(Kernel) Regularized Generalized Canonical Correlation Analysis

Arthur Tenenhaus

Michel Tenenhaus

Strasbourg, 17/12/2010

References

- Paper

Arthur & Michel Tenenhaus

Regularized Generalized CCA

Psychometrika (2011)

- R package

New package RGCCA with initial version 1.0
Title: Regularized Generalized Canonical Correlation Analysis
Version: 1.0
Date: 2010-06-08
Author: Arthur Tenenhaus
Repository: CRAN
Date/Publication: 2010-10-15 14:58:02
More information about RGCCA at CRAN
Path: /cran/new | permanent link

Economic inequality and political instability Data from Russett (1964), in GIFI

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)

	X1			X2		X3			
	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	3.27	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

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

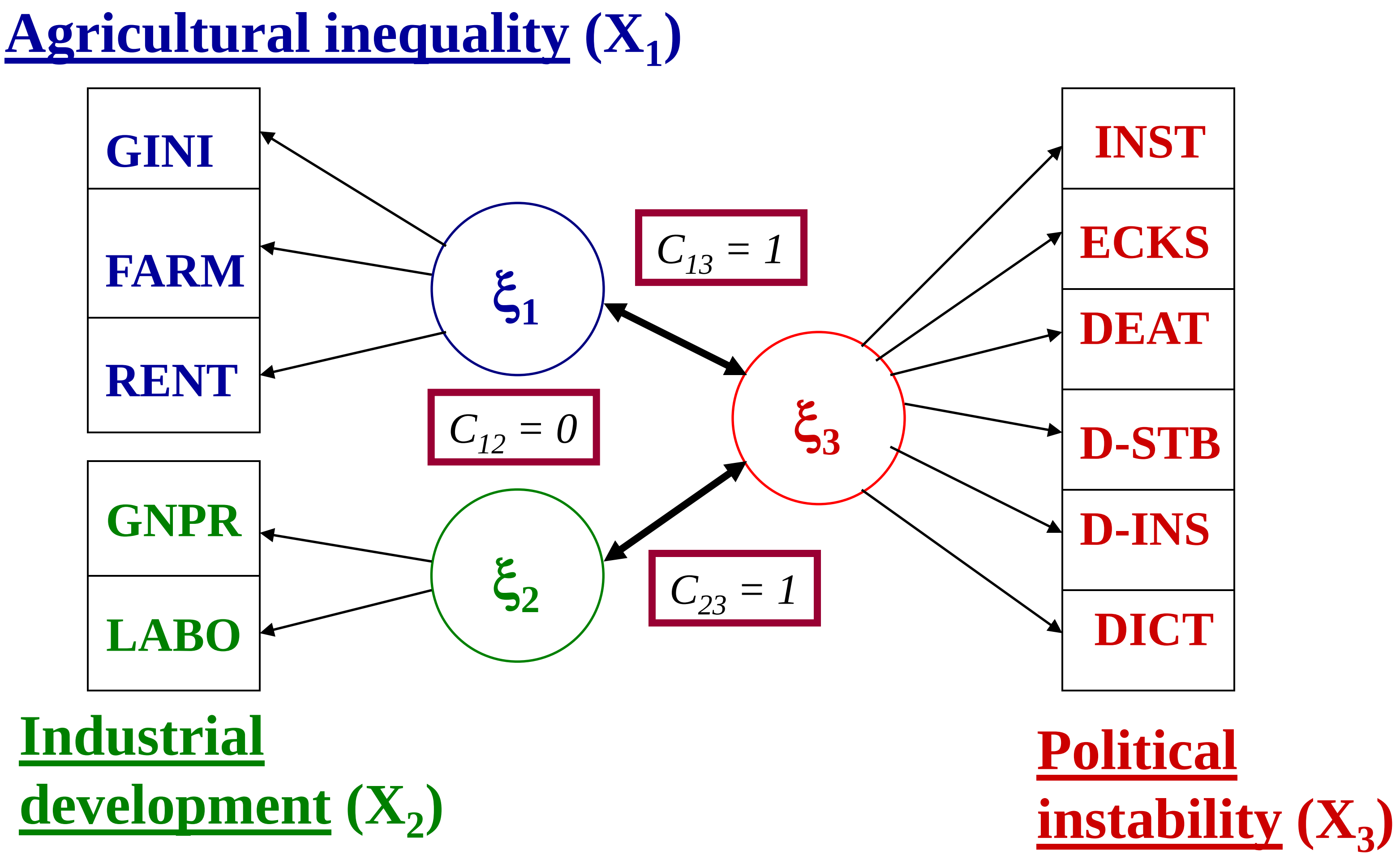
INST : Instability of executive (45-61)

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

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

DEMO : Stable democracy (1), Unstable democracy (2) or Dictatorship (3)

Structural relation between blocks



Block components

$$Y_1 = X_1 w_1 = w_{11} GINI + w_{12} FARM + w_{13} RENT$$

$$Y_2 = X_2 w_2 = w_{21} GNPR + w_{22} LABO$$

$$Y_3 = X_3 w_3 = w_{31} INST + w_{32} ECKS + w_{33} DEATH$$
$$+ w_{34} D\text{-}STB + w_{35} D\text{-}UNST$$
$$+ w_{36} DICT$$

Some modified multi-block methods

SUMCOR (Horst, 1961)	$\underset{j,k}{Max\sum} Cor(X_j w_j, X_k w_k)$
S[illegible]	$\sum_{j,k}$ [illegible]
SABSCOR (Mathes, 1993, Hanafi, 2004)	$\underset{j,k}{Max\sum} \lvert Cor(X_j w_j, X_k w_k) \rvert$
MAXDIFF (Van de Geer, 1984) [SUMCOV]	$\underset{\text{All } \lVert w_j \rVert = 1}{Max} \sum_{j,k} Cov(X_j w_j, X_k w_k)$
M[illegible] [SSQCOV]	$\underset{\text{All } \lVert w_j \rVert = 1}{Max} \sum_{j,k}$ [illegible]
SABSCOV (Krämer, 2007)	$\underset{\text{All } \lVert w_j \rVert = 1}{Max} \sum_{j,k} \lvert Cov(X_j w_j, X_k w_k) \rvert$

GENERALIZED CANONICAL CORRELATION ANALYSIS

GENERALIZED CANONICAL COVARIANCE ANALYSIS

Covariance-based criteria

c_{jk} = 1 if blocks are linked, 0 otherwise and c_{jj} = 0

SUMCOR	$\underset{\text{All } Var(X_j w_j)=1}{Max} \sum_{j,k} c_{jk} Cov(X_j w_j, X_k w_k)$
SSQCOR	$\underset{\text{All } Var(X_j w_j)=1}{Max} \sum_{j,k} c_{jk} Cov^2(X_j w_j, X_k w_k)$
SABSCOR	$\underset{\text{All } Var(X_j w_j)=1}{Max} \sum_{j,k} c_{jk} \lvert Cov(X_j w_j, X_k w_k) \rvert$
SUMCOV	$\underset{\text{All } \lVert w_j \rVert=1}{Max} \sum_{j,k} c_{jk} Cov(X_j w_j, X_k w_k)$
SSQCOV	$\underset{\text{All } \lVert w_j \rVert=1}{Max} \sum_{j,k} c_{jk} Cov^2(X_j w_j, X_k w_k)$
SABSCOV	$\underset{\text{All } \lVert w_j \rVert=1}{Max} \sum_{j,k} c_{jk} \lvert Cov(X_j w_j, X_k w_k) \rvert$

Starting point : RGGCA (for linear dependence)

$$\underset{\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_J}{\operatorname{argmax}} \sum_{j \neq k}^{J} c_{jk}\, \mathrm{g}\left(\operatorname{cov}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_k \mathbf{a}_k)\right)$$

Subject to the constraints $(1 - \tau_j)\operatorname{var}(\mathbf{X}_j \mathbf{a}_j) + \tau_j \|\mathbf{a}_j\|^2 = 1, j = 1, \ldots, J$

A monotone convergent algorithm related to this optimization problem will be described.

where

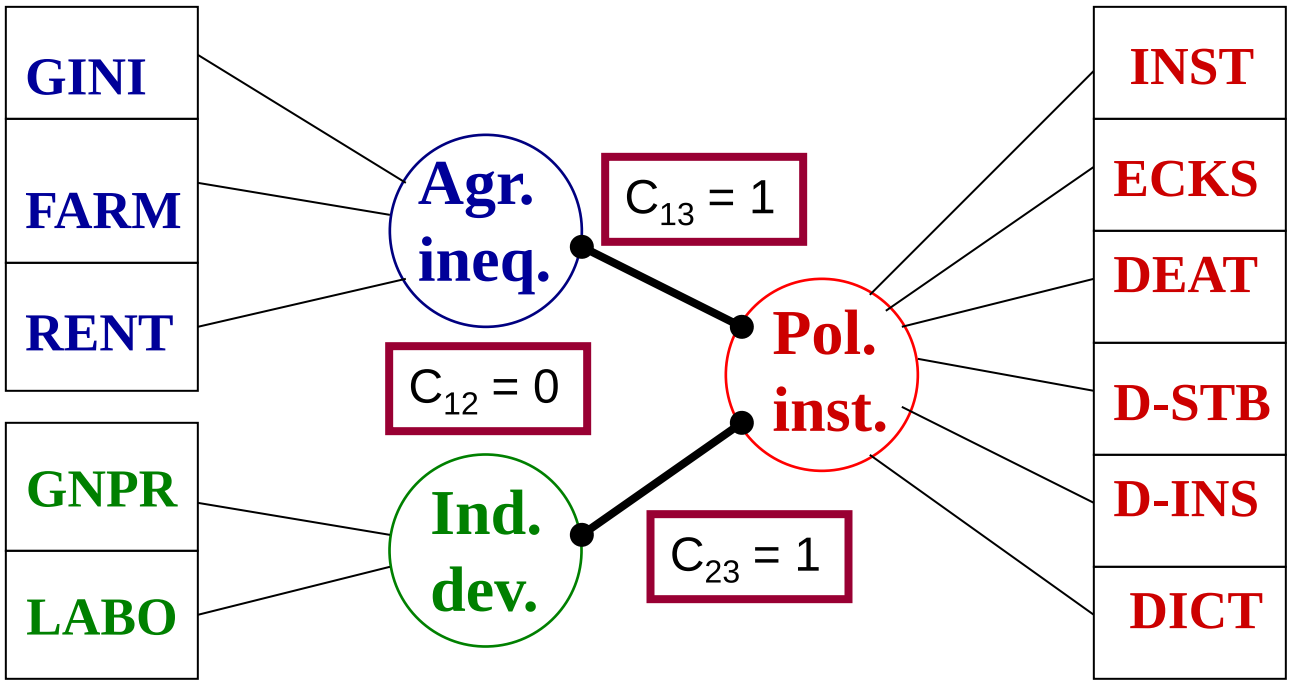
0 otherwise

$$g = \begin{cases} \text{identity} & \text{(Horst sheme)} \\ \text{square} & \text{(Factorial scheme)} \\ \text{abolute value} & \text{(Centroid scheme)} \end{cases}$$

and: $\tau_j =$ Shrinkage constant between 0 and 1

RGCCA applied to the Russett data

Agricultural inequality (X_1)



Industrial development (X_2)

Political instability (X_3)

$$\underset{a_1, a_2, a_3}{\text{Maximize}} \quad g(\text{Cov}(X_1 a_1, X_3 a_3)) + g(\text{Cov}(X_2 a_2, X_3 a_3))$$

$$\text{subject to the constraints} \quad \tau_j \|a_j\|^2 + (1-\tau_j) Var(X_j a_j) = 1, \quad j = 1, 2, 3$$

$0 \leq \tau_j \leq 1$, g = identity, square or absolute value

Construction of monotone convergent algorithms for these criteria

- Construct the Lagrangian function related to the optimization problem.
- Cancel the derivative of the Lagrangian function with respect to each w_j.
- Use the Wold's procedure to solve the stationary equations (≈ Gauss-Seidel algorithm).
- This procedure is monotonically convergent: the criterion increases at each step of the algorithm.

The general algorithm

Initial step

w_j

$$y_j = X_j w_j$$

Outer Estimation
(explains the block)

$$\tau_j \|w_j\|^2 + (1-\tau_j) Var(X_j w_j) = 1$$

$$\mathbf{z}_j = \sum_{k \neq j} e_{jk} \mathbf{y}_k$$

Inner Estimation
(explains relations between blocks)

Iterate until convergence of the criterion.

$$\mathrm{w}_j = \frac{\left(\left(1-\tau_j\right)\frac{1}{\boldsymbol{n}}\boldsymbol{X}_j^t\boldsymbol{X}_j + \tau_j \mathrm{I}\right)^{-1}\boldsymbol{X}_j^t \mathrm{z}_j}{\sqrt{\mathrm{z}_j^t \boldsymbol{X}_j \left(\left(1-\tau_j\right)\frac{1}{\boldsymbol{n}}\boldsymbol{X}_j^t\boldsymbol{X}_j + \tau_j \mathrm{I}\right)^{-1}\boldsymbol{X}_j^t \mathrm{z}_j}}$$

<u>Choice of weights e_{jh}:</u>

- Horst : $e_{jh} = c_{jh}$
- Centroid : $e_{jh} = c_{jh}\,\mathrm{sign}(cor(Y_h, Y_j))$
- Factorial : $e_{jh} = c_{jh}\,\mathrm{cov}(Y_h, Y_j)$

$c_{jh} = 1$ if blocks are linked, 0 otherwise and $c_{jj} = 0$

special cases of RGCCA (among others)

PLS regression:	Wold S., Martens & Wold H. (1983): The multivariate calibration problem in chemistry solved by the PLS method. In Proc. Conf. Matrix Pencils, Ruhe A. & Kåstrøm B. (Eds), March 1982, Lecture Notes in Mathematics, Springer Verlag, Heidelberg, p. 286-293**.**
Redundancy analysis :	Barker M. & Rayens W. (2003): Partial least squares for discrimination, *Journal of Chemometrics*, 17, 166-173.
Regularized CCA :	Vinod H. D. (1976): Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4, 147–166.
Inter-battery factor Analysis:	Tucker L.R. (1958): An inter-battery method of factor analysis, *Psychometrika*, vol. 23, n°2, pp. 111-136.
MCOA :	Chessel D. and Hanafi M. (1996): Analyse de la co-inertie de *K* nuages de points. *Revue de Statistique Appliquée*, 44, 35-60
SSQCOV:	Hanafi M. & Kiers H.A.L. (2006): Analysis of K sets of data, with differential emphasis on agreement between and within sets, *Computational Statistics & Data Analysis*, 51, 1491-1508.
SUMCOR:	Horst P. (1961): Relations among m sets of variables, *Psychometrika*, vol. 26, pp. 126-149.
SSQCOR:	Kettenring J.R. (1971): Canonical analysis of several sets of variables, *Biometrika*, 58, 433-451
MAXDIFF :	Van de Geer J. P. (1984): Linear relations among k sets of variables. *Psychometrik*a, 49, 70-94.
PLS path modeling: (mode B)	Tenenhaus M., Esposito Vinzi V., Chatelin Y.-M., Lauro C. (2005): PLS path modeling. *Computational Statistics and Data Analysis*, 48, 159-205.
Generalized orthogonal multiple co-inertia Analysis:	Vivien M. & Sabatier R. (2003): Generalized orthogonal multiple co-inertia analysis (-PLS): new multiblock component and regression methods, *Journal of Chemometrics*, 17, 287-301.
Caroll's RGCCA :	Takane Y., Hwang H. and Abdi H. (2008): Regularized Multiple-set Canonical Correlation Analysis, *Psychometrika*, 73 (4):753-775
Caroll 's GCCA :	Carroll, J.D. (1968): A generalization of canonical correlation analysis to three or more sets of variables, *Proc. 76th Conv. Am. Psych. Assoc*., pp. 227-228.

The two-block case: Regularized CCA

$$\text{Maximize Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$$
$$\text{subject to } \tau_j \|\mathbf{a}_j\| + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1$$

Special cases

Method	Criterion	Constraints
PLS regression	$\text{Maximize Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$	$\|\mathbf{a}_1\| = \|\mathbf{a}_2\| = 1$
Canonical Correlation Analysis	$\text{Maximize Cor}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$	$\text{Var}(\mathbf{X}_1\mathbf{a}_1) = \text{Var}(\mathbf{X}_2\mathbf{a}_2) = 1$
Redundancy analysis of X_1 with respect to X_2	Maximize $\text{Cor}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)\text{Var}(\mathbf{X}_1\mathbf{a}_1)^{1/2}$	$\|\mathbf{a}_1\| = 1$ $\text{Var}(\mathbf{X}_2\mathbf{a}_2) = 1$

Components X_1a_1 and X_2a_2 are well correlated.

1st component is stable

No stability condition for 2nd component

The two-block case: Regularized CCA

$$\text{Maximize } \text{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$$
$$\text{subject to } \tau_j \|\mathbf{a}_j\| + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1$$

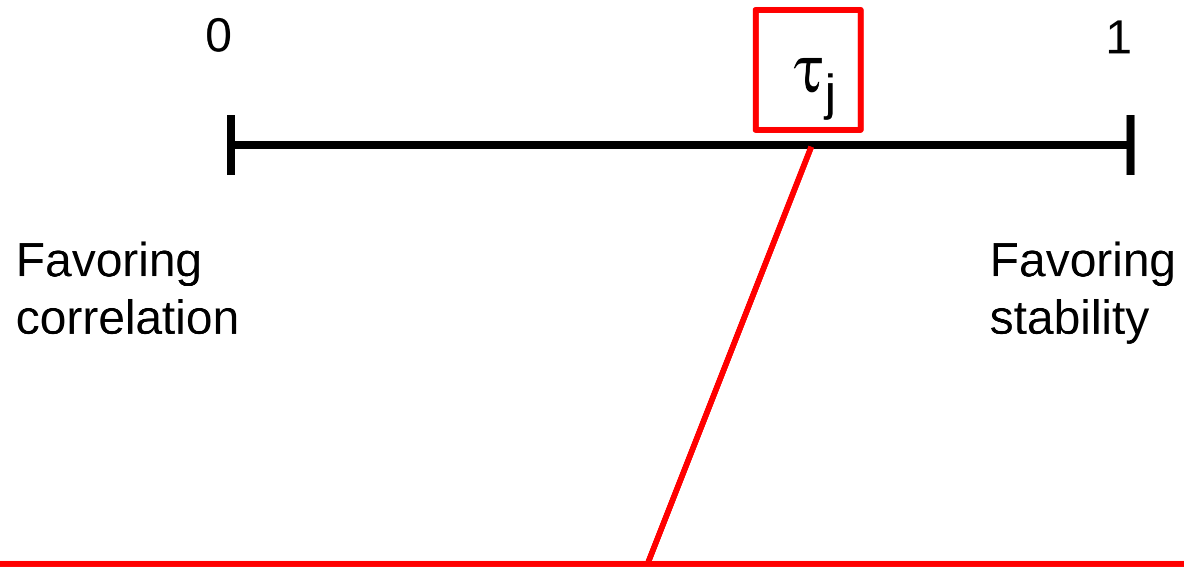
Special cases

Method	Criterion	Comments
PLS regression	$\underset{\|a_1\|=\|a_2\|=1}{\text{Maximize}}\ \text{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$	Is favoring too much stability with respect to correlation
Canonical Correlation Analysis	$\text{Maximize } \text{Cor}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$	Is favoring too much correlation with respect to stability

Choice of the shrinkage constant τ_j

$$\text{Maximize } \mathrm{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$$

$$\text{subject to } \tau_j \|\mathbf{a}_j\| + (1-\tau_j)\mathrm{Var}(\mathbf{X}_j\mathbf{a}_j) = 1$$



Schäfer and Strimmer (2005) give a formula for an optimal choice of τ_j.

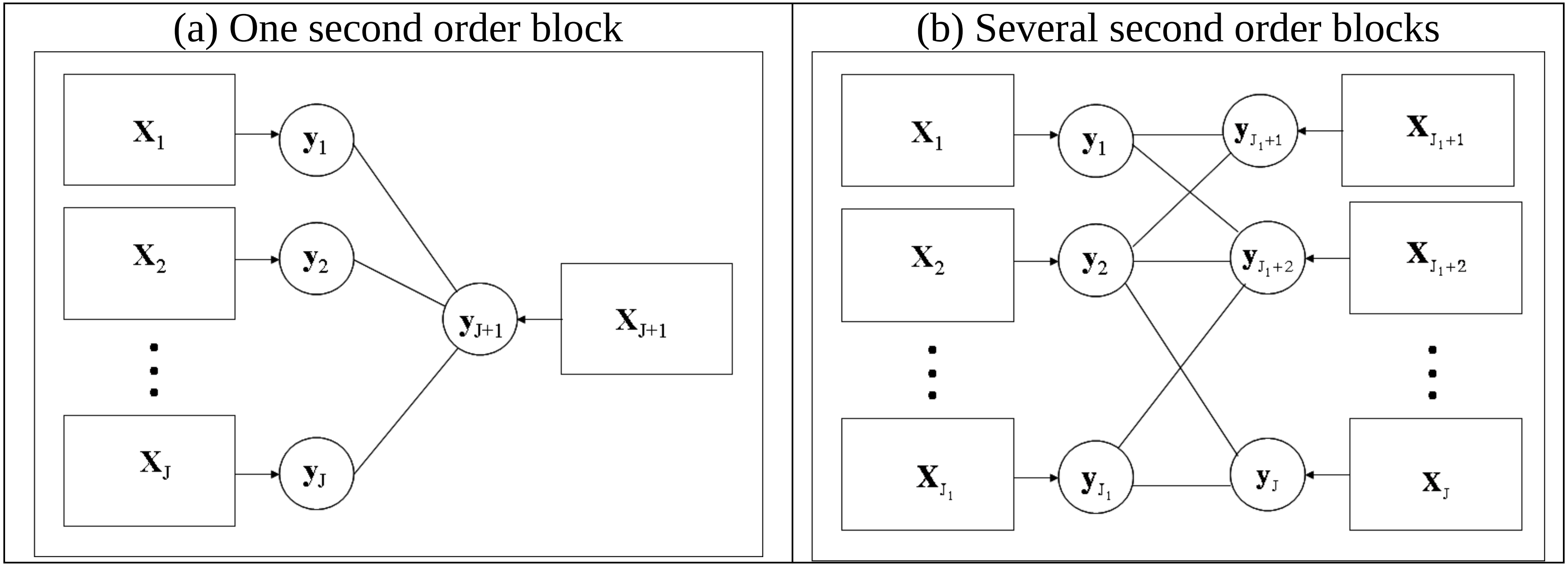
Special cases of Regularized generalized CCA

RGCCA and Multi-block data analysis

SUMCOR (Horst, 1961)	$\underset{Var(X_j a_j)=1}{Max} \sum_{j,k,j\neq k} \mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$
SSQCOR (Kettenring, 1971)	$\underset{Var(X_j a_j)=1}{Max} \sum_{j,k,j\neq k} \mathrm{Cor}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$
SABSCOR (Mathes, 1993, Hanafi, 2004)	$\underset{Var(X_j a_j)=1}{Max} \sum_{j,k,j\neq k} \left\|\mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)\right\|$
MAXDIFF (Van de Geer, 1984) [SUMCOV]	$\underset{\|a_j\|=1}{Max} \sum_{j,k,j\neq k} \mathrm{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$
MAXDIFF B (Hanafi & Kiers, 2006) [SSQCOV]	$\underset{\|a_j\|=1}{Max} \sum_{j,k,j\neq k} \mathrm{Cov}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$
SABSCOV (Krämer, 2007)	$\underset{\|a_j\|=1}{Max} \sum_{j,k,j\neq k} \left\|\mathrm{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)\right\|$

Special cases of Regularized generalized CCA

Hierarchical models



Very often:

$$X_1, \ldots, X_{J_1} = \text{Predictors}$$
$$X_{J_1+1}, \ldots, X_J = \text{Responses}$$

Special cases of Regularized generalized CCA

Hierarchical model : one 2nd order block

Method	Criterion	Constraints
Hierarchical PLS regression	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} g(\text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1}))$	$\lVert\mathbf{a}_j\rVert = 1,\ j = 1,\ldots,J+1$
Hierarchical Canonical Correlation Analysis	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} g(\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1}))$	$\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1,\ j = 1,\ldots,J+1$
Hierarchical Redundancy analysis of the $\mathbf{X}_j$'s with respect to $\mathbf{X}_{J+1}$	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} g(\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})\text{Var}(\mathbf{X}_j\mathbf{a}_j)^{1/2})$	$\lVert\mathbf{a}_j\rVert = 1,\ j = 1,\ldots,J$ $\text{Var}(\mathbf{X}_{J+1}\mathbf{a}_{J+1}) = 1$
Hierarchical Redundancy analysis of $\mathbf{X}_{J+1}$ with respect to the $\mathbf{X}_j$'s	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} g(\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})\text{Var}(\mathbf{X}_{J+1}\mathbf{a}_{J+1})^{1/2})$	$\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1,\ j = 1,\ldots,J$ $\lVert\mathbf{a}_{J+1}\rVert = 1$

Stable predictors and good prediction

Good predictors and stable response

g = identity, square or absolute value

Special cases of Regularized generalized CCA

Hierarchical model : one 2nd order block
Factorial scheme : g = square function

Concordance analysis (Hanafi & Lafosse, 2001)

$$\text{Maximize} \sum_{j=1}^{J} \text{Cov}^2(\mathbf{X}_j\mathbf{M}_j\mathbf{b}_j, \mathbf{X}_{J+1}\mathbf{M}_{J+1}\mathbf{b}_{J+1})$$

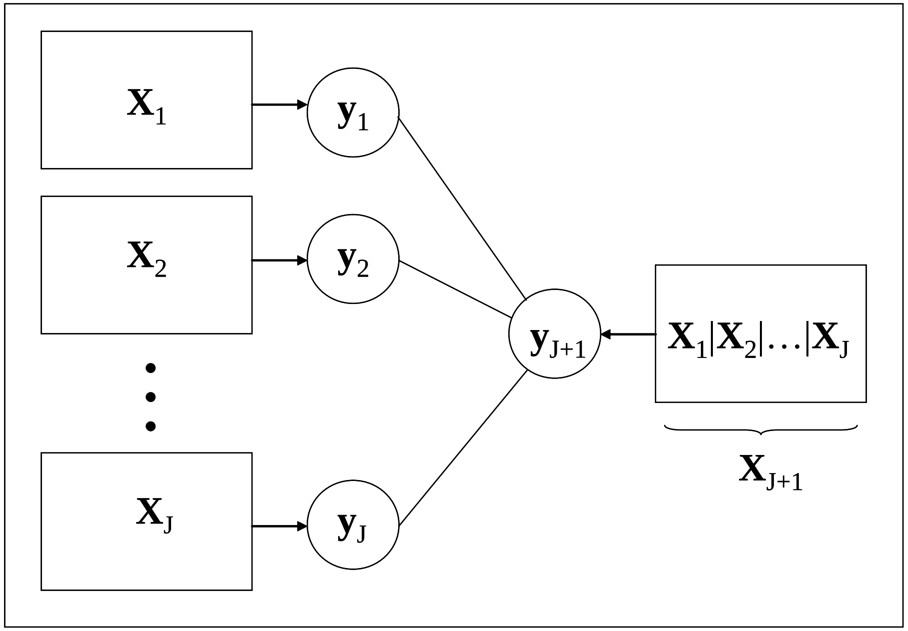
$$\text{subject to } \mathbf{b}_j^t\mathbf{M}_j\mathbf{b}_j = 1,\ j = 1,\ldots,J+1$$

The previous methods are found again for the metrics M_j equal to identity or Mahalanobis

Special cases of Regularized generalized CCA

Hierarchical model : one 2nd order block

$$\mathbf{X}_{J+1} = \left[\mathbf{X}_1, \ldots, \mathbf{X}_J\right]$$

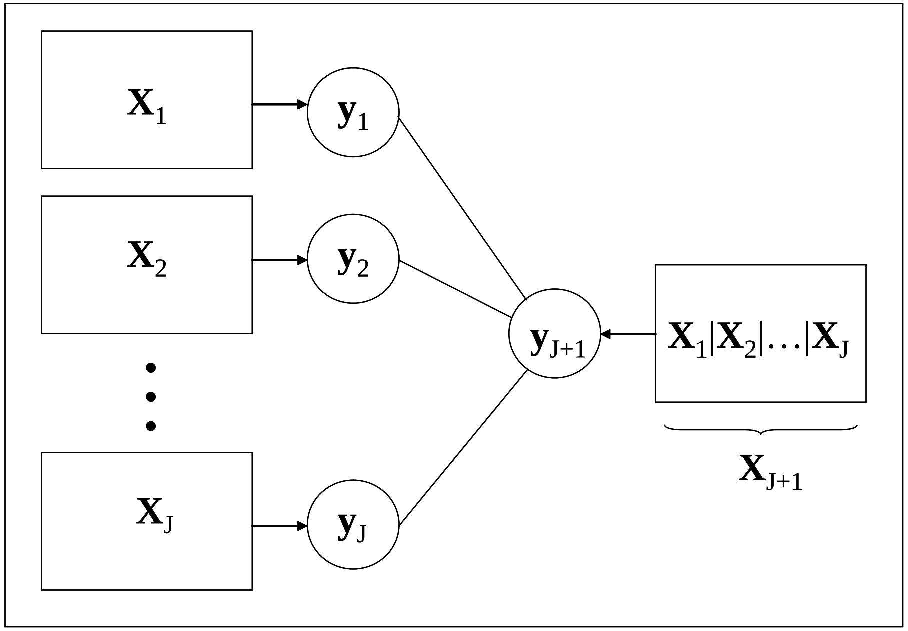


Method	Criterion	Constraints
SUMCOR (Horst, 1961)	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} \text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})$ or $\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} \left\|\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})\right\|$	$\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1, \ j = 1, \ldots, J+1$
Generalized CCA (Carroll, 1968a,b)	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J_1} \text{Cor}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1}) + \sum_{j=J_1+1}^{J} \text{Cov}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})$	$\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1, \ j = 1, \ldots, J_1, J+1$ $\left\|\mathbf{a}_j\right\| = 1, \ j = J_1+1, \ldots, J$

Special cases of Regularized generalized CCA

Hierarchical model : one 2nd order block

$$\mathbf{X}_{J+1} = \left[\mathbf{X}_1, \ldots, \mathbf{X}_J\right]$$



Multiple Co-inertia Analysis (Chessel & Hanafi, 1996)

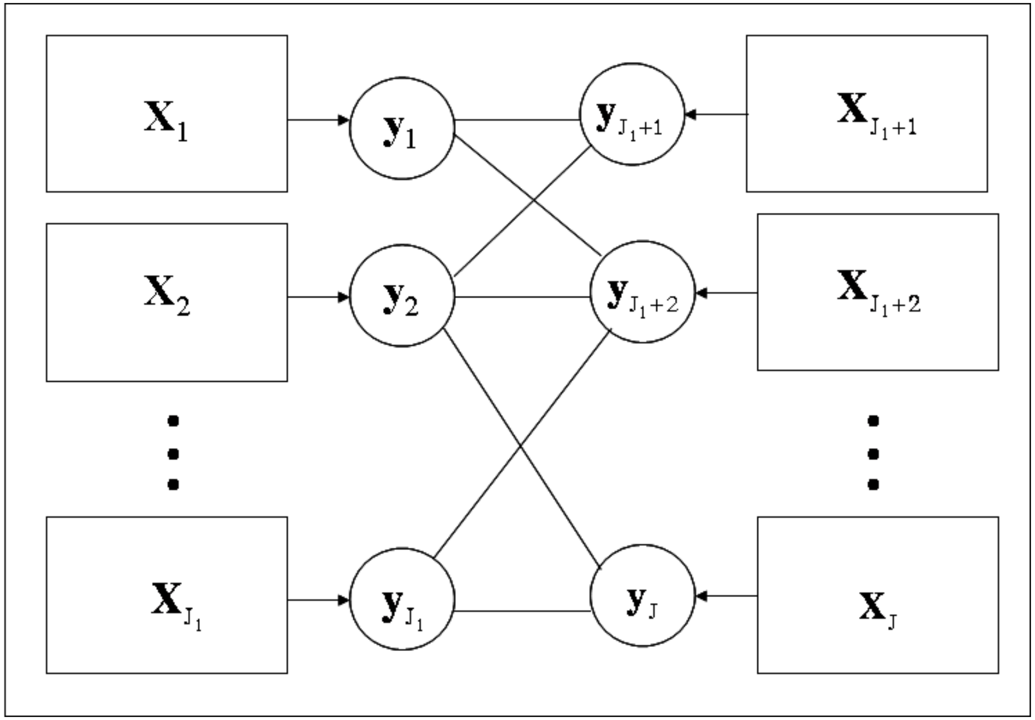
$$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} \text{Cov}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})$$

$$\text{subject to } \left\|\mathbf{a}_j\right\| = 1,\ j = 1,\ldots,J,\ Var(\mathbf{X}_{J+1}\mathbf{a}_{J+1}) = 1$$

Special case of Carroll's GCCA

Special cases of Regularized generalized CCA

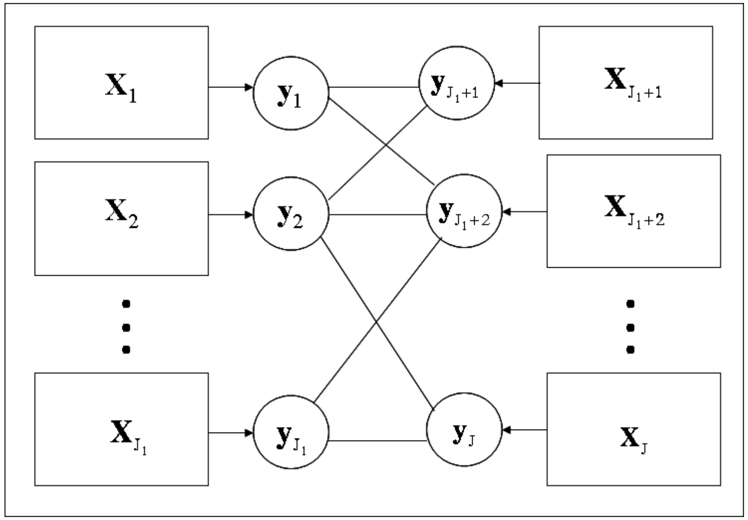
Hierarchical model : several 2nd order blocks



$c_{jk} = 1$ if response block X_k is connected to predictor block X_j,
$= 0$ otherwise

Special cases of Regularized generalized CCA

Hierarchical model : several 2nd order blocks



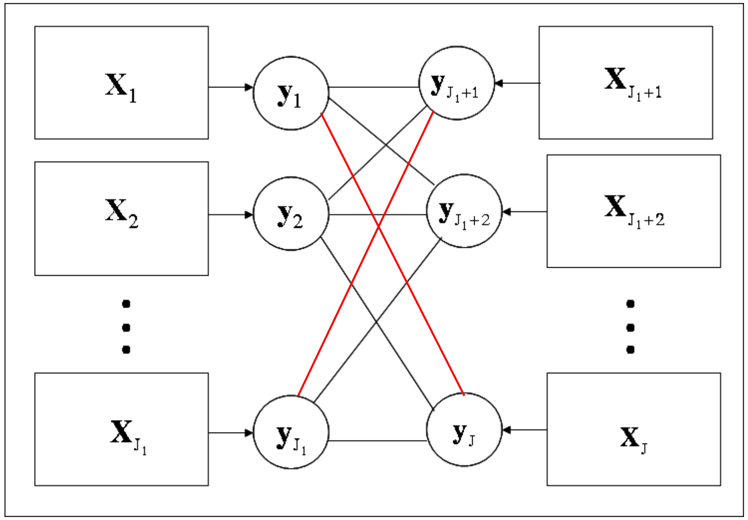
$$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_J}{\text{Maximize}} \quad \sum_{j=1}^{J_1} \sum_{k=J_1+1}^{J} c_{jk}\, g(\text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k))$$

$$\text{subject to the constraints: } \tau_j \left\|\mathbf{a}_j\right\|^2 + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1, \quad j = 1,\ldots, J$$

g = identity, square or absolute value

Special cases of Regularized generalized CCA

Generalized orthogonal multiple co-inertia analysis (Vivien & Sabatier, 2003)

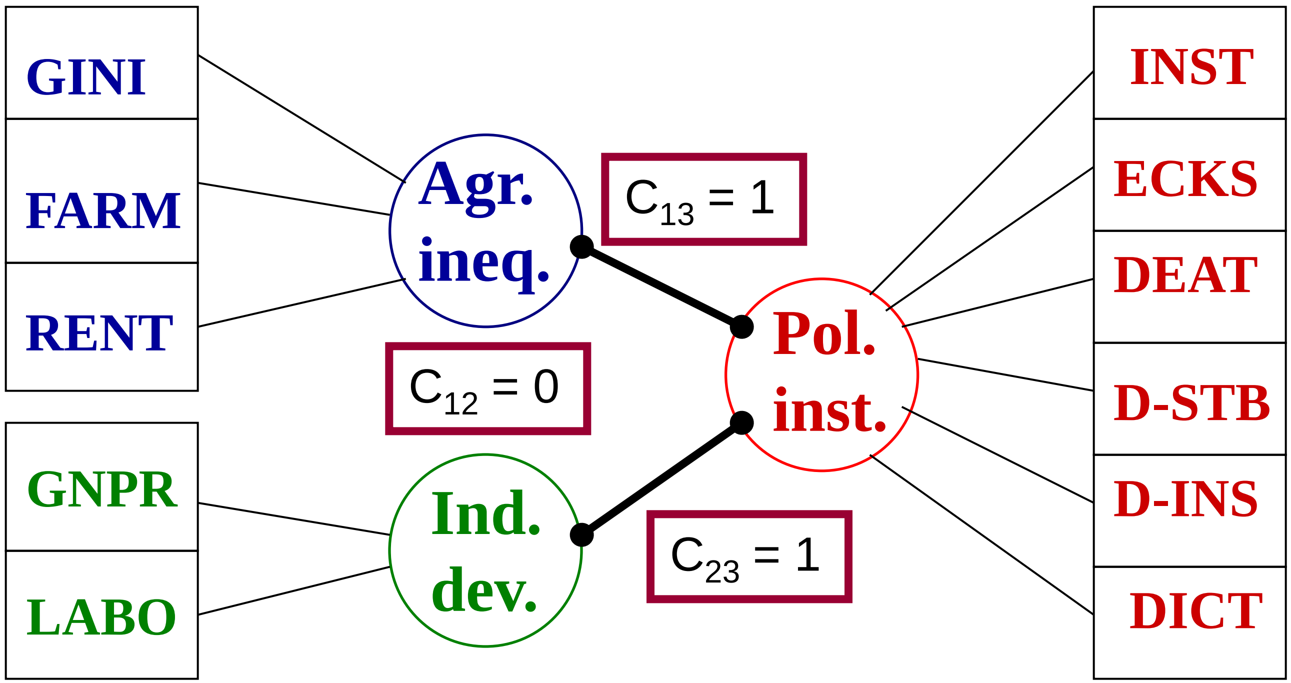


$$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_J}{\text{Maximize}} \sum_{j=1}^{J_1} \sum_{k=J_1+1}^{J} \text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$$

subject to the constraints: $\|\mathbf{a}_j\| = 1,\ \ j = 1,\ldots,J$

RGCCA applied to the Russett data

Agricultural inequality (X_1)



Industrial development (X_2)

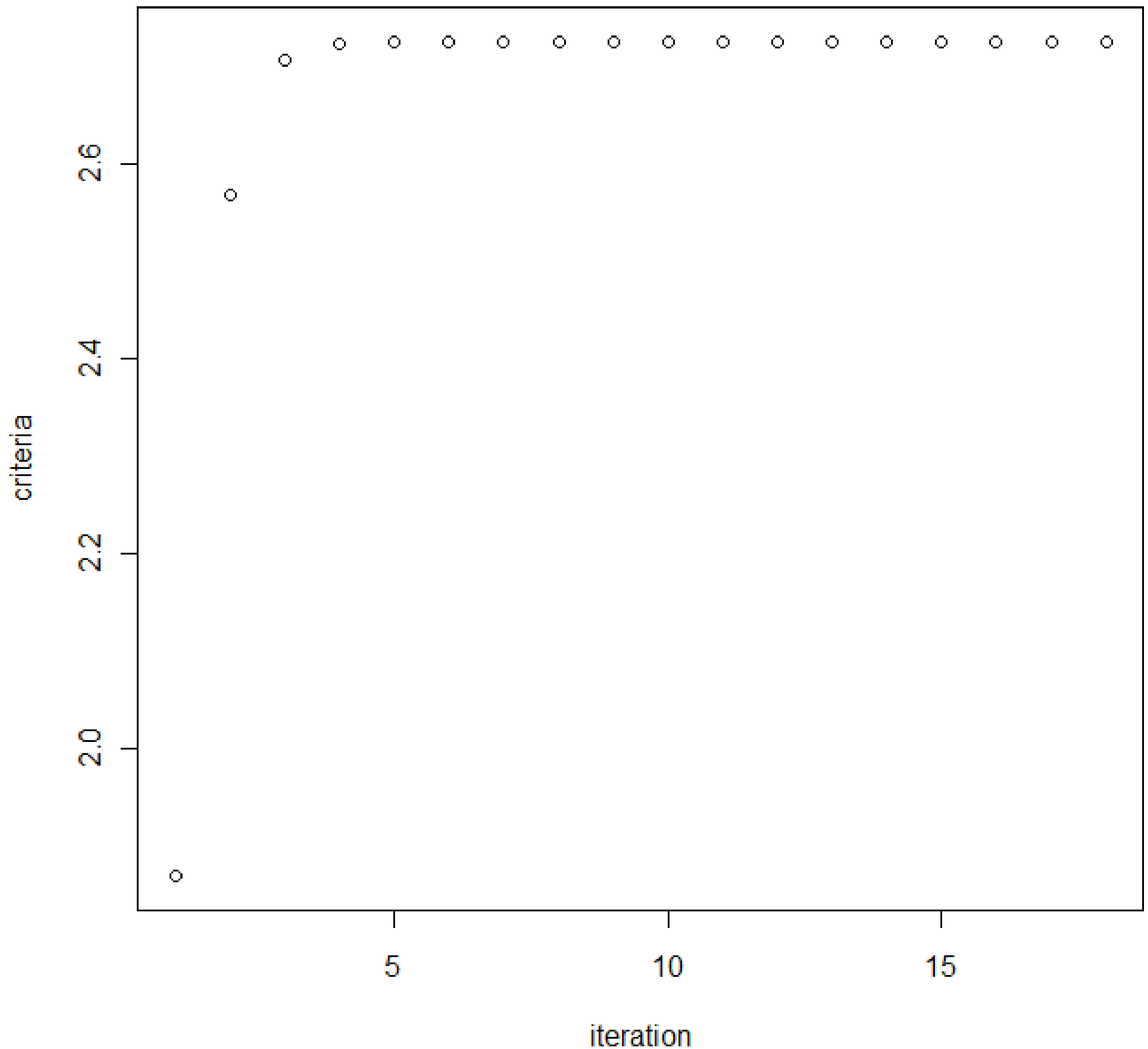
Political instability (X_3)

$$\underset{a_1, a_2, a_3}{\text{Maximize}} \quad \mathrm{g}(\mathrm{Cov}(X_1 a_1, X_3 a_3)) + \mathrm{g}(\mathrm{Cov}(X_2 a_2, X_3 a_3))$$

$$\text{subject to the constraints} \quad \tau_j \|a_j\|^2 + (1-\tau_j) Var(X_j a_j) = 1, \quad j = 1, 2, 3$$

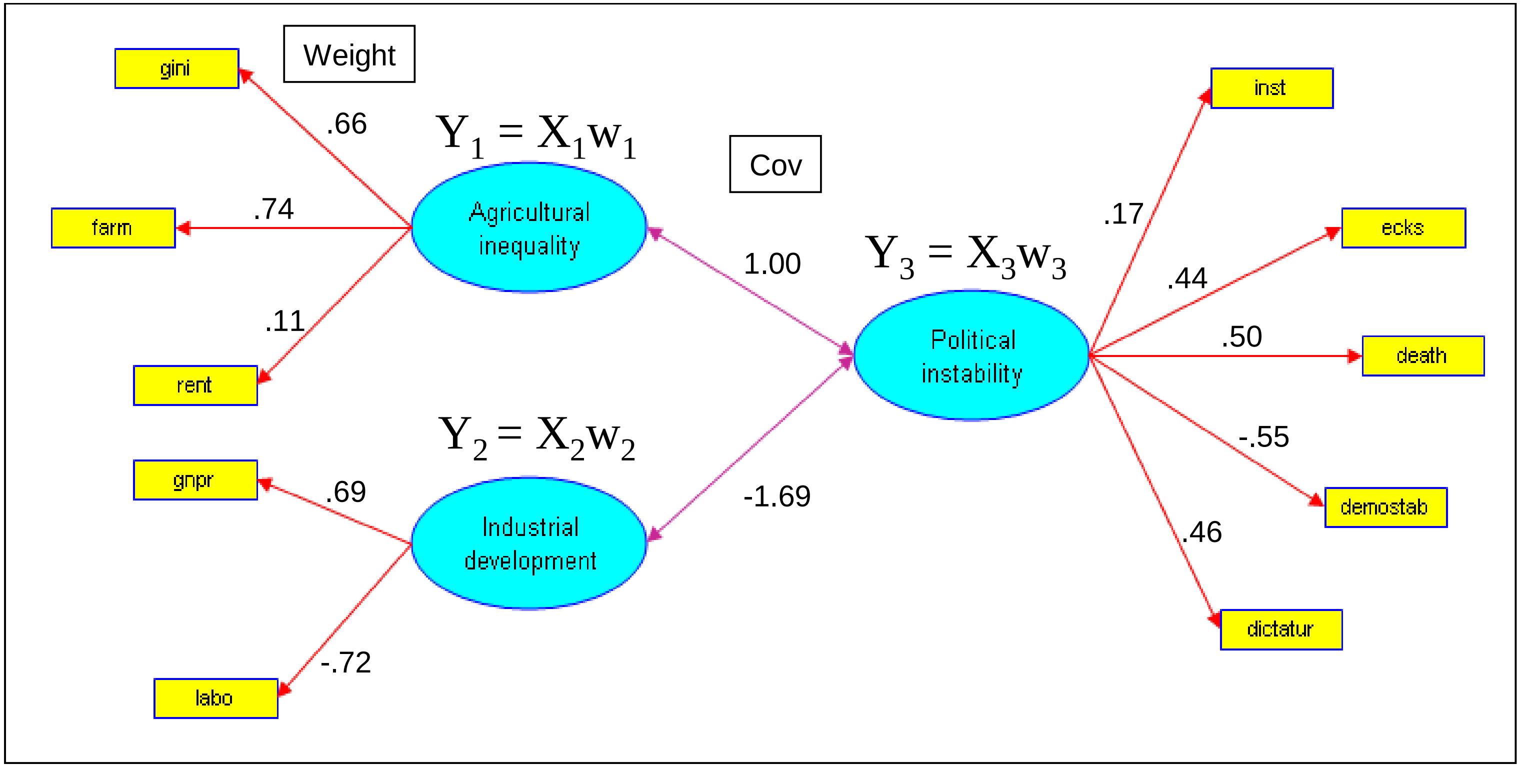
$0 \leq \tau_j \leq 1$, g = identity, square or absolute value

Monotone convergence of the algorithm



RGCCA: all $\tau_i = 1$ - centroid scheme

$$Max\left[\left|Cov(X_1w_1, X_3w_3)\right| + \left|Cov(X_2w_2, X_3w_3)\right|\right] = 2.69$$



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)

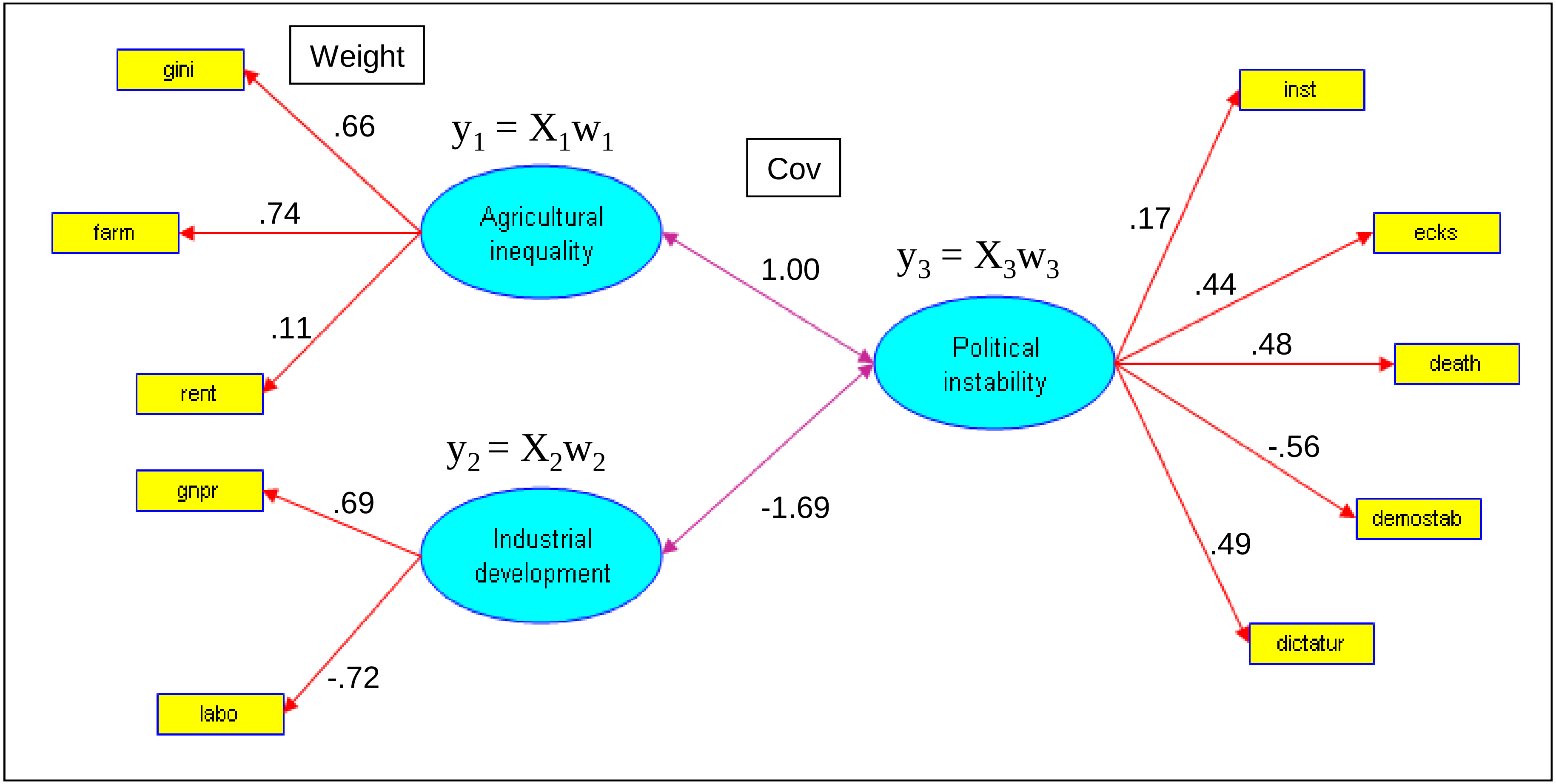
ECKS : Nb of violent internal war incidents (46-61)

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

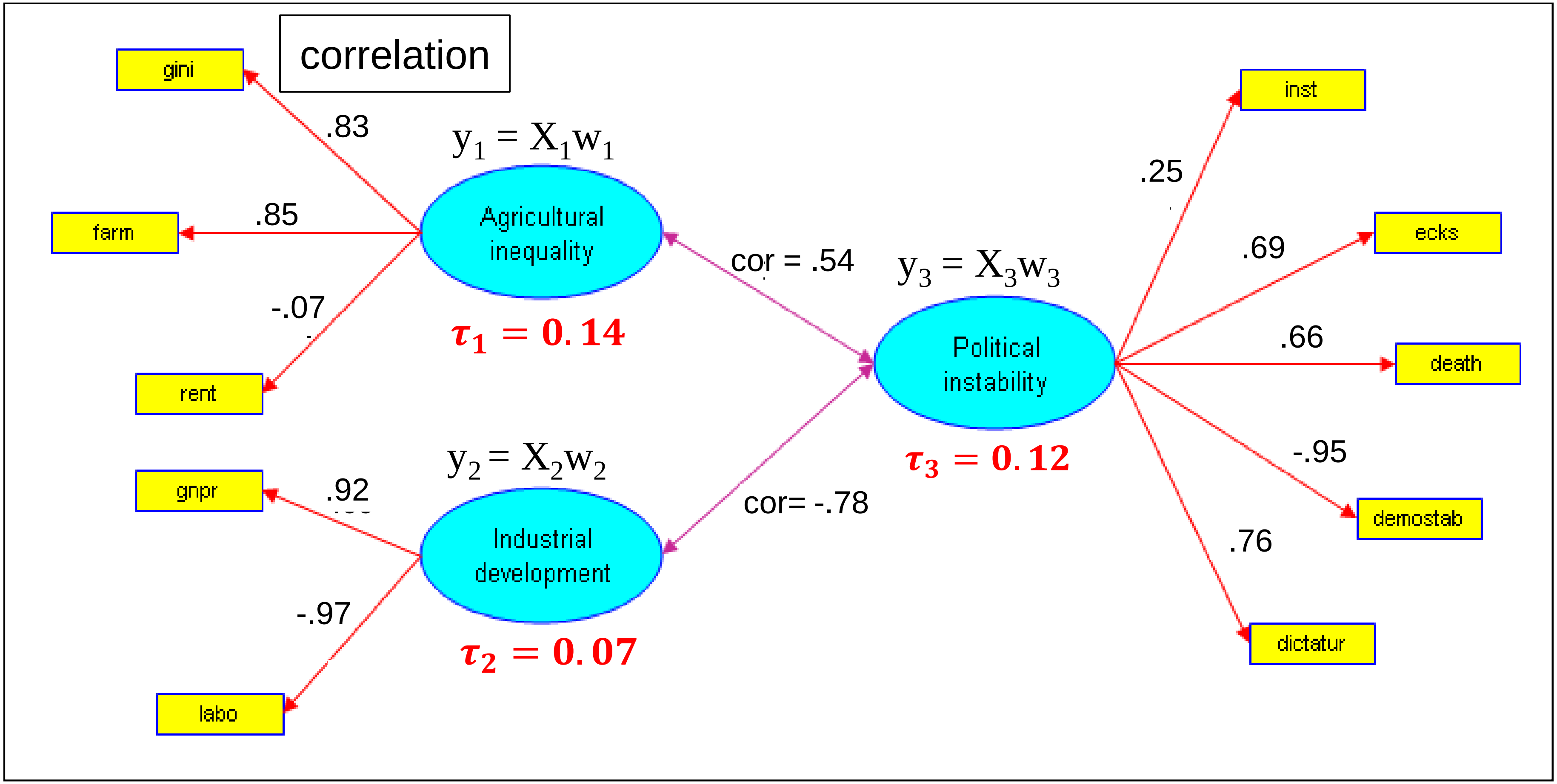
DEMO : Stable democracy, Unstable democracy or Dictatorship

RGCCA: all $\tau_i = 1$ - factorial scheme

$$Max\left[Cov^2(X_1w_1, X_3w_3) + Cov^2(X_2w_2, X_3w_3)\right] = 3.86$$

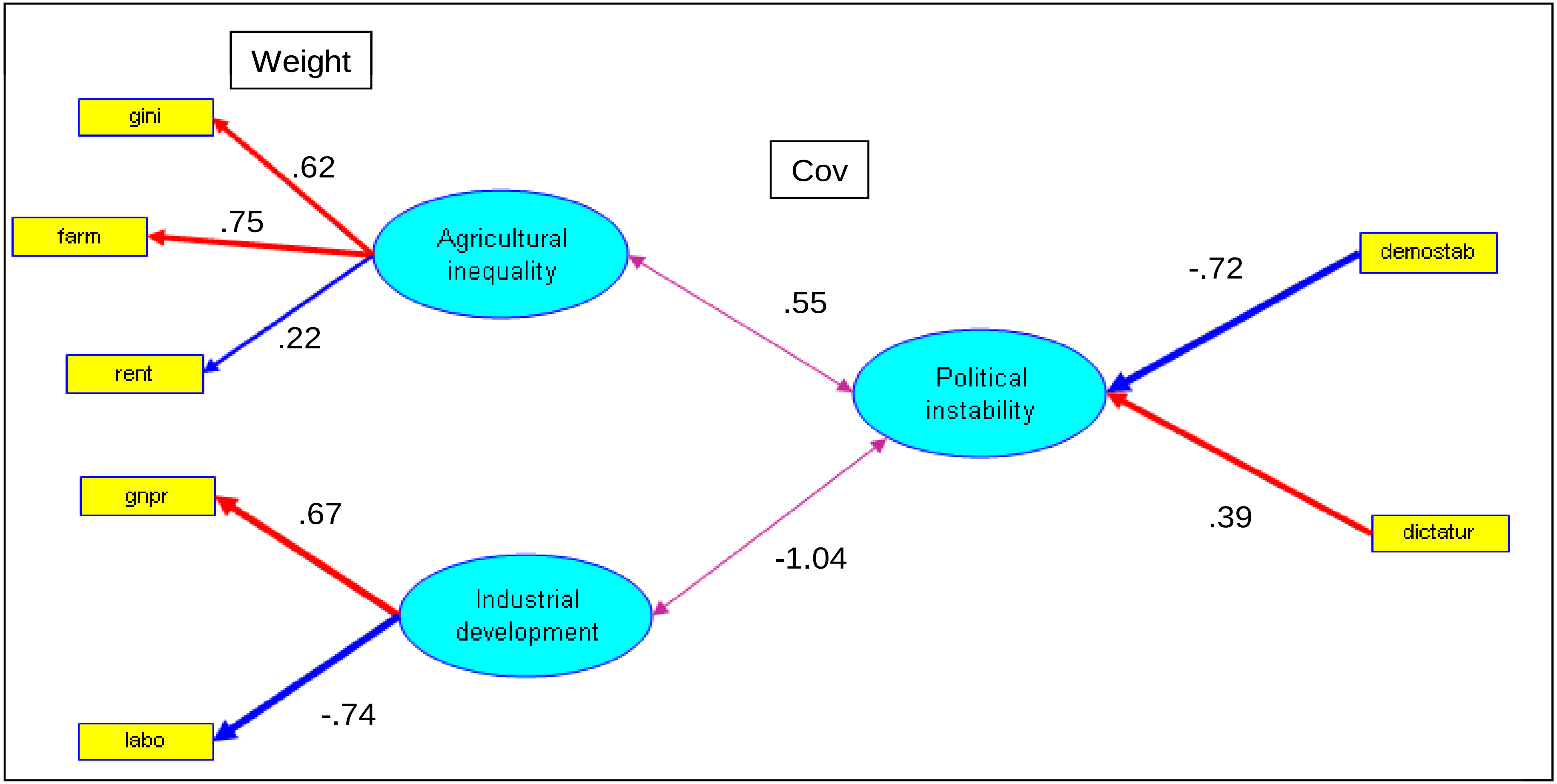


RGCCA: $\tau_i = \tau_i^*$ - centroid scheme



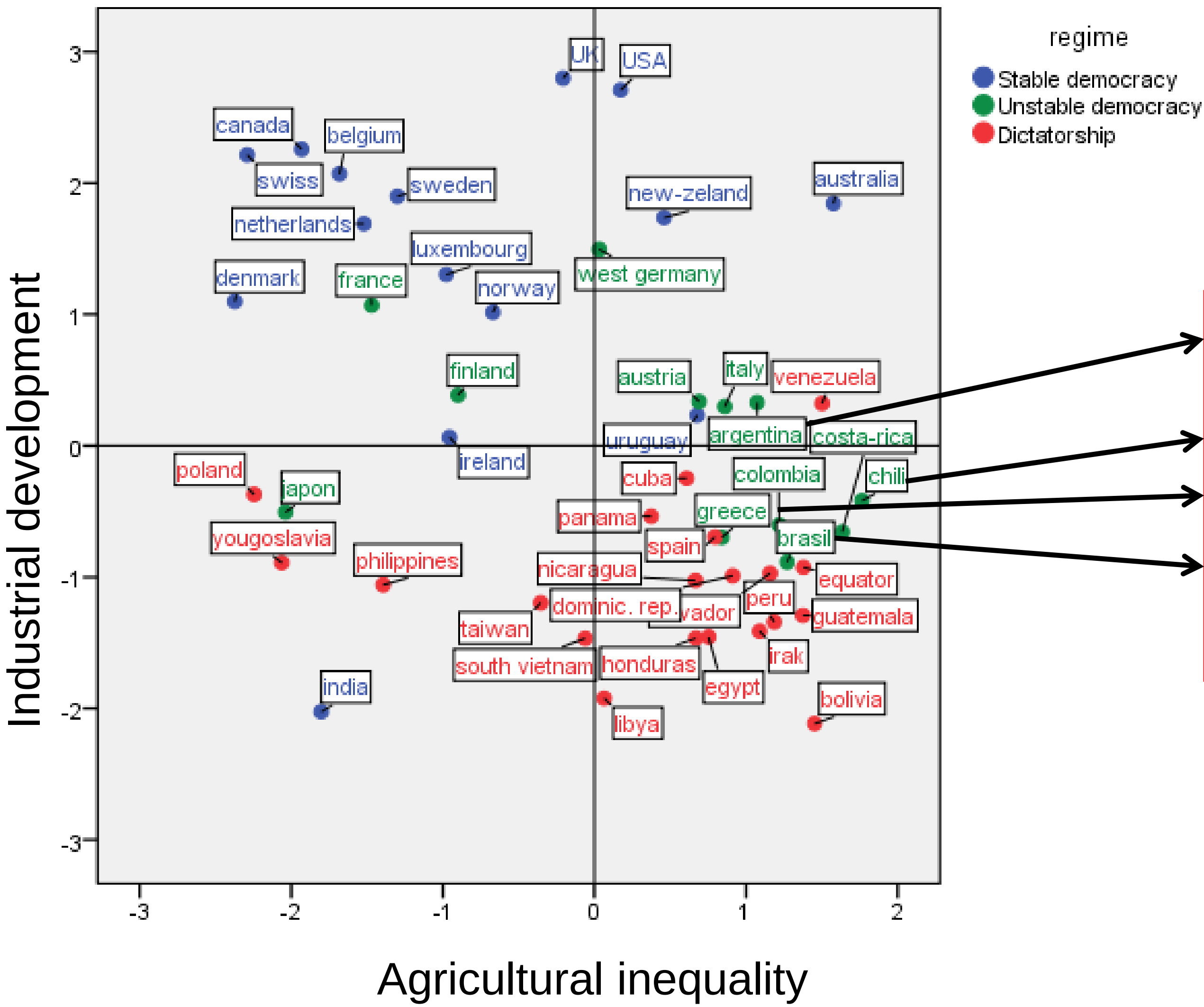
Generalized Barker & Rayens PLS-DA

$\tau_1 = \tau_2 = 1$ and $\tau_3 = 0$ - Factorial scheme



$$Max\left[Cor^2(X_1w_1, X_3w_3)*Var(X_1w_1)+Cor^2(X_2w_2, X_3w_3)*Var(X_2w_2)\right]=1.39$$

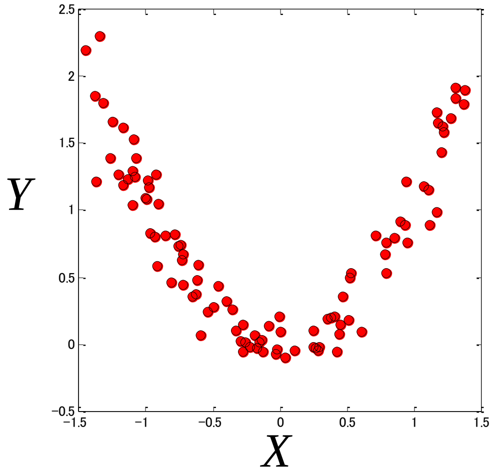
Generalized Barker & Rayens PLS-DA



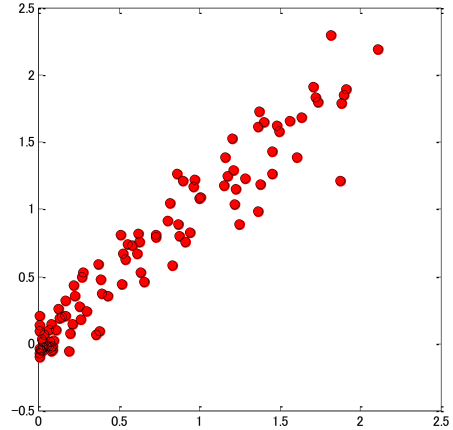
These countries have known a period of dictatorship after 1964.

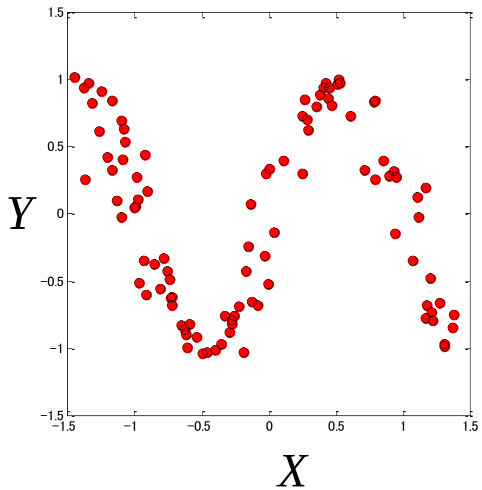
nonlinear dependence

cor(X, Y) = 0.17



cor(X^2, Y) = 0.96





cor(X, Y) = -0.06

cor(X^2, Y) = 0.09

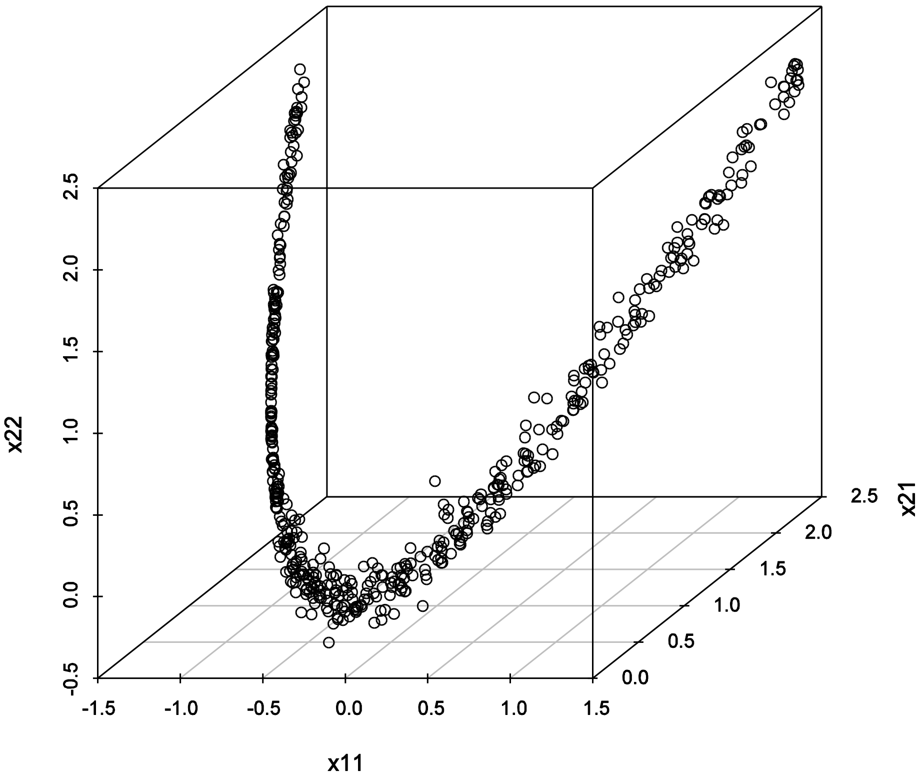
cor(X^3, Y) = -0.38

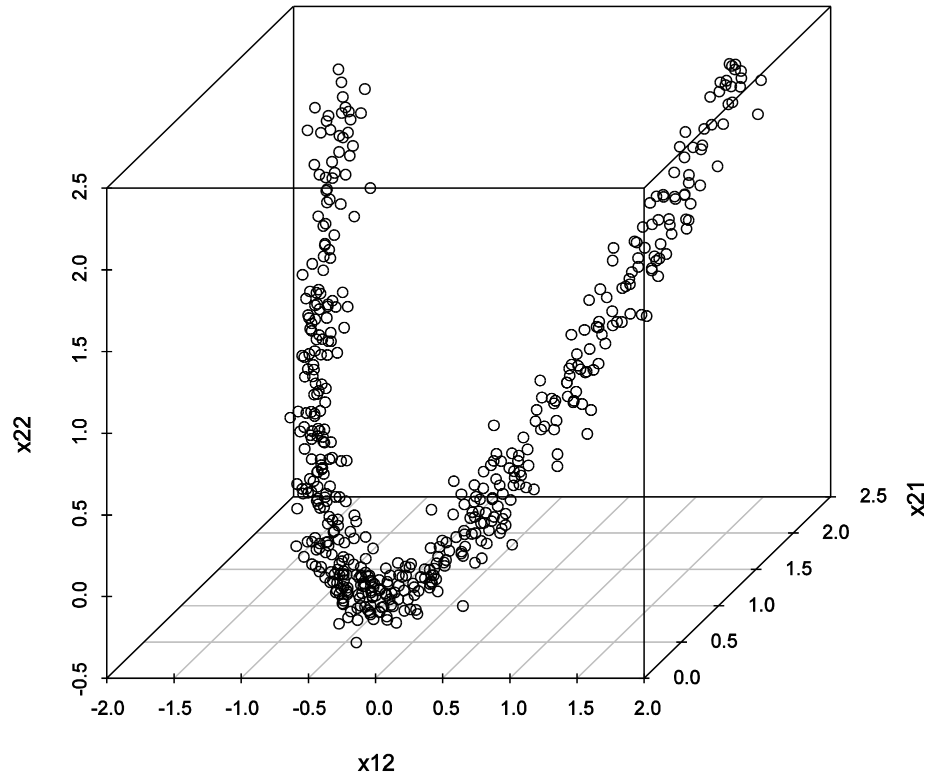
cor($\sin(\pi X)$, Y) = 0.93

RGCCA in action : toy example

Simulated data :

- Number of observations : 500
- Number of blocks : 2 (2 variables per block)



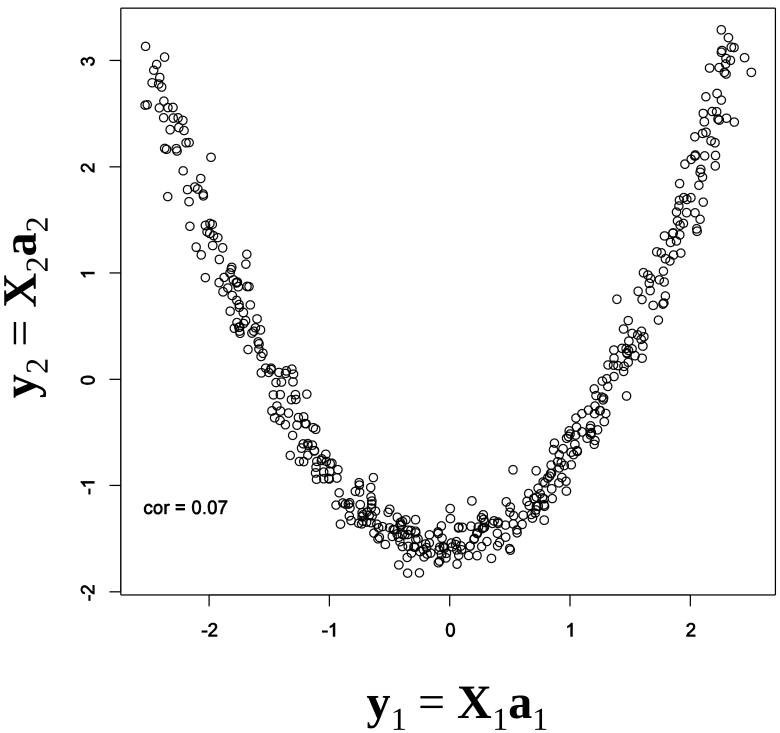


X_{11} with block $\mathbf{X}_2 = [X_{21}, X_{22}]$

X_{12} with block $\mathbf{X}_2 = [X_{21}, X_{22}]$

RGCCA results

RGCCA in action with $\tau_1 = \tau_2 = 1$ and factorial scheme



RGCCA fails to captures nonlinear dependences between blocks

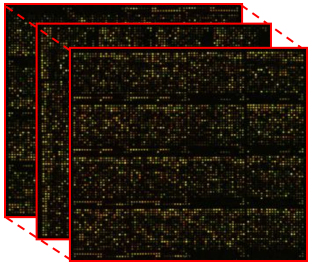
First motivation of Kernel GCCA :

recover nonlinear relationships between blocks

Glioma Cancer Data

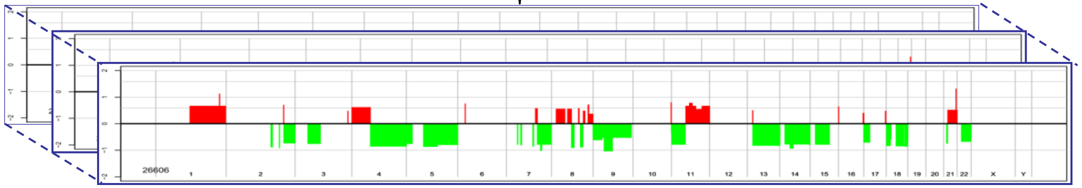
(From the Department of Pediatric Oncology of the Gustave Roussy Institute, 2009)

Transcriptomic data (X_1)



outcome (X_3)

	Gene 1	Gene 2	…	Gene 27982	CGH1	…	CGH 3268	Outcome
Patient 1	0.18	-0.21		-0.73	0.00		-0.55	X
Patient 2	1.15	-0.45		0.27	-0.30		0.00	X
Patient 3	1.35	0.17		0.22	0.33		0.64	Y
⋮								
⋮								
Patient 36	1.39	0.18		-0.17	0.00		0.43	Z



CGH data (X_2)

RGCCA in action in a very high dimensional blocks settings

- Requires to invert J matrices

$$\mathbf{M}_j = (1-\tau_j)\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I}_{p_j} \qquad j = 1,\dots,J$$

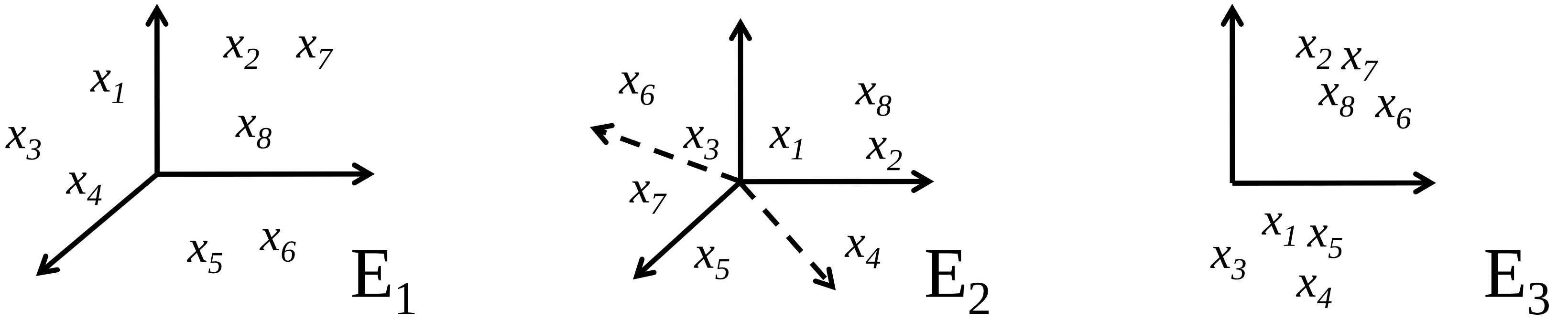
where $\mathbf{M}_j \in \mathbb{R}^{p_j \times p_j}$ $\qquad j = 1,\dots,J$

Here $p_1 = 27982$, $p_2 = 3268$ and $p_3 = 2$

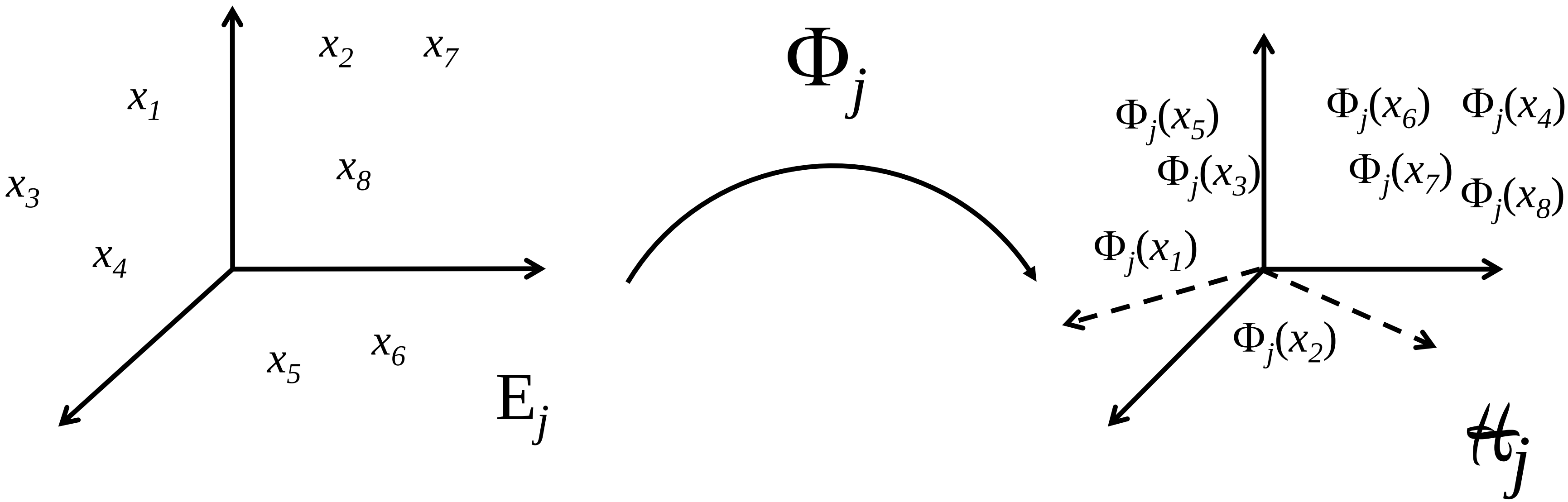
Second motivation of Kernel GCCA :
Capable of dealing with very high dimensional blocks settings

what we have …

- A set of *n* observations characterized by different point of view



Kernel GCCA: an idea



Apply RGCCA on the feature spaces $\mathcal{H}_1, \ldots, \mathcal{H}_J$

- How to choose $\Phi_1, \ldots \Phi_J$ $\rightarrow$ kernel method

$$\Phi_j(x) = k_j(., x) \qquad j = 1, \ldots, J$$

where k_j: $E_j \times E_j$ is a positive definite kernel function and $\mathcal{H}_j$ associated reproducing kernel Hilbert space.

Kernel GCCA (two-block case)

For any two directions $f_1 \in \mathcal{H}_1$ and $f_2 \in \mathcal{H}_2$, define the projections y_{1i} of $\Phi_1\left(x_1^{(i)}\right)$ on f_1 and y_{2i} of $\Phi_2\left(x_2^{(i)}\right)$ on f_2 by:

$$y_{1i} = \langle f_1, \Phi_1\left(x_1^{(i)}\right)\rangle \quad \text{and} \quad y_{2i} = \langle f_2, \Phi_2(x_2^{(i)})\rangle$$

The goal of KGCCA (two-block case) is to find $f_1 \in \mathcal{H}_1$ and $f_2 \in \mathcal{H}_2$ that maximize the following optimization problem:

$$\underset{f_1 \in H_1, f_2 \in H_2}{\text{argmax}} \ \text{cov}(\mathbf{y}_1, \mathbf{y}_2)$$

$$\text{s.c.} \ (1-\tau_j)\text{var}\left(\mathbf{y}_j\right) + \tau_j \left\|f_j\right\|_{\mathcal{H}_j}^2 = 1, \quad j = 1,2$$

Kernel GCCA (two-block case)

It is always possible to express f_j as follows: $f_j = \sum_{k=1}^{n} \alpha_j^{(k)} \Phi_j\left(x_j^{(k)}\right)$

We deduce that y_{ji} can be re-expressed as follows:

$$y_{ji} = \langle f_j, \Phi_j\left(x_j^{(i)}\right)\rangle = \sum_{k=1}^{n} \alpha_j^{(k)} \langle \Phi_j\left(x_j^{(k)}\right), \Phi_j\left(x_j^{(i)}\right)\rangle$$

Let us note $\mathbf{K}_\mathrm{j}$ the $n \times n$ kernel matrix defined by:

$$(\mathbf{K}_j)_{ki} = k\left(x_j^{(k)}, x_j^{(i)}\right) = \langle \Phi_j\left(x_j^{(k)}\right), \Phi_j\left(x_j^{(i)}\right)\rangle$$

We deduce that $\mathbf{y}_j$ is defined by: $\boxed{\mathbf{y}_j = \mathbf{K}_j \boldsymbol{\alpha}_j}$

Kernel GCCA (two-block case)

Then the previous optimization problem:

$$\underset{f_1 \in H_1, f_2 \in H_2}{\operatorname{argmax}} \ \operatorname{cov}(\mathbf{y}_1, \mathbf{y}_2)$$

$$\text{s.c. } (1-\tau_j)\operatorname{var}\left(\mathbf{y}_j\right) + \tau_j \left\| f_j \right\|_{\mathcal{H}_j}^2 = 1, \quad j = 1,2$$

can be write down only in terms of kernel matrices as follows:

$$\underset{\alpha_1, \alpha_2}{\operatorname{argmax}} \ \frac{1}{n} \boldsymbol{\alpha}_1^t \mathbf{K}_1 \mathbf{K}_2 \boldsymbol{\alpha}_2$$

$$\text{s.c. } (1-\tau_j)\frac{1}{n} \boldsymbol{\alpha}_j^t \mathbf{K}_j^2 \boldsymbol{\alpha}_\mathbf{j} + \tau_j \boldsymbol{\alpha}_j^t \mathbf{K}_j \boldsymbol{\alpha}_j = 1 \quad j = 1, 2$$

Kernel KGCCA (J-block case)

$$\underset{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_J}{\operatorname{argmax}} \sum_{j\neq k}^{J} c_{jk}\ \mathrm{g}\left(\frac{1}{n}\boldsymbol{\alpha}_j^t \mathbf{K}_j \mathbf{K}_k \boldsymbol{\alpha}_k\right)$$

$$\text{s.c.} \quad (1-\tau_j)\frac{\mathbf{1}}{\mathbf{n}}\boldsymbol{\alpha}_j^t \mathbf{K}_j^2 \boldsymbol{\alpha}_{\mathbf{j}} + \tau_j \boldsymbol{\alpha}_j^t \mathbf{K}_j \boldsymbol{\alpha}_j = 1 \quad j = 1,\ldots,J$$

Requires to invert J matrices

$$\mathbf{N}_j = (1-\tau_j)\,\frac{1}{n}\mathbf{K}_j^2 + \tau_j \mathbf{K}_j \qquad j = 1,\ldots,J$$

where $\mathbf{N}_j \in \mathbb{R}^{n\times n}$ $\qquad j = 1,\ldots,J$

Problem : $\mathbf{K}_j$ (and thus $\mathbf{N}_j$) is not necessary of full rank
(for instance when a centered Gram matrix is considered)

(incomplete) Cholesky Decomposition

- Find a lower triangular matrix $\mathbf{R}_j$ such that $\mathbf{K}_j = \mathbf{R}_j^t \mathbf{R}_j$

  where $\mathbf{R}_j \in \mathbb{R}^{\text{rank}(\mathbf{K}_j) \times n}$

$$\underset{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_J}{\text{argmax}} \sum_{j \neq k}^{J} c_{jk}\, \mathrm{g}\left(\frac{1}{n} \boldsymbol{\alpha}_j^t \mathbf{R}_j^t \mathbf{R}_j \mathbf{R}_k^t \mathbf{R}_k \boldsymbol{\alpha}_k\right)$$

$$\text{s.c.} \quad (1 - \tau_j) \frac{1}{n} \boldsymbol{\alpha}_j^t \mathbf{R}_j^t \mathbf{R}_j \mathbf{R}_j^t \mathbf{R}_j \boldsymbol{\alpha}_j + \tau_j \boldsymbol{\alpha}_j^t \mathbf{R}_j^t \mathbf{R}_j \boldsymbol{\alpha}_j = 1, \quad j = 1, \ldots, J$$

The KGCCA optimization problem

$$\underset{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_J}{\text{argmax}} \sum_{j,k}^{J} c_{jk}\ \mathrm{g}\left(\frac{1}{n}\mathbf{w}_j^t \mathbf{R}_j \mathbf{R}_k^t \mathbf{w}_k\right) \qquad (\mathbf{w}_j = \mathbf{R}_j \boldsymbol{\alpha}_j)$$

$$\text{s.c.} \qquad (1-\tau_j)\frac{1}{n}\mathbf{w}_j^t \mathbf{R}_j \mathbf{R}_j^t \mathbf{w}_j + \tau_j \mathbf{w}_j^t \mathbf{w}_j = 1, \quad j = 1, \ldots, J$$

Apply the initial algorithm of RGCCA on $\mathbf{R}_1^t, \ldots, \mathbf{R}_J^t$ to obtain the Latent variables outer estimation.

In the glioma Cancer Data : $\mathbf{X}_1 \in \mathbb{R}^{36\times 27982} \rightarrow \mathbf{R}_1^t \in \mathbb{R}^{36\times 35}$

$\mathbf{X}_2 \in \mathbb{R}^{36\times 3268} \rightarrow \mathbf{R}_2^t \in \mathbb{R}^{36\times 34}$

Construction of monotone convergent algorithms for these criteria

The Kernel GCCA algorithm

Initial step → $\mathbf{w}_j$

$$\mathrm{y}_{\boldsymbol{j}} = \mathrm{R}_{\boldsymbol{j}}^{\boldsymbol{t}} \mathrm{w}_{\boldsymbol{j}}$$

Outer Estimation
(explains the block)

$$\mathbf{w}_j^t \left[(1-\tau_j)\frac{1}{n}\mathbf{R}_j\mathbf{R}_j^t + \tau_j \mathbf{I}_{rank(\mathbf{K}_j)} \right] \mathbf{w}_j = 1$$

$$\mathbf{z}_j = \sum_{k \neq j} e_{jk} \mathbf{y}_k$$

Inner
Estimation
(explains
relation
between
block)

Iterate until
convergence
of the criterion

$$\mathrm{w}_{\boldsymbol{j}} = \frac{\left((1-\tau_{\boldsymbol{j}})\frac{1}{\boldsymbol{n}}\mathrm{R}_{\boldsymbol{j}}\mathrm{R}_{\boldsymbol{j}}^{\boldsymbol{t}} + \tau_{\boldsymbol{j}} \mathrm{I}_{\mathrm{rank}(\mathrm{K}_{\boldsymbol{j}})} \right)^{-1} \mathrm{R}_{\boldsymbol{j}} \mathrm{z}_{\boldsymbol{j}}}{\sqrt{\mathrm{z}_{\boldsymbol{j}}^{\boldsymbol{t}} \mathrm{R}_{\boldsymbol{j}}^{\boldsymbol{t}} \left((1-\tau_{\boldsymbol{j}})\frac{1}{\boldsymbol{n}}\mathrm{R}_{\boldsymbol{j}}\mathrm{R}_{\boldsymbol{j}}^{\boldsymbol{t}} + \tau_{\boldsymbol{j}} \mathrm{I}_{\mathrm{rank}(\mathrm{K}_{\boldsymbol{j}})} \right)^{-1} \mathrm{R}_{\boldsymbol{j}} \mathrm{z}_{\boldsymbol{j}}}}$$

Choice of weights e_{jh}:

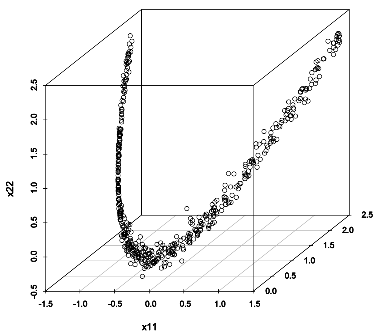
- Horst : $e_{jk} = c_{jk}$
- Centroid : $e_{jk} = c_{jk}\,\mathrm{sign}\left(\mathrm{cor}\left(\mathbf{y}_j, \mathbf{y}_k\right)\right)$
- Factorial : $e_{jk} = c_{jk}\,\mathrm{cov}\left(\mathbf{y}_j, \mathbf{y}_k\right)$

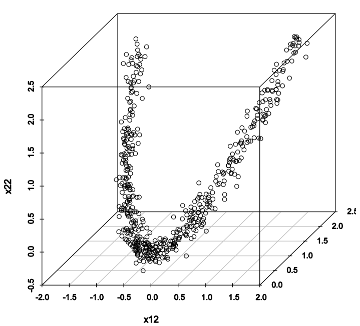
c_{jk} = 1 if blocks are linked, 0 otherwise and c_{jj} = 0

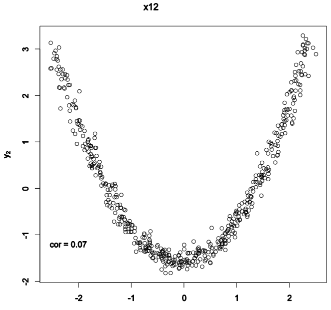
KGCCA in action : toy example

Simulated data :

- Number of observations : 500
- Number of blocks : 2 (2 variables per block)



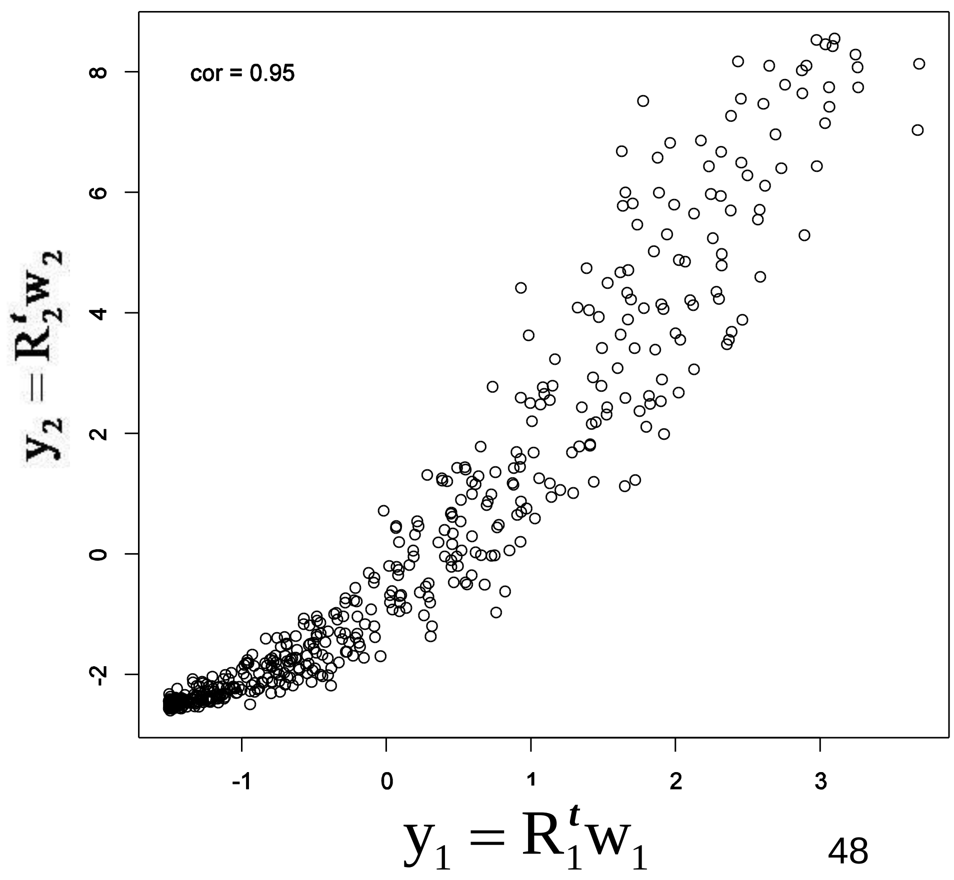




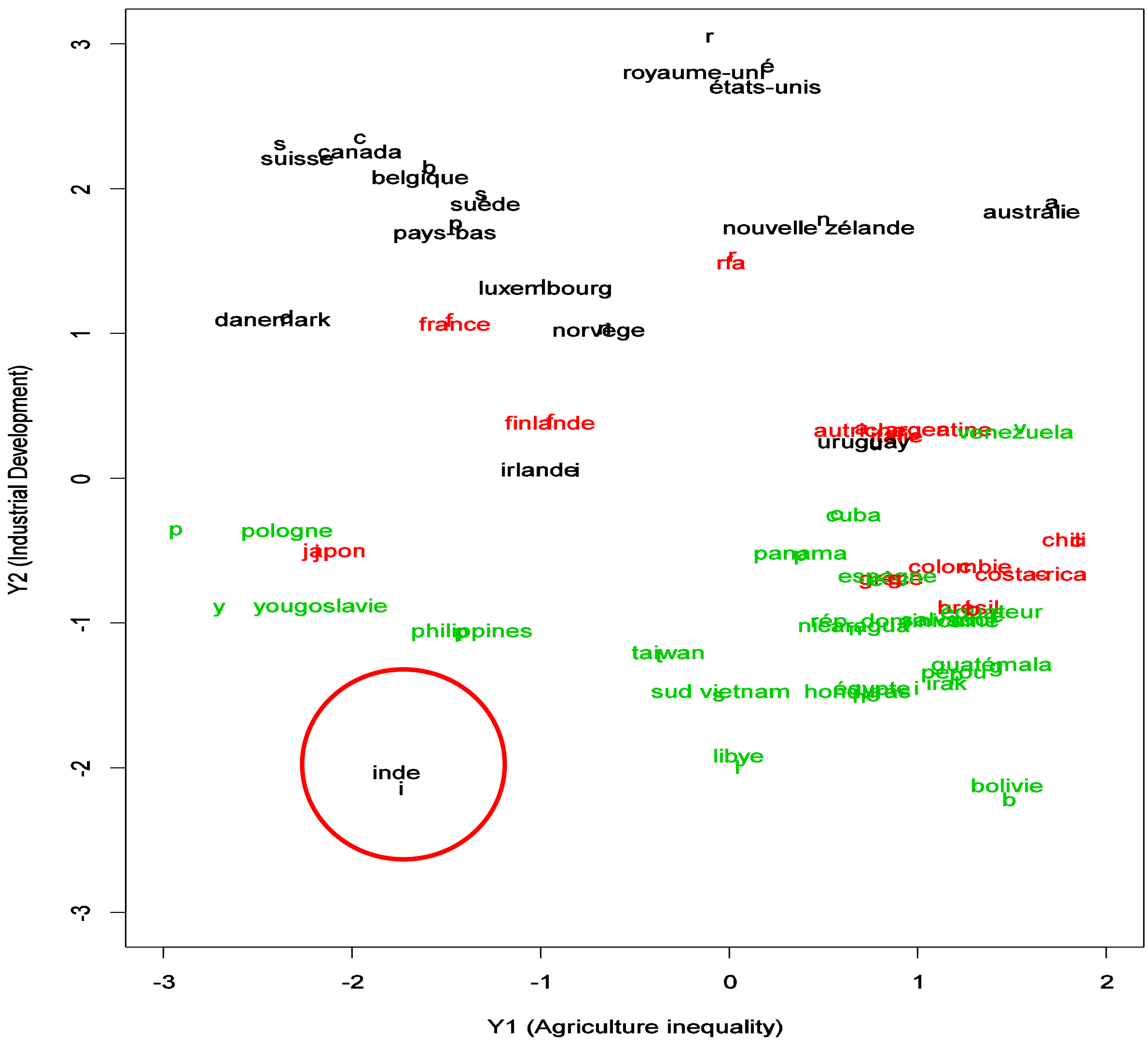
$\tau_1 = \tau_2 = 1$

factorial scheme

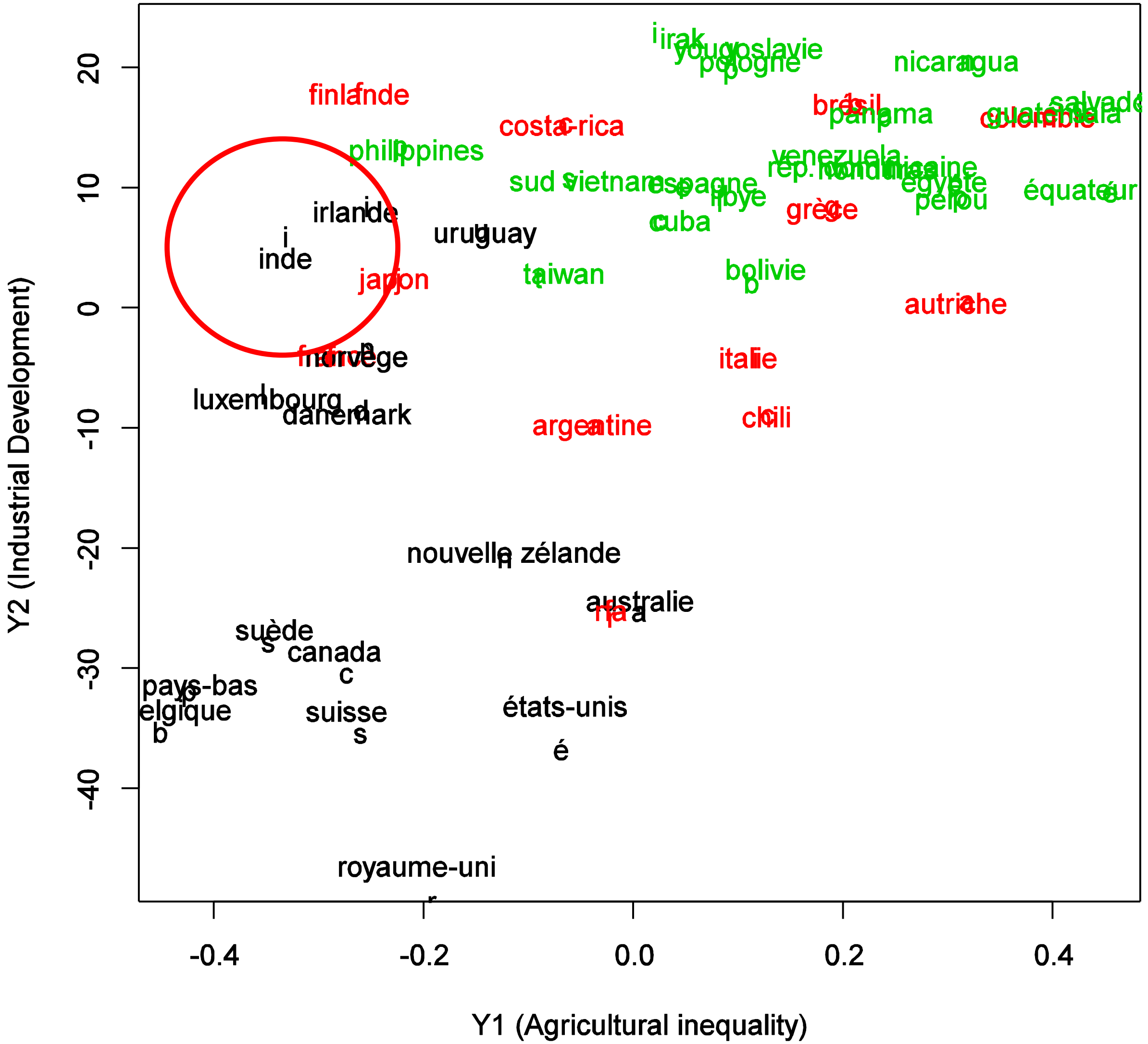
polynomial
Kernel of degree 2



RGCCA results (factorial scheme, $\tau_j = 1$)



KGCCA results (factorial scheme, τ_j =1, spline kernel)

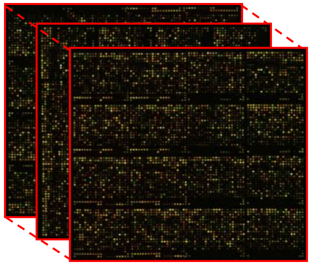


$$k(x,y) = 1 + xy + xy\min(x,y) - \frac{x+y}{2}\min(x,y)^2 + \frac{1}{3}\min(x,y)^3$$

Glioma Cancer Data

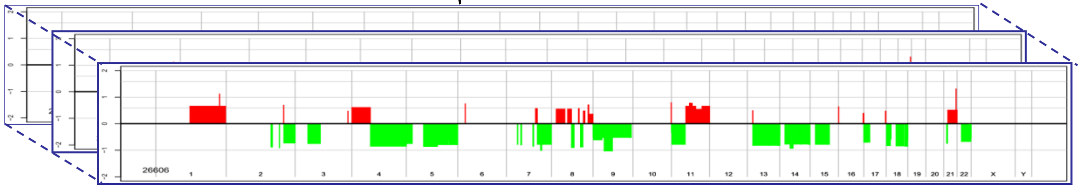
(From the Department of Pediatric Oncology of the Gustave Roussy Institute, 2009)

Transcriptomic data (X_1)



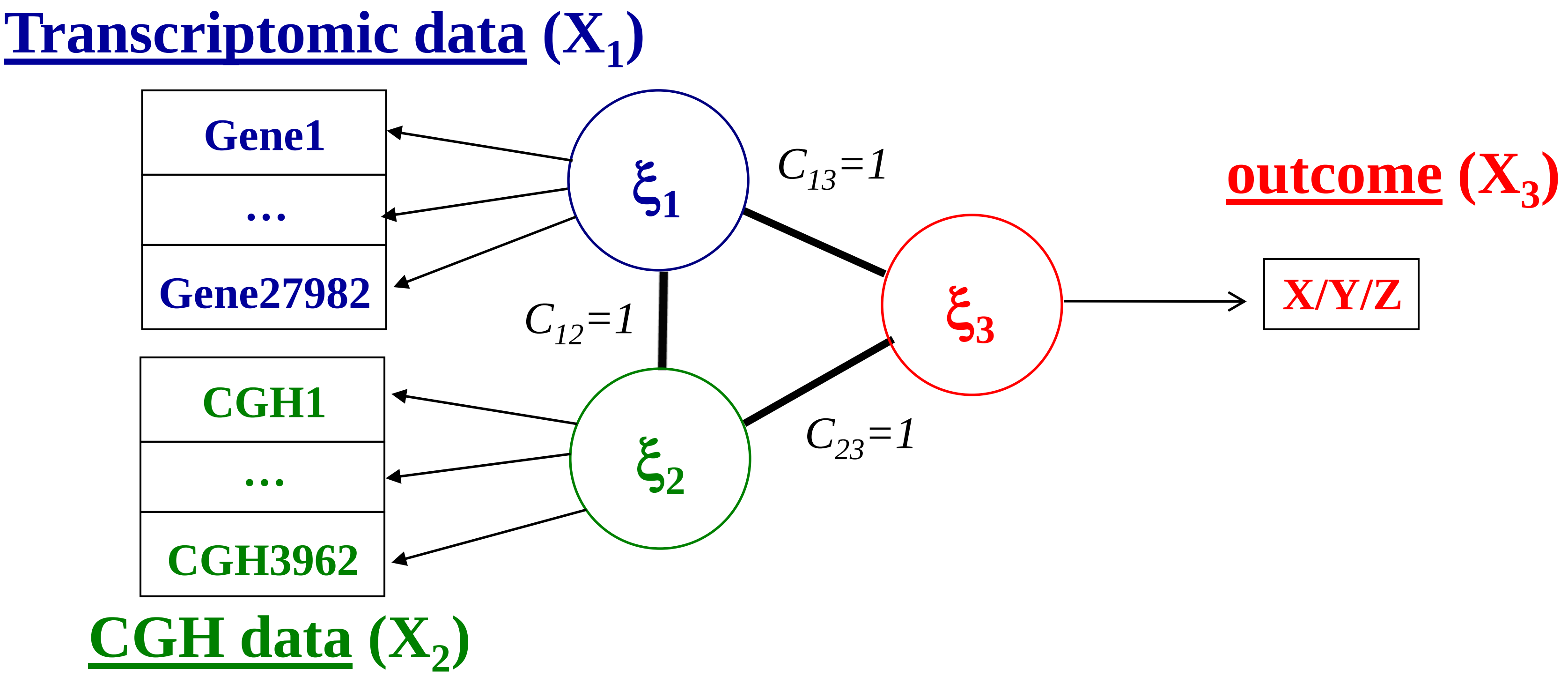
outcome (X_3)

	Gene 1	Gene 2	…	Gene 27982	CGH1	…	CGH 3268	Outcome
Patient 1	0.18	-0.21		-0.73	0.00		-0.55	X
Patient 2	1.15	-0.45		0.27	-0.30		0.00	X
Patient 3	1.35	0.17		0.22	0.33		0.64	Y
⋮								
⋮								
Patient 36	1.39	0.18		-0.17	0.00		0.43	Z



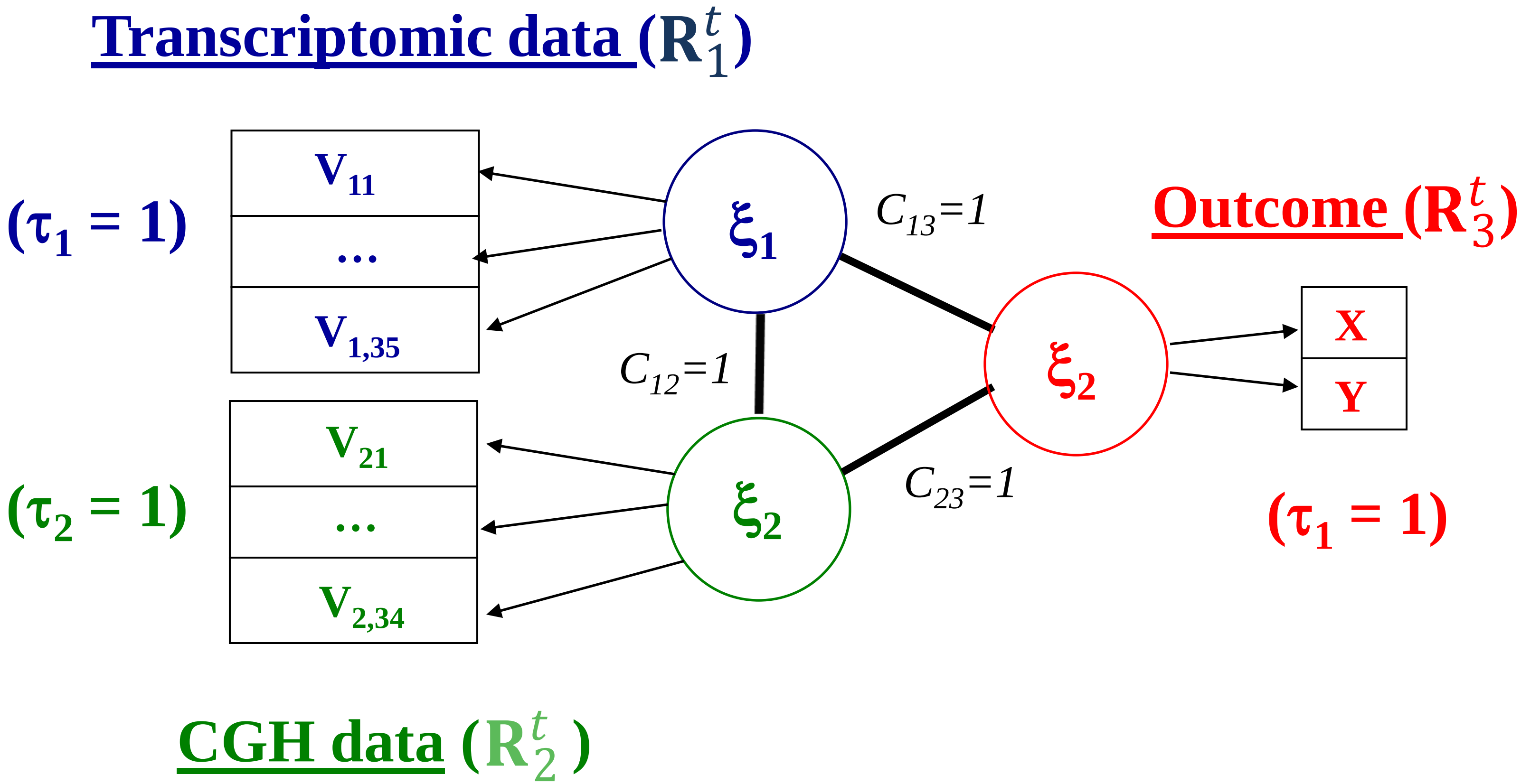
CGH data (X_2)

Glioma Cancer : from a RGCCA point of view

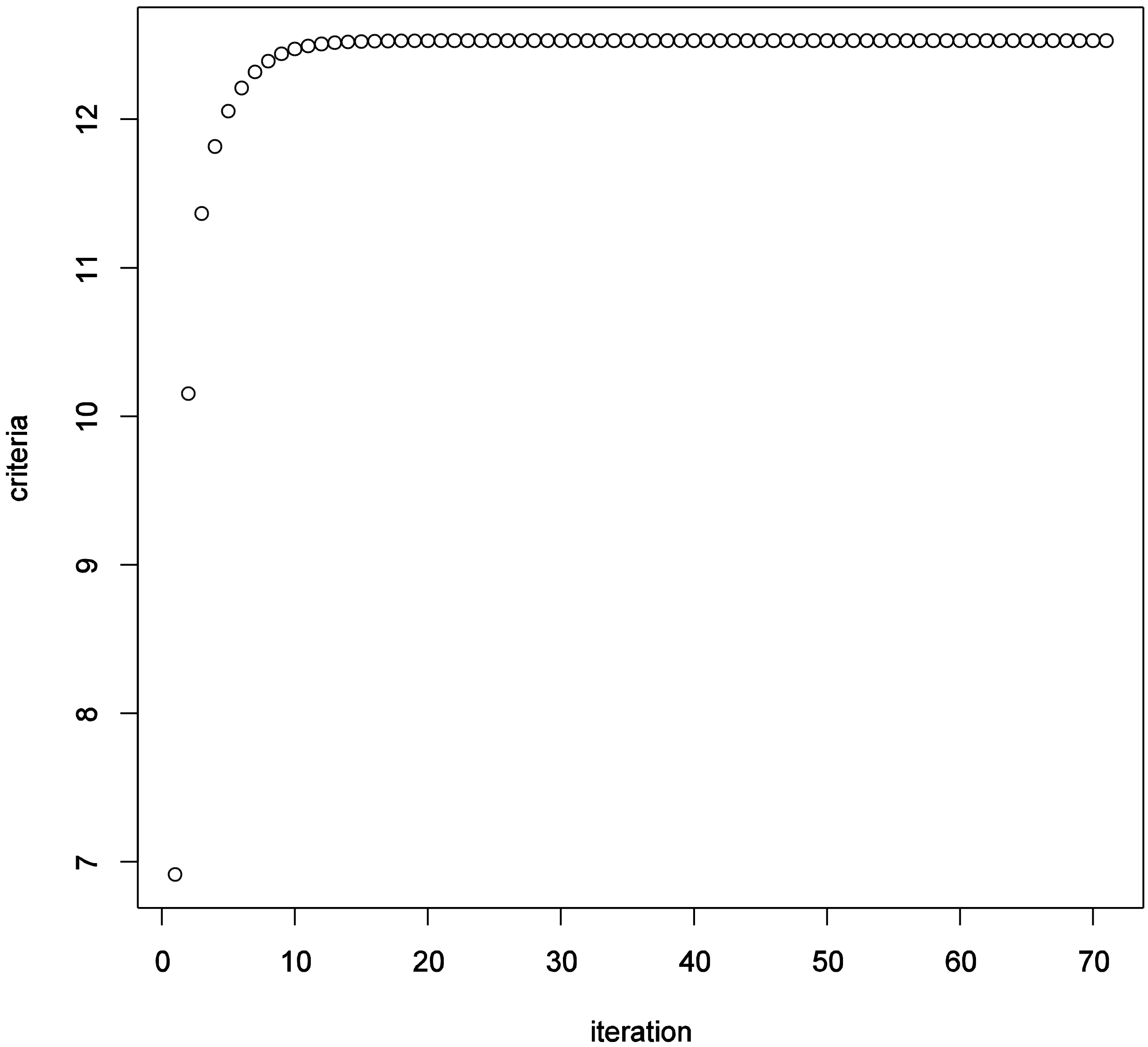


Glioma Cancer : from a KGCCA point of view

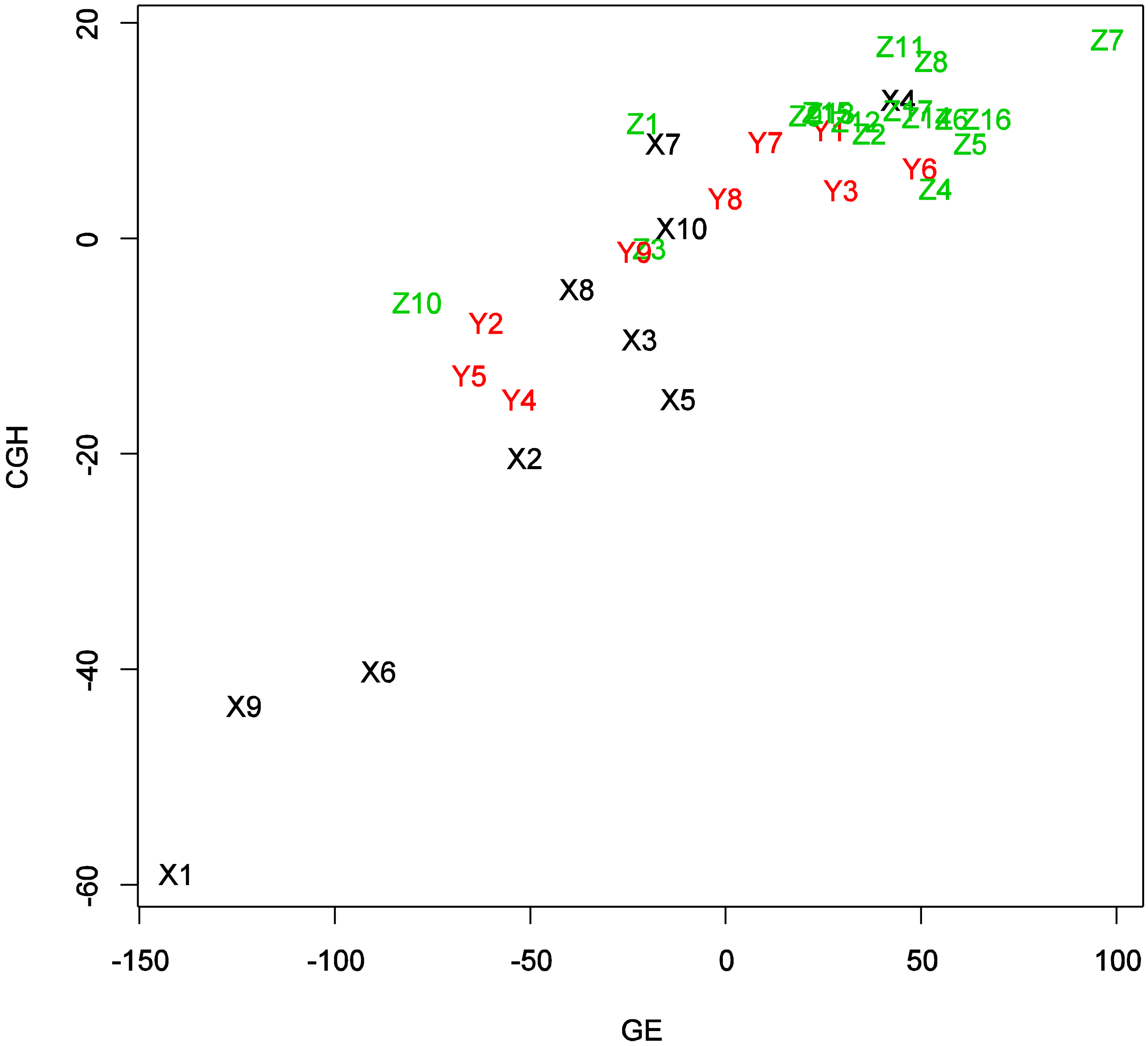
Factorial scheme, linear kernel, $\tau_1 = 1$



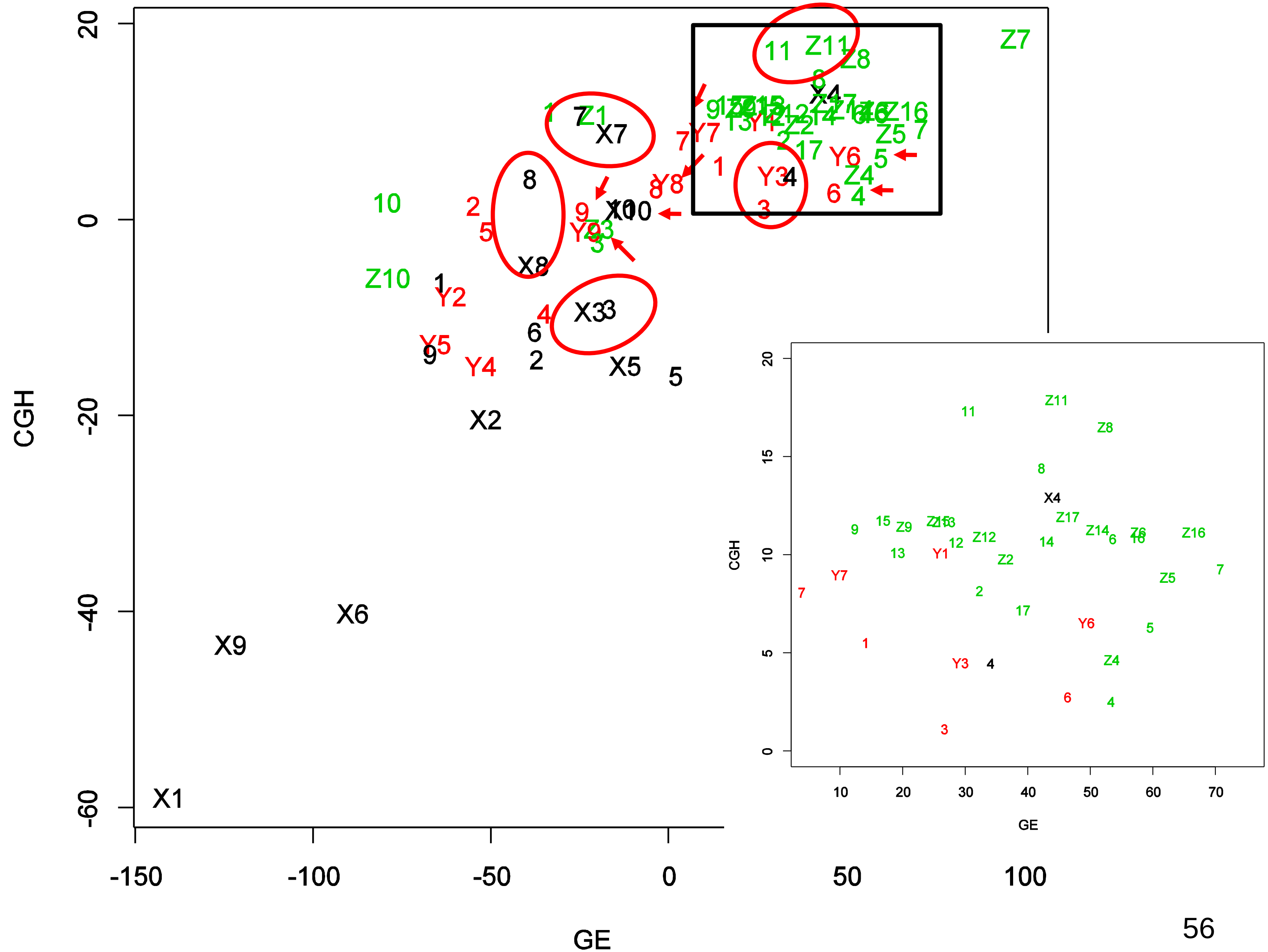
Monotone convergence of the KGCCA algorithm



$\mathbf{y}_1$ versus $\mathbf{y}_2$ (learning phase)



$\mathbf{y}_1$ versus $\mathbf{y}_2$ (leave one out phase)



As conclusion: special cases of KGCCA

Redundancy analysis[2,3] **:** Barker M. & Rayens W. (2003): Partial least squares for discrimination, *Journal of Chemometrics*, 17, 166-173.

Caroll GCCA : Carroll, J.D. (1968): A generalization of canonical correlation analysis to three or more sets of variables, *Proc. 76th Conv. Am. Psych. Assoc*., pp. 227-228.

MCOA : Chessel D. and Hanafi M. (1996): Analyse de la co-inertie de *K* nuages de points. *Revue de Statistique Appliquée*, 44, 35-60

SSQCOV: Hanafi M. & Kiers H.A.L. (2006): Analysis of K sets of data, with differential emphasis on agreement between and within sets, *Computational Statistics & Data Analysis*, 51, 1491-1508.

SUMCOR[4]**:** Horst P. (1961): Relations among m sets of variables, *Psychometrika*, vol. 26, pp. 126-149.

SSQCOR: Kettenring J.R. (1971): Canonical analysis of several sets of variables, *Biometrika*, 58, 433-451

Caroll's RGCCA : Takane Y., Hwang H. and Abdi H. (2008): Regularized Multiple-set Canonical Correlation Analysis, *Psychometrika*, 73 (4):753-775

PLS path modeling: Tenenhaus M., Esposito Vinzi V., Chatelin Y.-M., Lauro C. (2005): PLS path modeling. *Computational Statistics and Data Analysis*, 48, 159-205.

Inter-battery factor Analysis[5]**:** Tucker L.R. (1958): An inter-battery method of factor analysis, *Psychometrika*, vol. 23, n°2, pp. 111-136.

MAXDIFF : Van de Geer J. P. (1984): Linear relations among k sets of variables. *Psychometrika*, 49, 70-94.

Regularized CCA : Vinod H. D. (1976): Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4, 147–166.

Generalized orthogonal multiple co-inertia Analysis: Vivien M. & Sabatier R. (2003): Generalized orthogonal multiple co-inertia analysis (-PLS): new multiblock component and regression methods, *Journal of Chemometrics*, 17, 287-301.

PLS regression[1]**:** Wold S., Martens & Wold H. (1983): The multivariate calibration problem in chemistry solved by the PLS method. In Proc. Conf. Matrix Pencils, Ruhe A. & Kåstrøm B. (Eds), March 1982, Lecture Notes in Mathematics, Springer Verlag, Heidelberg, p. 286-293.

(1) **Rosipal R., Trejo L.J.,** Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert space, Journal of Machine Learning Research, 2:97- 123, 2001.
(2) **Rosipal R., Trejo L.J., Matthews B.** Kernel PLS-SVC for Linear ans Nonlinear Classification, Twentieth International Conference on Machine Learning, Washington DC, 640-647, 2003.
(3) **Takane Y., Hwang H.,** Regularized linear and kernel redundancy analysis, Computational Statistics & Data Analysis 52: 394–405, 2007
(4) **Bach F. R. and Jordan M. I.** Kernel independent component analysis. Journal of Machine Learning Research, 3:1–48, 2002.
(5) **A. Gretton, A. Smola, O. Bousquet, R. Herbrich, A. Belitski, M. Augath, Y. Murayama, J. Pauls, B. Schölkopf, and N. Logothetis**. Kernel constrained covariance for dependence measurement, AISTATS, volume 10, 2005b.