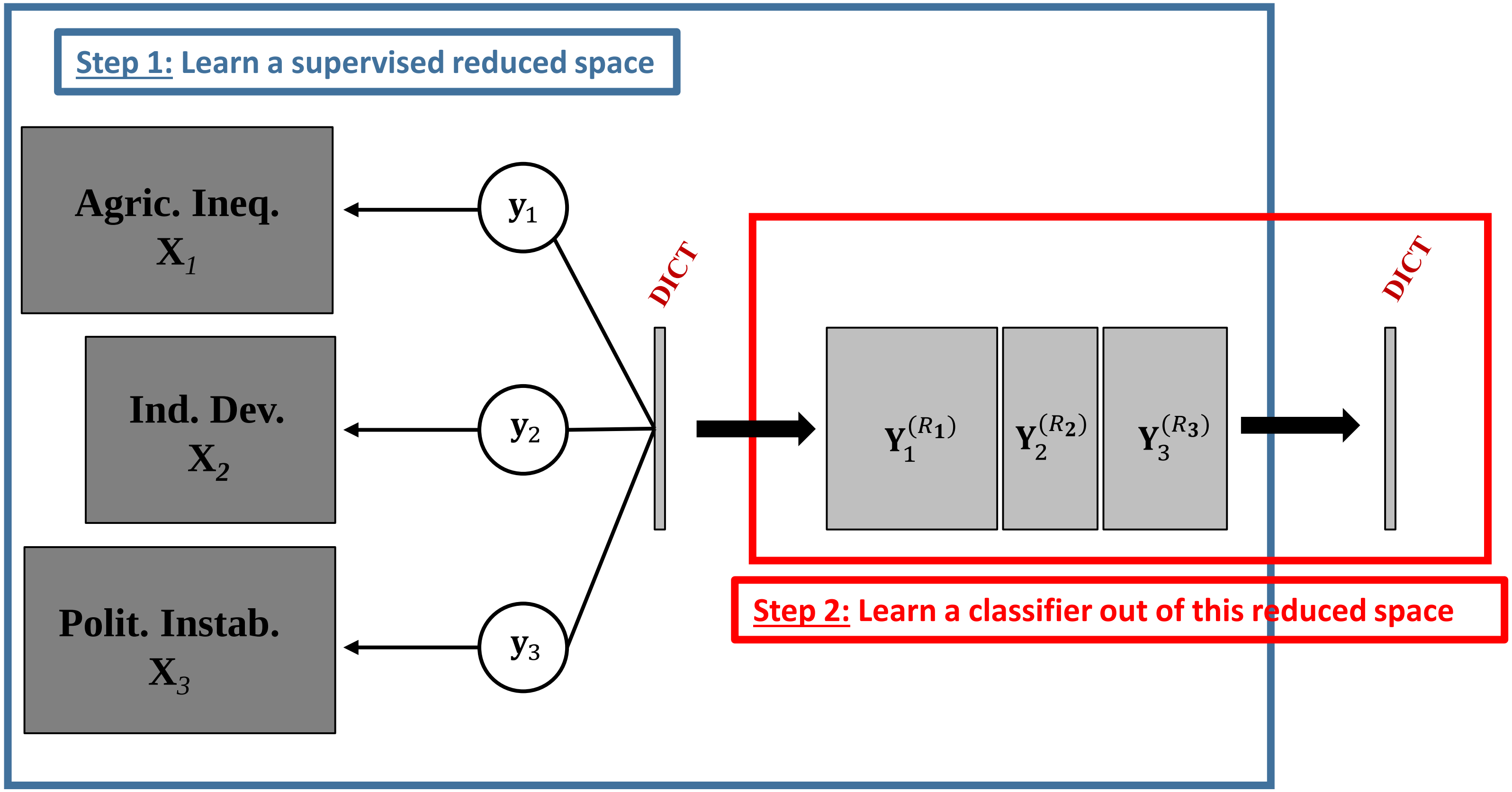

```
Tuning parameters (tau) used:
   Agric DevInd Polit
1  1.000  1.000 1.000
2  0.889  0.889 0.889
3  0.778  0.778 0.778
4  0.667  0.667 0.667
5  0.556  0.556 0.556
6  0.444  0.444 0.444
7  0.333  0.333 0.333
8  0.222  0.222 0.222
9  0.111  0.111 0.111
10 0.000  0.000 0.000

   Tuning parameters Criterion Permuted criterion     sd zstat p-value
1     1.00/1.00/1.00     0.708             0.0671 0.0402 15.93       0
2     0.89/0.89/0.89     0.758             0.0738 0.0433 15.81       0
3     0.78/0.78/0.78     0.814             0.0819 0.0467 15.66       0
4     0.67/0.67/0.67     0.878             0.0919 0.0508 15.49       0
5     0.56/0.56/0.56     0.953             0.1046 0.0555 15.27       0
6     0.44/0.44/0.44     1.040             0.1216 0.0613 14.98       0
7     0.33/0.33/0.33     1.144             0.1456 0.0685 14.57       0
8     0.22/0.22/0.22     1.273             0.1837 0.0783 13.92       0
9     0.11/0.11/0.11     1.449             0.2586 0.0942 12.64       0
10    0.00/0.00/0.00     1.934             0.5953 0.1660  8.06       0
The best combination is: 1.00/1.00/1.00 for a z score of 15.9 and a p-value of 0
```

Supervised RGCCA



Supervised RGCCA



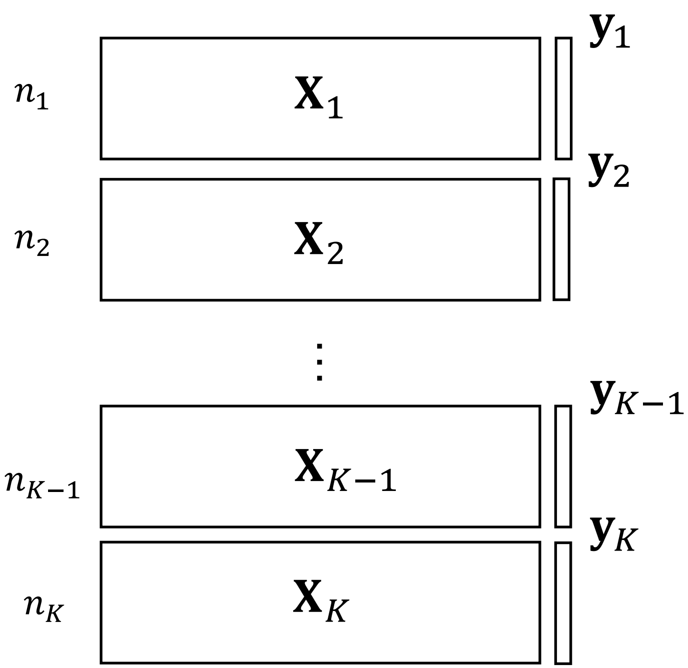
Standard Cross-Validation (K-Fold, LOO) can be performed to tune hyperparameters.

Choice of the shrinkage parameter: τ_j
(Cross-validation)

Leitmotiv: a good model should predict efficiently samples not used for its construction.

K-fold cross-validation

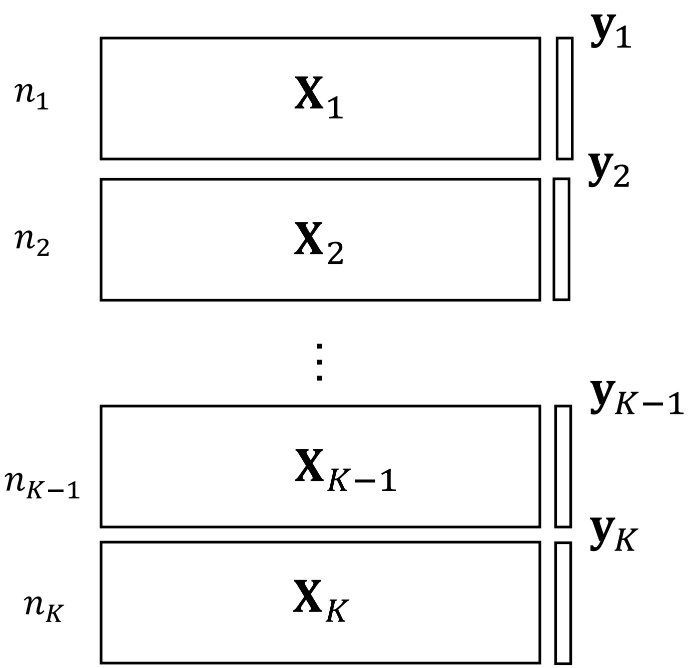
K-fold cross validation. Split the data into K segments of equal size.



$$\hat{f}_{-K}^{\tau}(\mathbf{X}_K) = \hat{\mathbf{y}}_K \quad \Rightarrow \mathrm{errCV}_{K,\tau}$$

K-fold cross-validation

K-fold cross validation. Split the data into K segments of equal size.

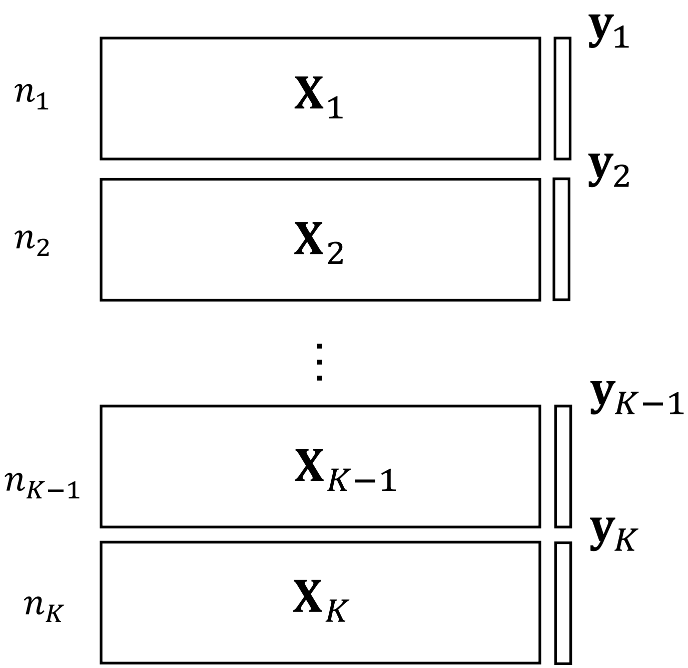


$$\hat{f}^{\tau}_{-K}(\mathbf{X}_K) = \hat{\mathbf{y}}_K \quad \Rightarrow \mathrm{errCV}_{K,\tau}$$

$$\hat{f}^{\tau}_{-(K-1)}(\mathbf{X}_{K-1}) = \hat{\mathbf{y}}_{K-1} \quad \Rightarrow \mathrm{errCV}_{K-1,\tau}$$

K-fold cross-validation

K-fold cross validation. Split the data into K segments of equal size.



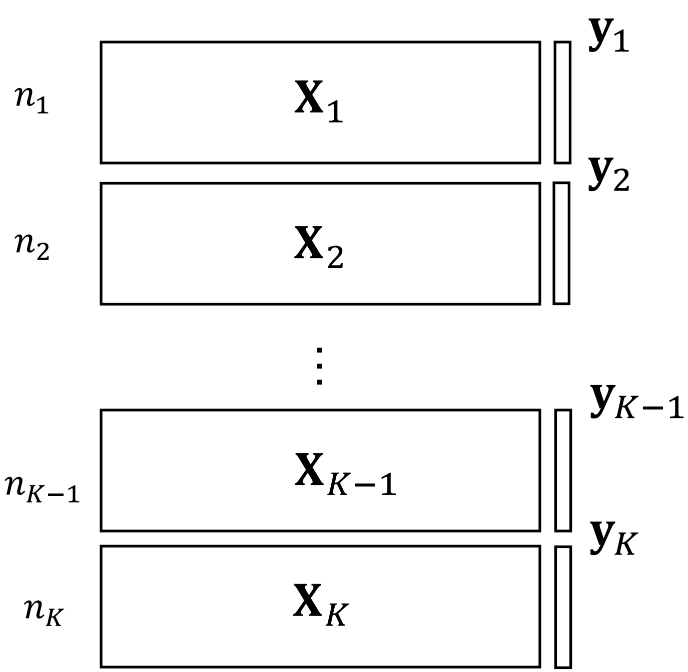
$$\hat{f}^{\tau}_{-K}(\mathbf{X}_K) = \hat{\mathbf{y}}_K \quad \Rightarrow \mathrm{errCV}_{K,\tau}$$

$$\hat{f}^{\tau}_{-(K-1)}(\mathbf{X}_{K-1}) = \hat{\mathbf{y}}_{K-1} \quad \Rightarrow \mathrm{errCV}_{K-1,\tau}$$

$$\hat{f}^{\tau}_{-2}(\mathbf{X}_2) = \hat{\mathbf{y}}_2 \quad \Rightarrow \mathrm{errCV}_{2,\tau}$$

K-fold cross-validation

K-fold cross validation. Split the data into K segments of equal size.



$$\hat{f}^{\tau}_{-K}(\mathbf{X}_K) = \hat{\mathbf{y}}_K \quad \Rightarrow \mathrm{errCV}_{K,\tau}$$

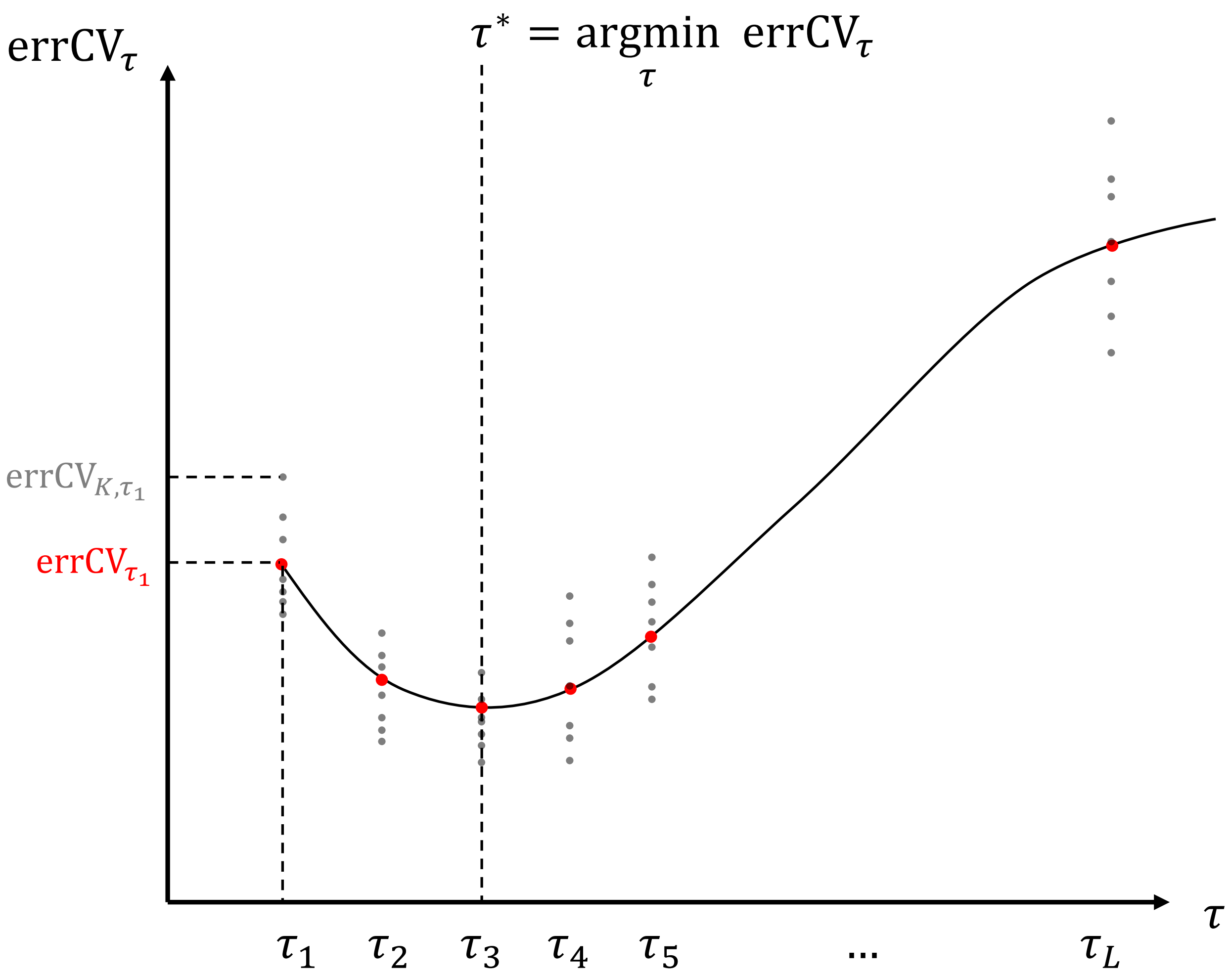
$$\hat{f}^{\tau}_{-(K-1)}(\mathbf{X}_{K-1}) = \hat{\mathbf{y}}_{K-1} \quad \Rightarrow \mathrm{errCV}_{K-1,\tau}$$

$$\hat{f}^{\tau}_{-2}(\mathbf{X}_2) = \hat{\mathbf{y}}_2 \quad \Rightarrow \mathrm{errCV}_{2,\tau}$$

$$\hat{f}^{\tau}_{-1}(\mathbf{X}_1) = \hat{\mathbf{y}}_1 \quad \Rightarrow \mathrm{errCV}_{1,\tau}$$

$$\mathrm{errCV}_{\tau} = K^{-1} \sum_{k=1}^{K} \mathrm{errCV}_{k,\tau}$$

Model selection by cross validation: τ^*



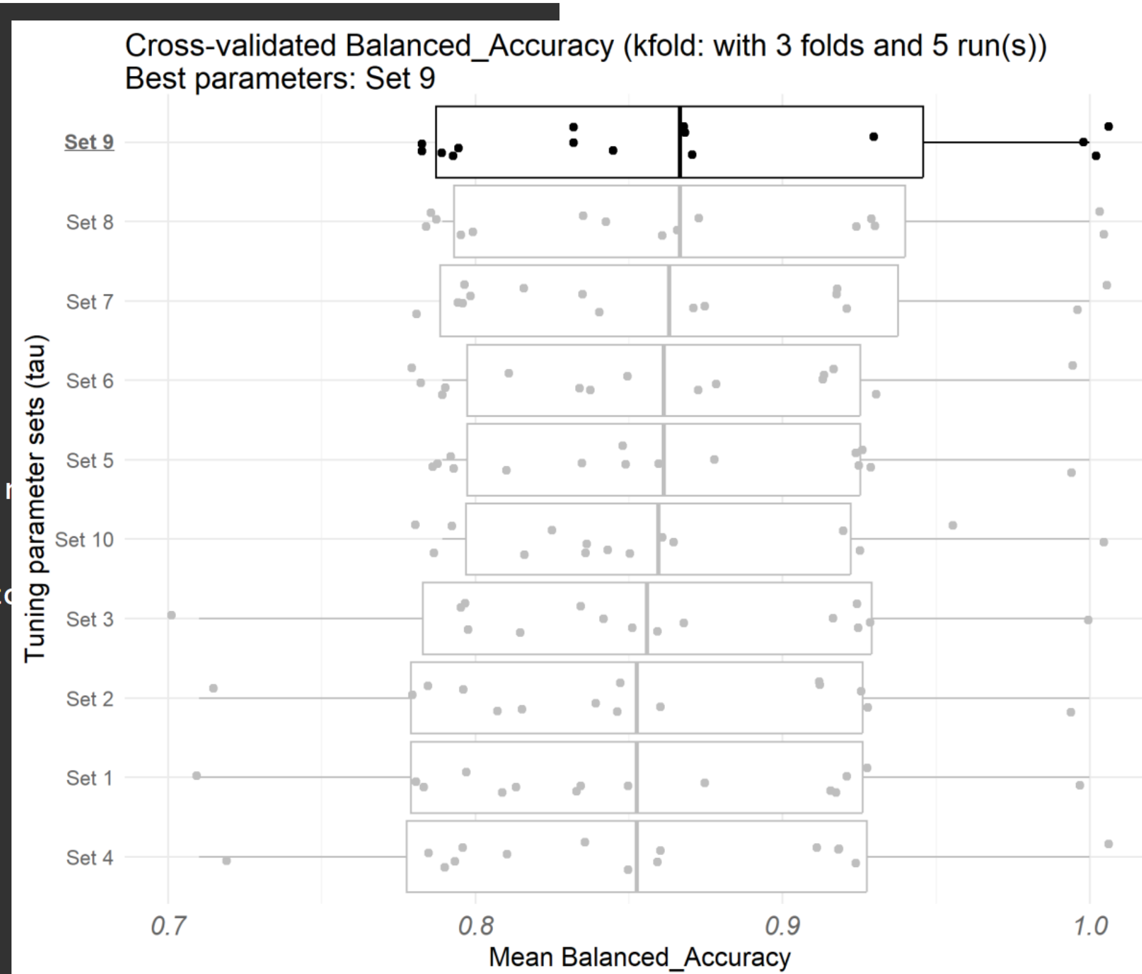
Model selection by cross-validation

```
Tuning parameters (tau) used:
   Agric   Ind Polit Regime
1  1.000 1.000 1.000      0
2  0.889 0.889 0.889      0
3  0.778 0.778 0.778      0
4  0.667 0.667 0.667      0
5  0.556 0.556 0.556      0
6  0.444 0.444 0.444      0
7  0.333 0.333 0.333      0
8  0.222 0.222 0.222      0
9  0.111 0.111 0.111      0
10 0.000 0.000 0.000      0

Validation: kfold with 3 folds and 5 r
Prediction model: lda

   Tuning parameters Mean Balanced_Acc
1              Set 1
2              Set 2
3              Set 3
4              Set 4
5              Set 5
6              Set 6
7              Set 7
8              Set 8
9              Set 9
10            Set 10

The best combination is: Set 9 for a mean Balanced_Accuracy of 0.867
```



Consensus space with RGCCA

The goal is to find jointly a global component $\mathbf{y}$ and block components $\mathbf{y}_1 = \mathbf{X}_1\mathbf{w}_1, \ldots, \mathbf{y}_J = \mathbf{X}_J\mathbf{w}_J$.

- The global block component is obtained by considering the following optimization problem

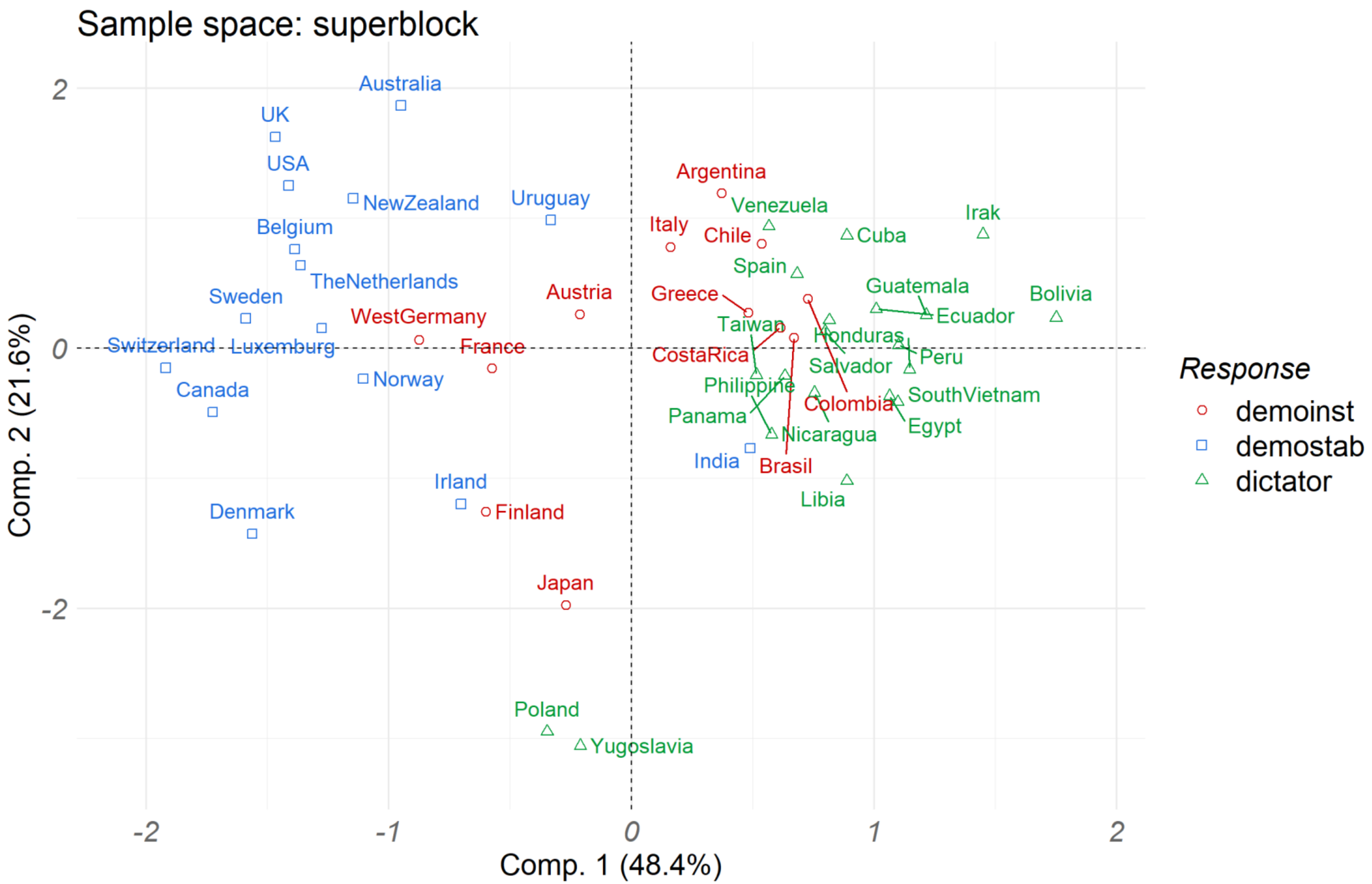
$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_{J+1}} \sum_{j=1}^{J} g\left(\operatorname{cov}\left(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_{J+1}\mathbf{w}_{J+1}\right)\right) \text{ s.t. } \mathbf{w}_j^\top \mathbf{M}_j \mathbf{w}_j = 1, \forall j$$

- <u>**Important result**</u>**:** The optimal global component is obtained as linear combination of the variable of the so-called superblock defined as: $\mathbf{X}_{J+1} = [\mathbf{X}_1 | \ldots | \mathbf{X}_J]$

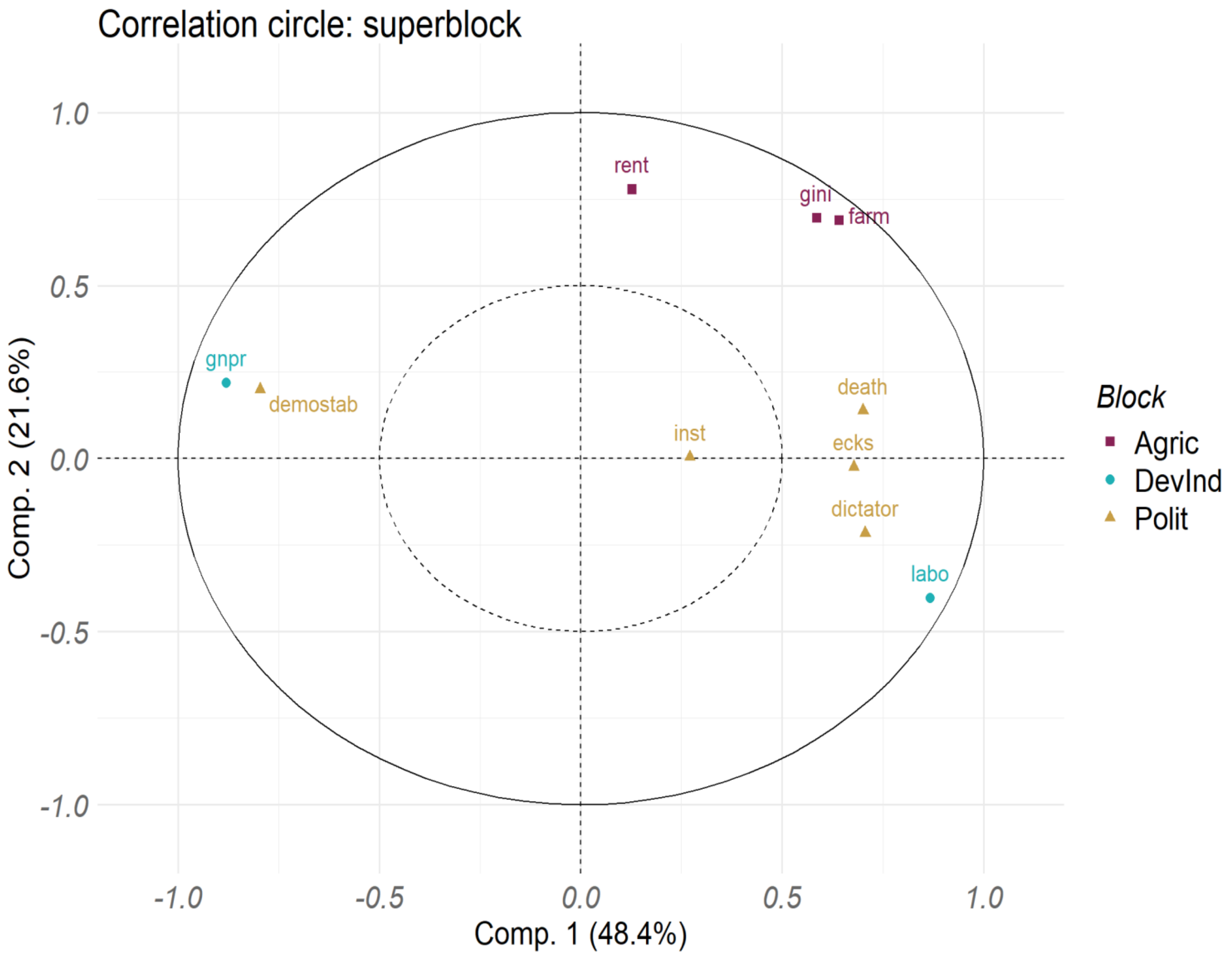
RGCCA as a general framework for multiblock analysis

Methods	$g(x)$	τ_j	$\mathbf{C}$	Orthogonality
Generalized CCA `gcca/maxvar/maxvar-b`	x^2	$\tau_j = 0, j = 1, \ldots, J+1$		Comp
(mixed) Generalized CCA `rgcca`	x^2	$\tau_j = 0, j = 1, \ldots, J_1$; $\tau_j = 1, j = J_1 + 1, \ldots, J$		Comp
Multiple co-inertia analysis `mcoa/mcia`	x^2	$\tau_j = 1, j = 1, \ldots, J$; $\tau_{J+1} = 0$		Weight
Multiple factor analysis `mfa`	x^2	$\tau_j = 1, j = 1, \ldots, J+1$	$\begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}$	Comp
Consensus PCA(1) `cpca-1`	x	$\tau_j = 1, j = 1, \ldots, J$; $\tau_{J+1} = 0$		Comp
Consensus PCA(2) `cpca-2/maxvar-a`	x^2	$\tau_j = 1, j = 1, \ldots, J$; $\tau_{J+1} = 0$		Comp
Hierarchical PCA `hpca/cpca-4`	x^4	$\tau_j = 1, j = 1, \ldots, J$; $\tau_{J+1} = 0$		Comp

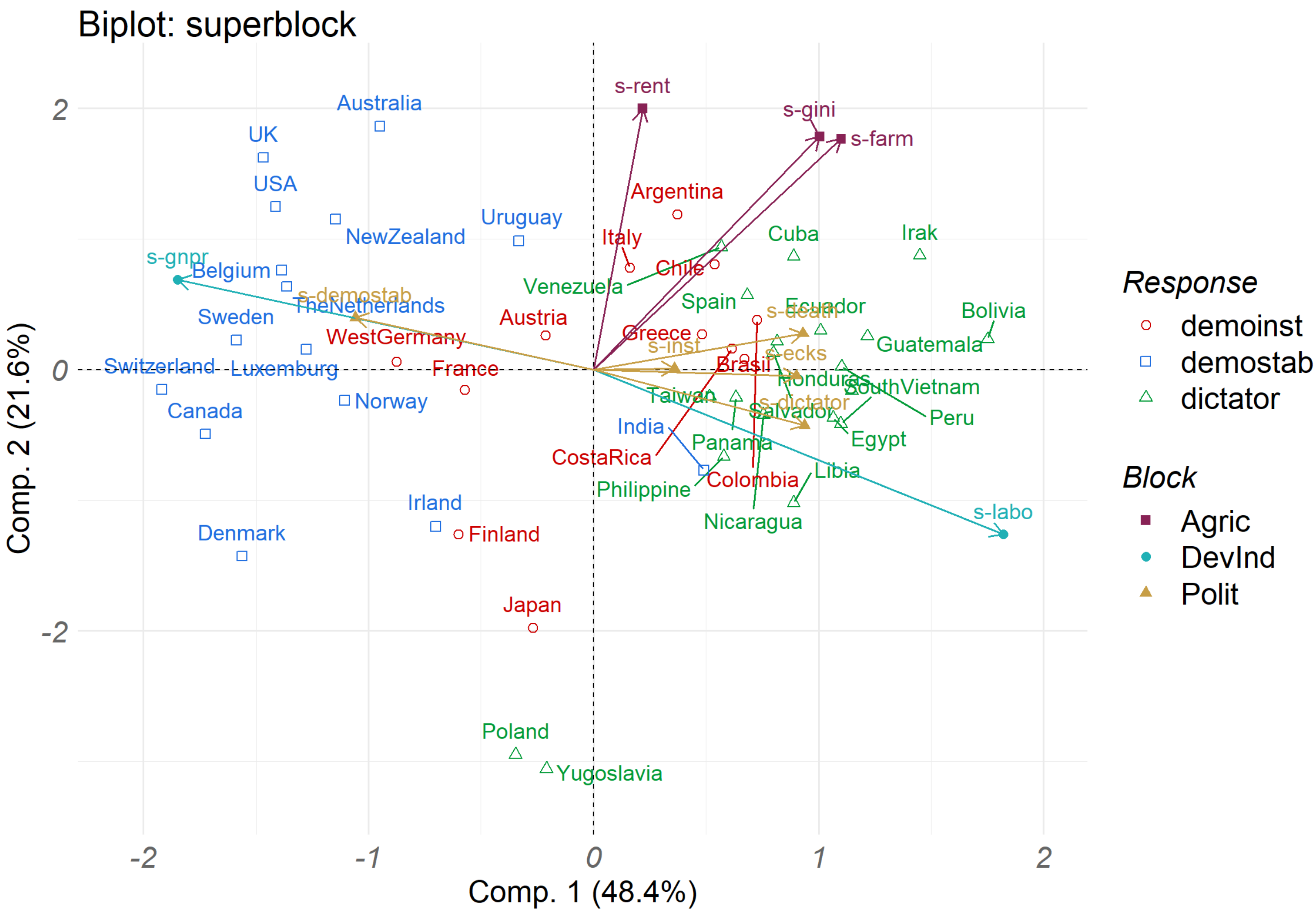
Consensus space of Russett (sample plot)



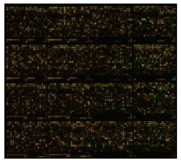
Consensus space of Russett (correlation circle)



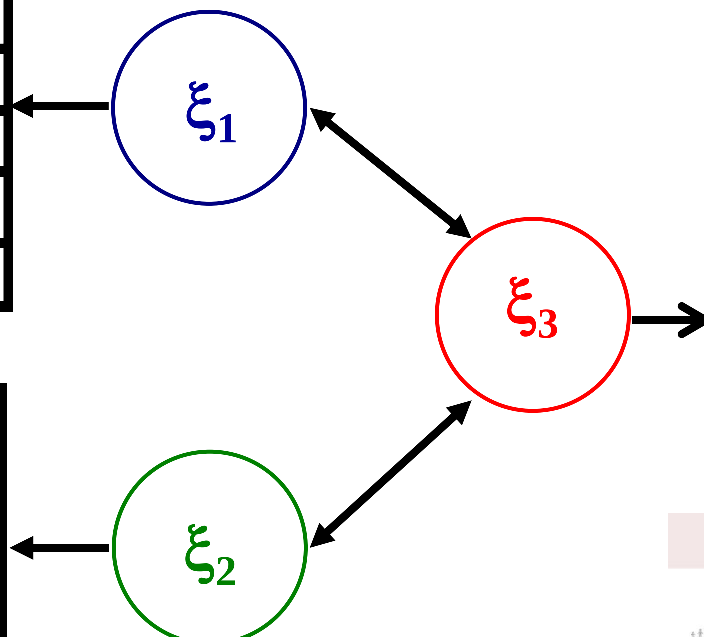
Consensus space of Russett (biplot)



Pediatric glioma data



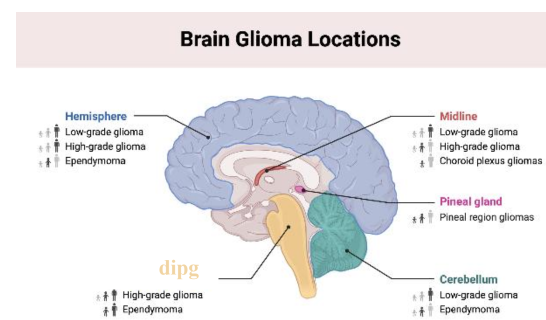
Jacques Grill



X_1	Gene 1	…	Gene 15201
Patient 1	0.18		-0.73
Patient 2	1.15		0.27
Patient 3	1.35		0.22
⋮			
Patient 53	1.39		-0.17

X_2	CGH1	…	CGH 1229
Patient 1	0.00		-0.55
Patient 2	-0.30		0.00
Patient 3	0.33		0.64
⋮			
Patient 53	0.00		0.43

X_3	Localisation
Patient 1	midl
Patient 2	dipg
Patient 3	hemi
⋮	
Patient 53	midl



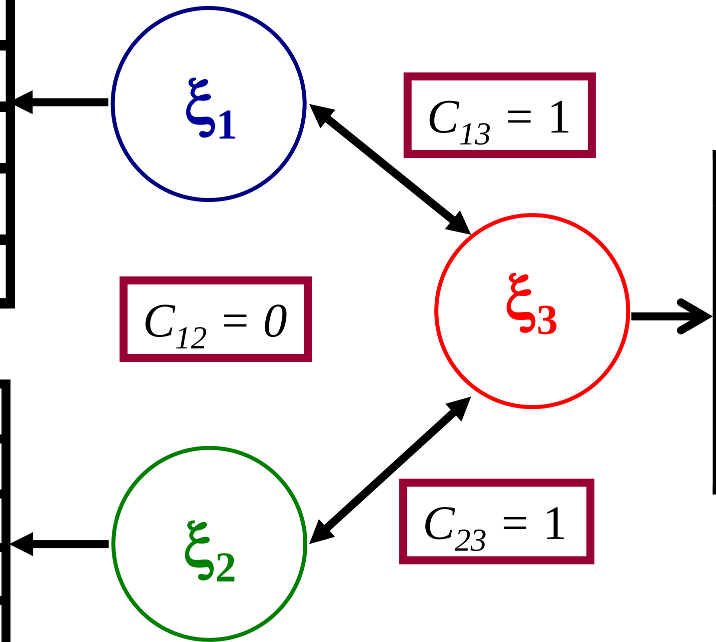
Glioma Cancer Data: from an RGCCA viewpoint

(Department of Pediatric Oncology of the Gustave Roussy Institute)

RGCCA with factorial scheme - $\tau_1 = 1$, $\tau_2 = 1$ and $\tau_3 = 0$

	Gene 1	…	Gene 15201
Patient 1	0.18		-0.73
Patient 2	1.15		0.27
Patient 3	1.35		0.22
⋮			
Patient 53	1.39		-0.17

	CGH1	…	CGH 1229
Patient 1	0.00		-0.55
Patient 2	-0.30		0.00
Patient 3	0.33		0.64
⋮			
Patient 53	0.00		0.43

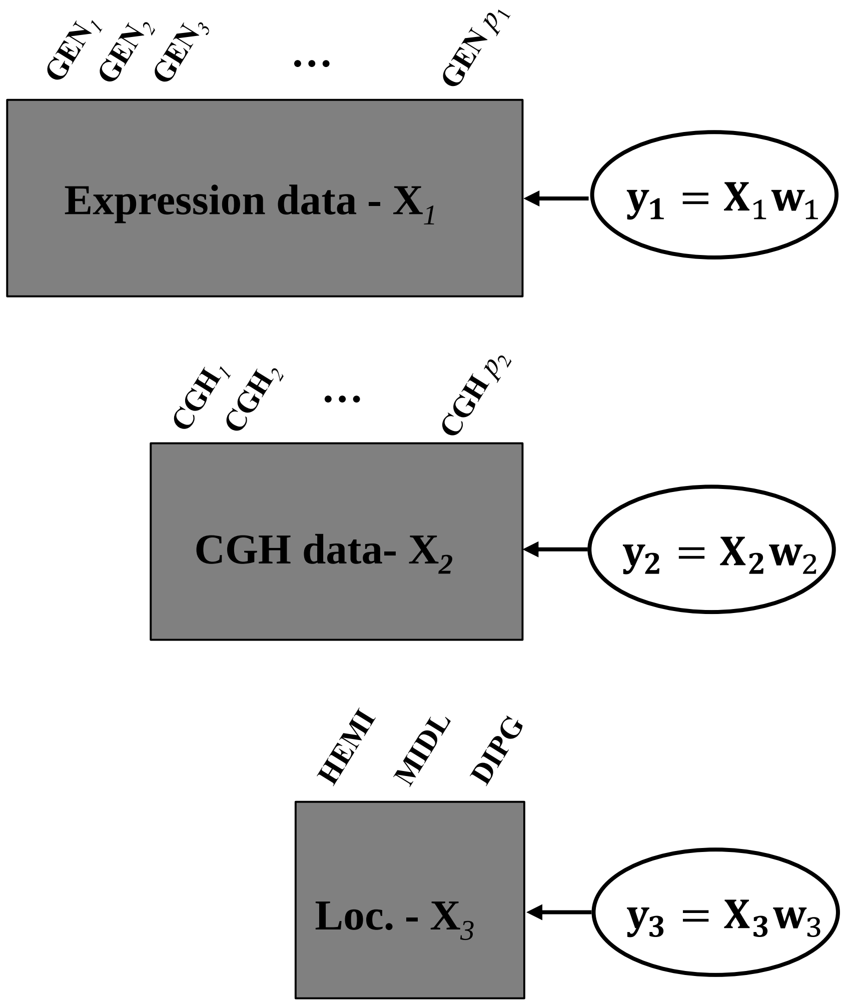


	Hemisphere	DIPG
Patient 1	1	0
Patient 2	0	0
Patient 3	0	1
⋮		
Patient 53	1	0

High dimensional block settings $\Longrightarrow$ dual algorithm for RGCCA

THE R CORNER: KGCCA ON GLIOMA DATA

Multiblock component methods with sparsity



Block components should verified three properties at the same time:

1. Block components well explain their own block.

2. Block components are as correlated as possible for connected blocks.

3. Block components are built from sparse $\mathbf{w}_j$

RGCCA for multiblock analysis

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_J} \sum_{j,k}^{J} c_{jk}\, g\left(\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\right)$$

$$\text{s.t.} \quad \|\mathbf{w}_j\|_2^2 = 1 \quad \& \quad \|\mathbf{w}_j\|_1 \leq s_j = 1, \ldots, J$$

Tenenhaus A., Philippe C., Guillemot V, et al..(2014). Variable Selection for Generalized Canonical Correlation Analysis, Biostatistics, 15 (3) : 569-583
Tenenhaus A. and Guillemot V. (2017): RGCCA Package. http://cran.project.org/web/packages/RGCCA/index.html

Block components

$$\mathbf{y}_1 = \mathbf{X}_1\mathbf{w}_1 = w_{11}\mathbf{Gene}_1 + \cdots + w_{1,15201}\mathbf{Gene}_{15201}$$

$$\mathbf{y}_2 = \mathbf{X}_2\mathbf{w}_2 = w_{21}\mathbf{CGH}_1 + \cdots + w_{2,1909}\mathbf{CGH}_{1909}$$

$$\mathbf{y}_3 = \mathbf{X}_3\mathbf{w}_3 = w_{31}\mathbf{Hemisphere} + w_{32}\mathbf{DIPG}$$

Block components should verify three properties at the same time:

(i) Block components explain their block well.

(ii) Block components are as correlated as possible for connected blocks.

(iii) Block components are built from sparse $\mathbf{w}_j$

Block relaxation: from $\mathbf{w}^s$ to $\mathbf{w}^{s+1}$

$$\mathbf{w}^s = \left(\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_J^s\right)$$

$$\underset{\|\mathbf{w}_1\|_2=1\ \&\ \|\mathbf{w}_1\|_2 \le s_1}{\operatorname{argmax}} h\left(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_J^s\right)$$

$$\to \mathbf{w}_1^{s+1}$$

$$\underset{\|\mathbf{w}_2\|_2=1\ \&\ \|\mathbf{w}_2\|_1 \le s_2}{\operatorname{argmax}} h\left(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \mathbf{w}_3^s \dots, \mathbf{w}_J^s\right)$$

$$\to \mathbf{w}_2^{s+1}$$

$$\vdots$$

$$\underset{\|\mathbf{w}_j\|_2=1\ \&\ \|\mathbf{w}_j\|_1 \le s_j}{\operatorname{argmax}} h\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{j-1}^{s+1}, \mathbf{w}_j, \mathbf{w}_{j+1}^s, \dots, \mathbf{w}_J^s\right)$$

$$\to \mathbf{w}_j^{s+1}$$

$$\vdots$$

$$\underset{\|\mathbf{w}_J\|_2=1\ \&\ \|\mathbf{w}_J\|_1 \le s_J}{\operatorname{argmax}} h\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{J-1}^{s+1}, \mathbf{w}_J\right)$$

$$\to \mathbf{w}_J^{s+1}$$

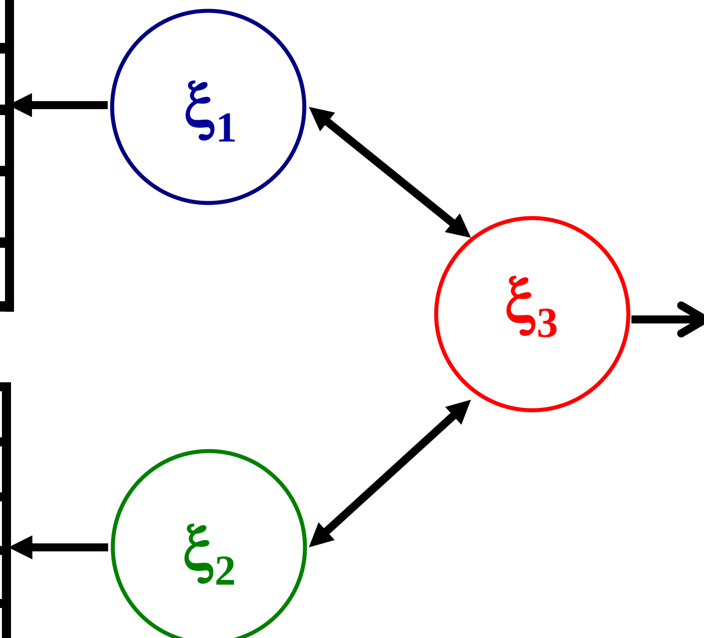
$$\mathbf{w}_j^{s+1} = \frac{\max\left(0, \left|\frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j\right| - \lambda_j\right)}{\left\|\max\left(0, \left|\frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j\right| - \lambda_j\right)\right\|_2}$$

$$\mathbf{w}^{s+1} = \left(\mathbf{w}_1^{s+1}, \mathbf{w}_2^{s+1}, \dots, \mathbf{w}_J^{s+1}\right)$$

THE R CORNER: SGCCA ON GLIOMA DATA

THE R CORNER: SGCCA ON GLIOMA DATA

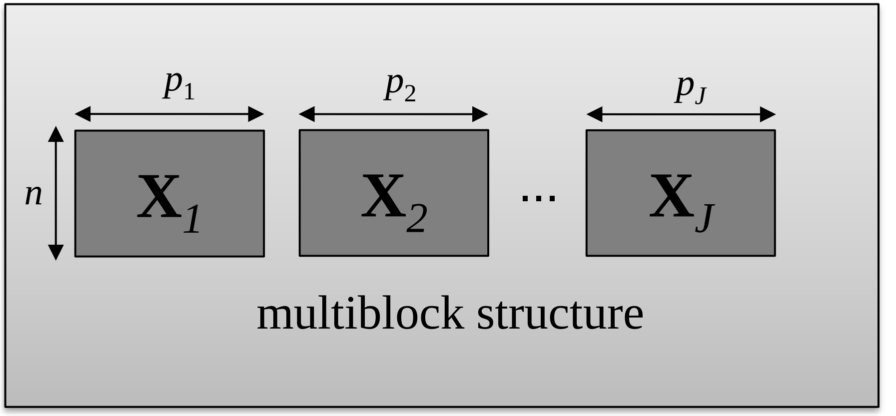
	Gene 1	…	Gene 15201
Patient 1	0.18		-0.73
Patient 2	1.15		0.27
Patient 3	1.35		0.22
⋮			
Patient 53	1.39		-0.17



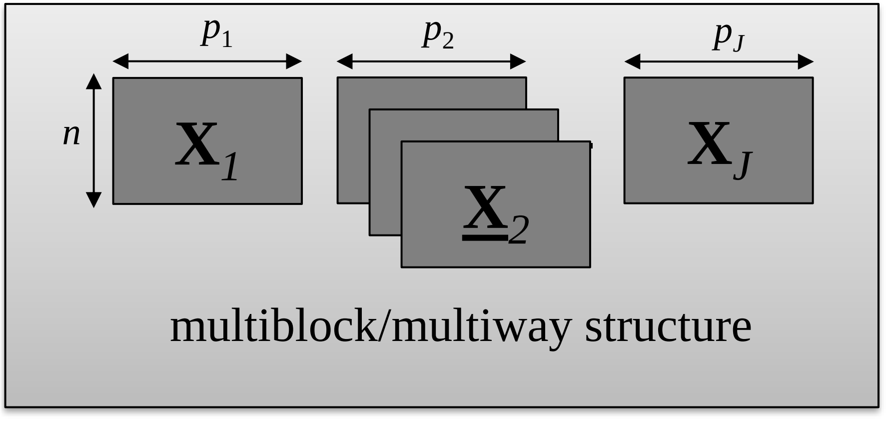
	Hemisphere	DIPG
Patient 1	1	0
Patient 2	0	0
Patient 3	0	1
⋮		
Patient 53	1	0

	CGH1	…	CGH 1229
Patient 1	0.00		-0.55
Patient 2	-0.30		0.00
Patient 3	0.33		0.64
⋮			
Patient 53	0.00		0.43

High dimensional block settings $\Longrightarrow$ sparse GCCA

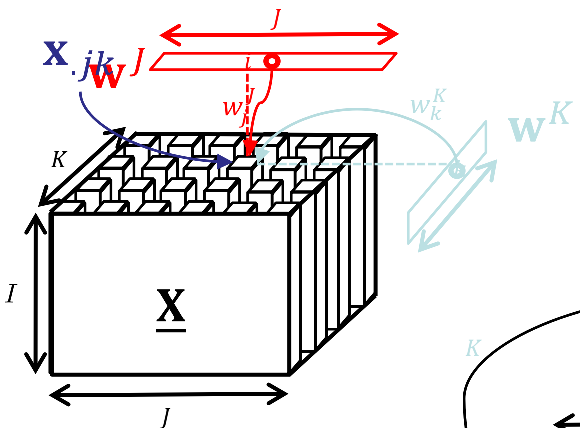


p_1
p_2
p_J
n
$\mathbf{X}_1$
$\mathbf{X}_2$
…
$\mathbf{X}_J$
multiblock structure



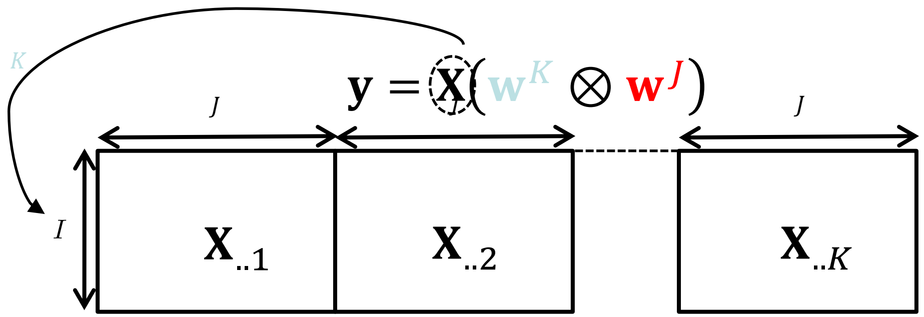
p_1
p_2
p_J
n
$\mathbf{X}_1$
$\underline{\mathbf{X}}_2$
$\mathbf{X}_J$
multiblock/multiway structure

How to handle multiway data in RGCCA



$$\mathbf{y} = \sum_{j=1}^{J} \sum_{k=1}^{K} w_k^K w_j^J \mathbf{x}_{.jk}$$

$$\mathbf{y} = \mathbf{X}\left(\mathbf{w}^K \otimes \mathbf{w}^J\right)$$



The COMA project

(Brain and Spine Institute, La pitié Salpêtrière Hospital)

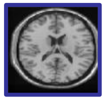
$4 \times p_1$ parameters to estimate

$\mathbf{w}_1$ $\mathbf{w}_2$ $\mathbf{w}_3$ $\mathbf{w}_4$

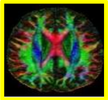
L_1 ? L_t

n ~100

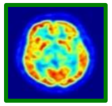
...nctional ...tropy (FA) — p_1 — $\mathbf{X}_1$



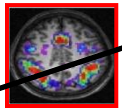
Mean Diffusivity (MD) — p_1 — $\mathbf{X}_2$



p_1 — $\mathbf{X}_3$



p_1 — $\mathbf{X}_4$



Final Phenotype

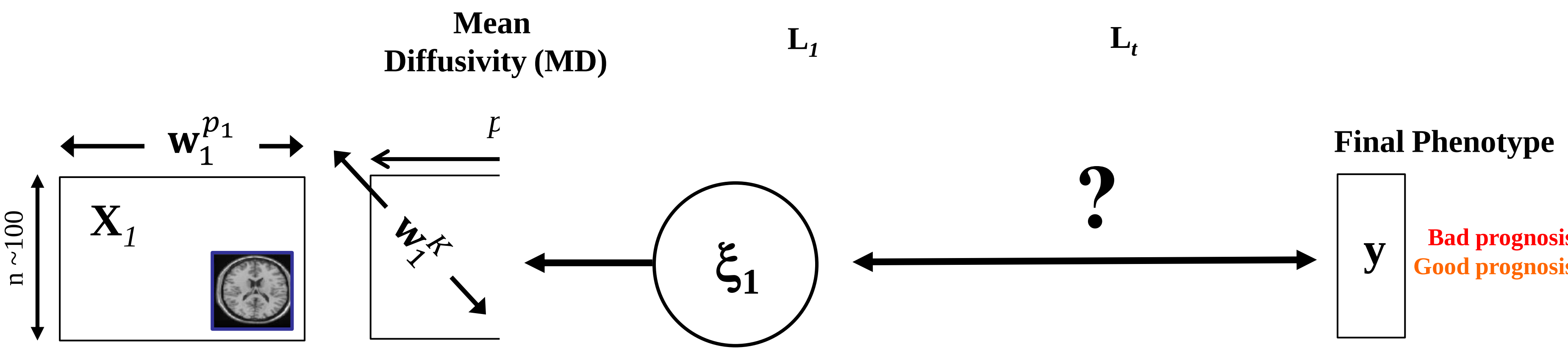
$\mathbf{y}$

Bad prognosis

Good prognosis

The COMA project

(Brain and Spine Institute, La pitié Salpêtrière Hospital)



$4 + p_1$ parameters
instead of $4 \times p_1$

$$\mathbf{y}_1 = \left(\sum_{k=1}^{K} w_{1k}^{K} \mathbf{X}_k \right) \mathbf{w}_1^{p_1}$$

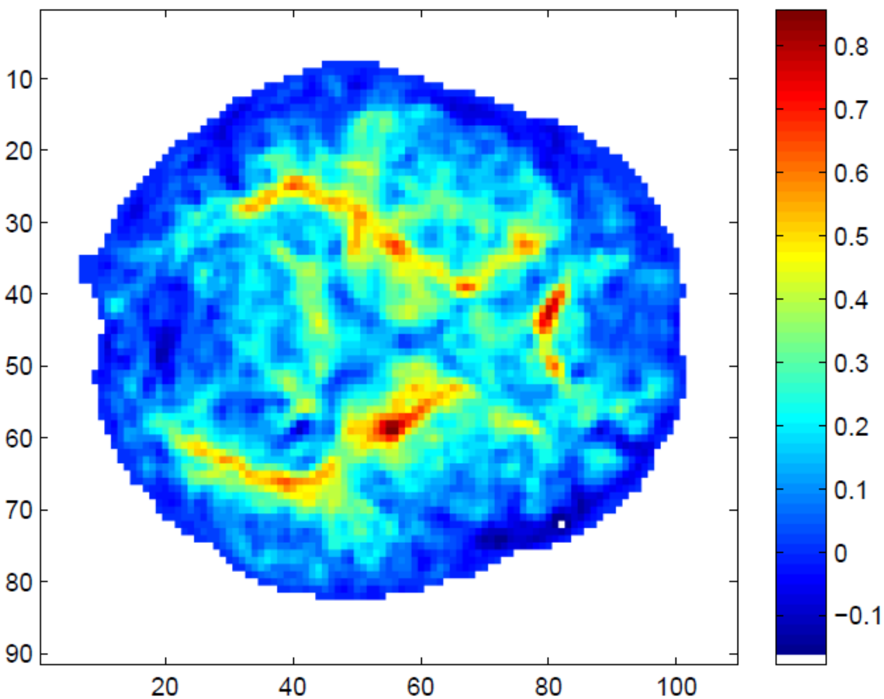
$\mathbf{X}_{(1)}$ $\mathbf{w}_1$

$$\max_{\mathbf{w}_1, \mathbf{w}_2} \operatorname{cov}\left(\mathbf{X}_{(1)} \mathbf{w}_1, \mathbf{y}\right) \quad \text{s.t.} \quad \begin{cases} \mathbf{w}_1^{\top} \mathbf{w}_1 = 1 \\ \mathbf{w}_1 = \mathbf{w}_1^{K} \otimes \mathbf{w}_1^{J} \end{cases}$$

COMA project

Contribution of the voxels and the modalities to predict the long term recovery of patients after traumatic brain injury can be studied separately.

Influence of spatial positions: $\mathbf{w}_1^{p_1}$

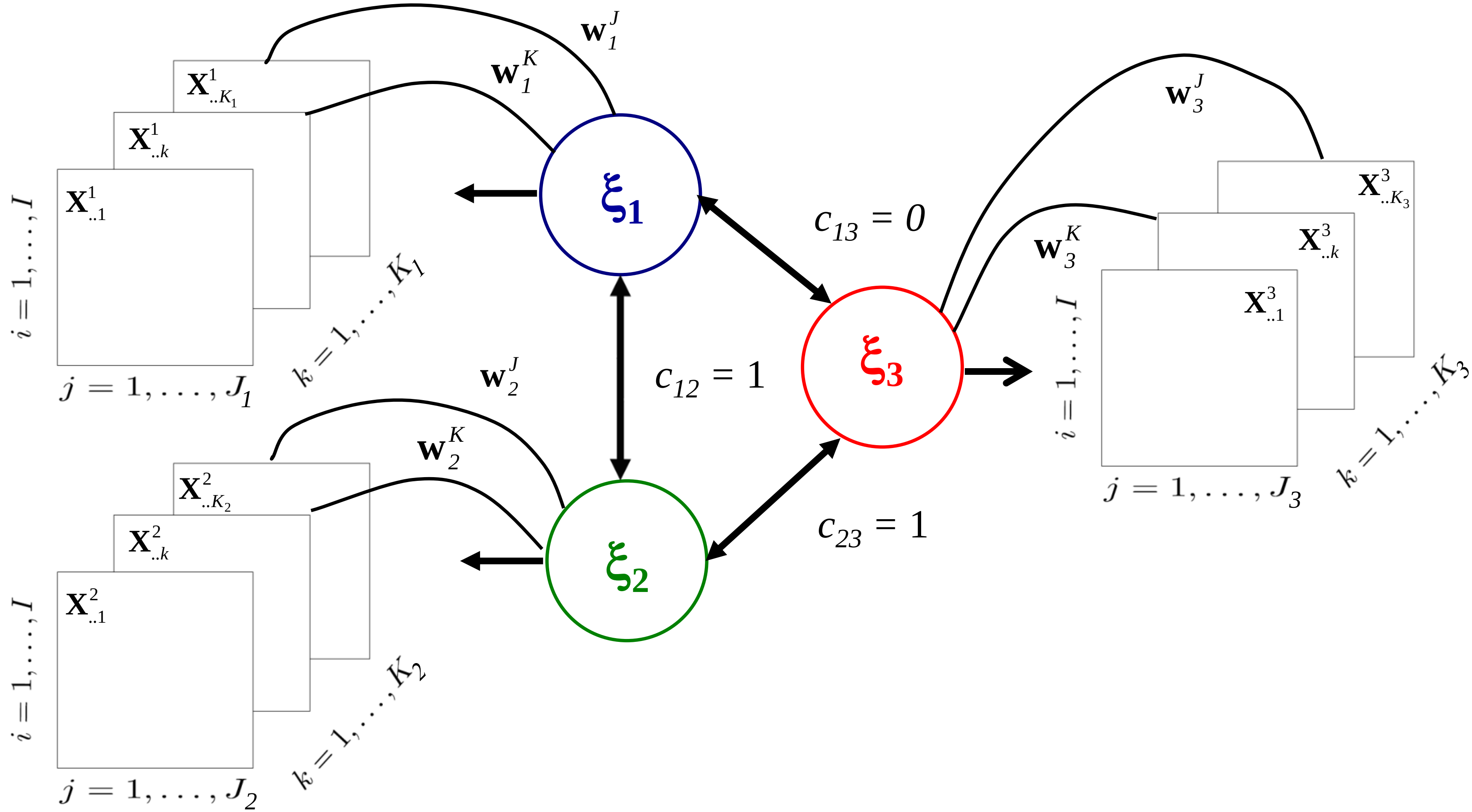


Discriminating voxels within the white matter bundles

Influence of spatial positions: $\mathbf{w}_1^{K}$

Modality	$\mathbf{w}_1^{K}$
FA	**0.9887**
MD	0.0036
$\mathbf{L_1}$	0.0046
$\mathbf{L_t}$	0.0031

MGCCA optimization problem



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_J} \sum_{j,k=1}^{J} c_{jk}\, g\left(\mathrm{cov}\left(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k\right)\right) \quad \text{s.t.} \quad \mathbf{w}_j^\top \mathbf{M}_j \mathbf{w}_j = 1 \text{ and } \mathbf{w}_j = \mathbf{w}_j^K \otimes \mathbf{w}_j^J, j = 1, \ldots, J$$

Block relaxation: from $\mathbf{w}^s$ to $\mathbf{w}^{s+1}$

$$\mathbf{w}^s = \left(\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_J^s\right)$$

$$\underset{\|\mathbf{w}_1\|_2=1\ \&\ \|\mathbf{w}_1\|_2 \le s_1}{\operatorname{argmax}} h\left(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_J^s\right)$$

$$\rightarrow \mathbf{w}_1^{s+1}$$

$$\underset{\|\mathbf{w}_2\|_2=1\ \&\ \|\mathbf{w}_2\|_1 \le s_2}{\operatorname{argmax}} h\left(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \mathbf{w}_3^s \dots, \mathbf{w}_J^s\right)$$

$$\rightarrow \mathbf{w}_2^{s+1}$$

$$\vdots$$

$$\underset{\|\mathbf{w}_j\|_2=1\ \&\ \|\mathbf{w}_j\|_1 \le s_j}{\operatorname{argmax}} h\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{j-1}^{s+1}, \mathbf{w}_j, \mathbf{w}_{j+1}^s, \dots, \mathbf{w}_J^s\right)$$

$$\rightarrow \mathbf{w}_j^{s+1}$$

$$\vdots$$

$$\underset{\|\mathbf{w}_J\|_2=1\ \&\ \|\mathbf{w}_J\|_1 \le s_J}{\operatorname{argmax}} h\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{J-1}^{s+1}, \mathbf{w}_J\right)$$

$$\rightarrow \mathbf{w}_J^{s+1}$$

$$\mathbf{w}_j^{s+1} = \mathbf{w}_j^K \otimes \mathbf{w}_j^J$$

where $\mathbf{w}_j^K$ and $\mathbf{w}_j^J$ are obtained as the first left and right singular vector of a certain matrix of dimension $K_j \times J_j$

$$\mathbf{w}^{s+1} = \left(\mathbf{w}_1^{s+1}, \mathbf{w}_2^{s+1}, \dots, \mathbf{w}_J^{s+1}\right)$$

Consensus space with RGCCA

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_{J+1}} \sum_{j=1}^{J} \left(\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_{J+1}\mathbf{w}_{J+1})\right)^m \text{ s.t. } \mathbf{w}_j^\top \mathbf{M}_j \mathbf{w}_j = 1, \forall j$$

The superblock component $\mathbf{y}_{J+1}$ is proportional to:

$$\mathbf{y}_{J+1} \propto \mathbf{X}_{J+1}\mathbf{M}_{J+1}^{-1}\mathbf{X}_{J+1}^\top \boxed{\sum_{j=1}^{J} \left(\mathrm{cov}(\mathbf{y}_j, \mathbf{y}_{J+1})\right)^{m-1} \mathbf{y}_j}$$

When $\mathbf{M}_{J+1} = n^{-1}\mathbf{X}_{J+1}^\top\mathbf{X}_{J+1}$, it reduces to

$$\mathbf{y}_{J+1} \propto \sum_{j=1}^{J} \left(\mathrm{cov}(\mathbf{y}_j, \mathbf{y}_{J+1})\right)^{m-1} \mathbf{y}_j$$

weighted sums of
block components

Tenenhaus M, Tenenhaus A, Groenen PJF, (2017) Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods, Psychometrika, vol. 82, no. 3, 737–777

Girka F, Camenen E, Caroline P, Gloaguen A, Guillemot V, Le Brusquet L, Tenenhaus A (2023): RGCCA Package. http://cran.project.org/web/packages/RGCCA/index.html

Consensus space with RGCCA

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_J} \sum_{j=1}^{J} \mathrm{cov}\left(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_{J+1} \mathbf{w}_{J+1}\right)^m \quad \mathrm{s.t.} \begin{cases} \left\|\mathbf{w}_j\right\| = \cdots = \left\|\mathbf{w}_J\right\| = 1 \\ \mathrm{var}\left(\mathbf{X}_{J+1} \mathbf{w}_{J+1}\right) = 1 \end{cases}$$

$m = 1$

$$\mathbf{y}_{J+1} \propto \sum_{j=1}^{J} \mathbf{y}_j$$

$m = 2$

$$\mathbf{y}_{J+1} \propto \sum_{j=1}^{J} \mathrm{cov}\left(\mathbf{y}_j, \mathbf{y}_{J+1}\right) \mathbf{y}_j$$

$m = 4$

$$\mathbf{y}_{J+1} \propto \sum_{j=1}^{J} \mathrm{cov}\left(\mathbf{y}_j, \mathbf{y}_{J+1}\right)^2 \mathbf{y}_j$$

Fairness and block selection behavior

Will a solution be accepted as a good one even if it is dominated by only a few of the J sets, ignoring the other sets? Or do we require that all J sets have equal share in the solution? (Van de Geer, 1984)

The stationary equations of CPCA(m) give some information:

weighted sums of block components

$$\mathbf{y}_{J+1} \propto \mathbf{X}_{J+1}\mathbf{M}_{J+1}^{-1}\mathbf{X}_{J+1}^{\top}\boxed{\sum_{j=1}^{J}\left(\mathrm{cov}\left(\mathbf{y}_j, \mathbf{y}_{J+1}\right)\right)^{m-1}\mathbf{y}_j}$$

The influence of the various blocks $\mathbf{X}_j$ on the solution is related to the scale of the block component $\mathbf{y}_k$ and to the weights $\mathrm{cov}\left(\mathbf{y}_j, \mathbf{y}_{J+1}\right)^{m-1}$.

Fairness and block selection behavior

$$\mathbf{y}_{J+1} \propto \mathbf{X}_{J+1}\mathbf{M}_{J+1}^{-1}\mathbf{X}_{J+1}^{\top} \boxed{\sum_{j=1}^{J} \left(\mathrm{cov}\left(\mathbf{y}_j, \mathbf{y}_{J+1}\right)\right)^{m-1} \mathbf{y}_j}$$

weighted sums of block components

► Methods with equal block component scales are fairer than methods with unequal block component scales

⇒ Correlation-based methods are fairer than covariance-based methods

► Methods with equal weights are fairer than methods with unequal weights.

⇒ Using $g(x) = x$ or $g(x) = |x|$ scheme functions lead to fair methods. Block selection behavior is favored by using the scheme function $g(x) = x^m$ where m is a positive even integer.

The core optimization of sGCCA

Proposition. Consider the optimization problem

$$\max_{\mathbf{w}_j} \frac{1}{n}\mathbf{w}_j^\top \mathbf{X}_j^\top \mathbf{z}_j \quad \text{subject to} \quad \left\|\mathbf{w}_j\right\|_2 = 1 \;\&\; \left\|\mathbf{w}_j\right\|_1 \leq s_j$$

The solution satisfies $\mathbf{w}_j = \frac{\mathcal{S}\left(\frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j, \lambda_j\right)}{\left\|\mathcal{S}\left(\frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j, \lambda_j\right)\right\|_2}$, with $\mathcal{S}$ corresponds to soft-thresholding operator and $\lambda_j = 0$ if this results in $\left\|\mathbf{w}_j\right\|_1 \leq s_j$; otherwise, s_j is chosen so that $\|\mathbf{u}_1\|_1 = s_j$.

Proof. (with subgradient)

First, we rewrite the optimization problem using Lagrange multipliers:

$$\mathcal{L} = \frac{1}{n}\mathbf{w}_j^\top \mathbf{X}_j^\top \mathbf{z}_j - \lambda_2\left(\left\|\mathbf{w}_j\right\|_2^2 - 1\right) - \lambda_1\left(\left\|\mathbf{w}_j\right\|_1 - s_j\right)$$

The core optimization of sGCCA

Considering the partial subgradient of the Lagrangian function with respect to $\mathbf{w}_j$ yields :

$$\partial_{\mathbf{w}_j}\mathcal{L} = \frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j - 2\lambda_{2j}\mathbf{w}_j - \lambda_{1j}\boldsymbol{\gamma}_j, \qquad j = 1, \dots, J$$

where the kth element γ_{jk} of $\boldsymbol{\gamma}_j$ is the subgradient of $\sum_{k=1}^{p_j}|w_{jk}|$ with respect to w_{jk}, defined as:

$$\gamma_{jk} = \begin{cases} \operatorname{sign}(w_{jk}) \text{ if } w_{jk} \neq 0 \\ [-1; 1] \quad \text{if } w_{jk} = 0 \end{cases}$$

Therefore, we obtain:

$$\partial_{w_{jk}}\mathcal{L} = \begin{cases} \frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j - 2\lambda_{2j}w_{jk} - \lambda_{1j}\operatorname{sign}(w_{jk}) & \text{if } w_{jk} \neq 0 \\ \left[\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j - \lambda_{1j} ; \frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j + \lambda_{1j}\right] & \text{if } w_{jk} = 0 \end{cases}$$

The core optimization of sGCCA

$$\partial_{w_{jk}}\mathcal{L} = \begin{cases} \frac{1}{n}\mathbf{x}_{jk}^{\top}\mathbf{z}_j - 2\lambda_{2j}w_{jk} - \lambda_{1j}\mathrm{sign}(w_{jk}) & \text{if } w_{jk} \neq 0 \\ \left[\frac{1}{n}\mathbf{x}_{jk}^{\top}\mathbf{z}_j - \lambda_{1j} \,; \frac{1}{n}\mathbf{x}_{jk}^{\top}\mathbf{z}_j + \lambda_{1j}\right] & \text{if } w_{jk} = 0 \end{cases}$$

At the optimum, we must have $0 \in \partial_{\mathbf{w}_j}\mathcal{L}$ and we get:

$$\begin{cases} \frac{1}{n}\mathbf{x}_{jk}^{\top}\mathbf{z}_j - 2\lambda_{2j}w_{jk} - \lambda_{1j}\mathrm{sign}(w_{jk}) = 0 & \text{if } w_{jk} \neq 0 \\ \left|\frac{1}{n}\mathbf{x}_{jk}^{\top}\mathbf{z}_j\right| < \lambda_{1j} & \text{if } w_{jk} = 0 \end{cases}$$

The core optimization of sGCCA

$$\begin{cases} \frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j - 2\lambda_{2j} w_{jk} - \lambda_{1j}\operatorname{sign}(w_{jk}) = 0 & \text{if } w_{jk} \neq 0 \\ \left|\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| < \lambda_{1j} & \text{if } w_{jk} = 0 \end{cases}$$

From $\operatorname{sign}\left(\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right) = \operatorname{sign}(w_{jk})$

$$w_{jk} = \begin{cases} \frac{1}{2\lambda_{2j}}\left(\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j - \lambda_{1j}\operatorname{sign}\left(\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right)\right) & \text{if } w_{jk} \neq 0 \\ 0 & \text{if } \left|\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| < \lambda_{1j} \end{cases}$$

$$w_{jk} = \begin{cases} \frac{1}{2\lambda_{2j}}\operatorname{sign}\left(\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right)\left(\left|\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| - \lambda_{1j}\right) & \text{if } w_{jk} \neq 0 \\ 0 & \text{if } \left|\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| < \lambda_{1j} \end{cases}$$

The core optimization of sGCCA

$$w_{jk} = \begin{cases} \dfrac{1}{2\lambda_{2j}} \operatorname{sign}\left(\dfrac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right)\left(\left|\dfrac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| - \lambda_{1j}\right) & \text{if } w_{jk} \neq 0 \\ 0 & \text{if } \left|\dfrac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| < \lambda_{1j} \end{cases}$$

$$w_{jk} = \frac{1}{2\lambda_{2j}} \operatorname{sign}\left(\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right) \max\left(0, \left|\frac{1}{n}\mathbf{x}_{jk}^\top \mathbf{z}_j\right| - \lambda_{1j}\right)$$

where $\lambda_{1\mathrm{j}}$ is chosen such that $\|\mathbf{w}_j\|_1 = s_j$ (e.g. binary search) and λ_2 such that $\|\mathbf{w}_j\|_2 = 1$.

$$\mathbf{w}_j = \frac{\mathcal{S}\left(\frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j, \lambda_{1j}\right)}{\left\|\mathcal{S}\left(\frac{1}{n}\mathbf{X}_j^\top \mathbf{z}_j, \lambda_{1j}\right)\right\|_2}, \qquad j = 1, \dots, J$$