(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

### 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

### Instability of executive (45-61)

- **INST**: Instability of executive

### Nb of violent internal war incidents (46-61)

- **ECKS**: Nb of violent internal war incidents

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

- **DEAT**: Nb of people killed as a result of civic group violence

### Stable democracy

- **D-STAB**: Stable democracy

### Unstable democracy

- **D-UNST**: Unstable democracy

### Dictatorship

- **DICT**: Dictatorship
## Economic inequality and political instability
(Data from Russett, 1964)

### Agricultural inequality

<table>
<thead>
<tr>
<th></th>
<th>Gini</th>
<th>Farm</th>
<th>Rent</th>
<th>Gnpr</th>
<th>Labo</th>
<th>Inst</th>
<th>Ecks</th>
<th>Deat</th>
<th>Demo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentine</td>
<td>86.3</td>
<td>98.2</td>
<td>32.9</td>
<td>374</td>
<td>25</td>
<td>13.6</td>
<td>57</td>
<td>217</td>
<td>2</td>
</tr>
<tr>
<td>Australie</td>
<td>92.9</td>
<td>99.6</td>
<td>3.27</td>
<td>1215</td>
<td>14</td>
<td>11.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Autriche</td>
<td>74.0</td>
<td>97.4</td>
<td>10.7</td>
<td>532</td>
<td>32</td>
<td>12.8</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>58.3</td>
<td>86.1</td>
<td>26.0</td>
<td>1046</td>
<td>26</td>
<td>16.3</td>
<td>46</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Yougoslavie</td>
<td>43.7</td>
<td>79.8</td>
<td>0.0</td>
<td>297</td>
<td>67</td>
<td>0.0</td>
<td>9</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Political instability

<table>
<thead>
<tr>
<th></th>
<th>Inst</th>
<th>Ecks</th>
<th>Deat</th>
<th>Demo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentine</td>
<td>13.6</td>
<td>57</td>
<td>217</td>
<td>2</td>
</tr>
<tr>
<td>Australie</td>
<td>11.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Autriche</td>
<td>12.8</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>16.3</td>
<td>46</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Yougoslavie</td>
<td>0.0</td>
<td>9</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Definitions

- **GINI**: Inequality of land distributions
- **FARM**: % farmers that own half of the land (> 50)
- **RENT**: % farmers that rent all their land
- **GNPR**: Gross national product per capita ($, 1955)
- **LABO**: % of labor force employed in agriculture
- **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

**Agricultural inequality** \((X_1)\)

- **GINI**
- **FARM**
- **RENT**
- **GNPR**
- **LABO**

\(\xi_1\)

\(C_{13} = 1\)

\(C_{12} = 0\)

**Industrial development** \((X_2)\)

**Political instability** \((X_3)\)

- **INST**
- **ECKS**
- **DEAT**
- **D-STB**
- **D-INS**
- **DICT**

\(C_{23} = 1\)
Block components

\[ Y_1 = X_1 w_1 = w_{11} GINI + w_{12} FARM + w_{13} RENT \]

\[ Y_2 = X_2 w_2 = w_{21} GNPR + w_{22} LABO \]

\[ Y_3 = X_3 w_3 = w_{31} INST + w_{32} ECKS + w_{33} DEATH \]
\[ + w_{34} D-STB + w_{35} D-UNST \]
\[ + w_{36} DICT \]
### Some modified multi-block methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMCOR (Horst, 1961)</td>
<td>( \text{Max} \sum_{j,k} \text{Cor}(X_{j\cdot w_{j}}, X_{k\cdot w_{k}}) )</td>
</tr>
<tr>
<td>SSQCOR (Mathes, 1993, Hanafi, 2004)</td>
<td>( \text{Max} \sum_{j,k}</td>
</tr>
<tr>
<td>SABSCOR (Mathes, 1993, Hanafi, 2004)</td>
<td>( \text{Max} \sum_{j,k} \text{Cor}(X_{j\cdot w_{j}}, X_{k\cdot w_{k}}) )</td>
</tr>
<tr>
<td>MAXDIFF (Van de Geer, 1984)</td>
<td>( \text{Max} \sum_{j,k} \text{Cov}(X_{j\cdot w_{j}}, X_{k\cdot w_{k}}) )</td>
</tr>
<tr>
<td>SABSCOV (Krämer, 2007)</td>
<td>( \text{Max} \sum_{j,k}</td>
</tr>
</tbody>
</table>

**GENERALIZED CANONICAL CORRELATION ANALYSIS**

**GENERALIZED CANONICAL COVARIANCE ANALYSIS**
Covariance-based criteria

\[ c_{jk} = 1 \text{ if blocks are linked, } 0 \text{ otherwise and } c_{jj} = 0 \]

<table>
<thead>
<tr>
<th>SUMCOR</th>
<th>( \underset{\text{Max}}{\sum_{j,k}} c_{jk} \text{Cov}(X_j w_j, X_k w_k) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSQCOR</td>
<td>( \underset{\text{Max}}{\sum_{j,k}} c_{jk} \text{Cov}^2(X_j w_j, X_k w_k) )</td>
</tr>
<tr>
<td>SABSCOR</td>
<td>( \underset{\text{Max}}{\sum_{j,k}} c_{jk}</td>
</tr>
<tr>
<td>SUMCOV</td>
<td>( \underset{\text{Max}}{\sum_{j,k}} c_{jk} \text{Cov}(X_j w_j, X_k w_k) )</td>
</tr>
<tr>
<td>SSQCOV</td>
<td>( \underset{\text{Max}}{\sum_{j,k}} c_{jk} \text{Cov}^2(X_j w_j, X_k w_k) )</td>
</tr>
<tr>
<td>SABSCOV</td>
<td>( \underset{\text{Max}}{\sum_{j,k}} c_{jk}</td>
</tr>
</tbody>
</table>
Starting point: RGGCA (for linear dependence)

\[
\argmax_{a_1, a_2, \ldots, a_J} \sum_{j \neq k}^J c_{j,k} g \left( \text{cov}(x_j a_j, x_k a_k) \right)
\]

Subject to the constraints:

\[ (1 - \tau_j) \text{var}(x_j a_j) + \tau_j \|a_j\|^2 = 1, \quad j = 1, \ldots, J \]

where:

\[ g = \begin{cases} 
\text{identity} & \text{(Horst scheme)} \\
\text{square} & \text{(Factorial scheme)} \\
\text{absolute value} & \text{(Centroid scheme)}
\end{cases} \]

and:

\[ \tau_j = \text{Shrinkage constant between 0 and 1} \]

A monotone convergent algorithm related to this optimization problem will be described.
RGCCA applied to the Russett data

**Agricultural inequality** \((X_1)\)

\[
\begin{align*}
&\text{Maximize } \quad g(\text{Cov}(X_1 a_1, X_3 a_3)) + g(\text{Cov}(X_2 a_2, X_3 a_3)) \\
&\text{subject to the constraints } \quad \tau_j \|a_j\|^2 + (1 - \tau_j) \text{Var}(X_j a_j) = 1, \quad j = 1, 2, 3 \\
&0 \leq \tau_j \leq 1, \quad g = \text{identity, square or absolute value}
\end{align*}
\]
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 ($\approx$ Gauss-Seidel algorithm).
• This procedure is monotonically convergent: the criterion increases at each step of the algorithm.
The general algorithm

Initial step

\[ y_j = X_jw_j \]

Outer Estimation (explains the block)

\[ \tau_j \|w_j\|^2 + (1 - \tau_j)\text{Var}(X_jw_j) = 1 \]

Iterate until convergence of the criterion.

\[ w_j = \frac{\left(1 - \tau_j\right)\frac{1}{n}X_j^tX_j + \tau_jI} {\sqrt{z_j^tX_j\left(1 - \tau_j\right)\frac{1}{n}X_j^tX_j + \tau_jI}} X_j^tZ_j \]

Inner Estimation (explains relations between blocks)

\[ z_j = \sum_{k \neq j} e_{jk}y_k \]

Choice of weights \( e_{jh} \):
- Horst: \( e_{jh} = c_{jh} \)
- Centroid: \( e_{jh} = c_{jh} \text{sign}(\text{cor}(Y_h, Y_j)) \)
- Factorial: \( e_{jh} = c_{jh} \text{cov}(Y_h, Y_j) \)

\( c_{jh} = 1 \) if blocks are linked, 0 otherwise and \( c_{jj} = 0 \)
<table>
<thead>
<tr>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SSQCOV:</strong></td>
<td>Hanafi M. &amp; Kiers H.A.L. (2006): Analysis of $K$ sets of data, with differential emphasis on agreement between and within sets, Computational Statistics &amp; Data Analysis, 51, 1491-1508.</td>
</tr>
<tr>
<td><strong>SUMCOR:</strong></td>
<td>Horst P. (1961): Relations among $m$ sets of variables, Psychometrika, vol. 26, pp. 126-149.</td>
</tr>
<tr>
<td><strong>SSQCOR:</strong></td>
<td>Kettingring J.R. (1971): Canonical analysis of several sets of variables, Biometrika, 58, 433-451</td>
</tr>
<tr>
<td><strong>Carroll’s RGCCA :</strong></td>
<td>Takane Y., Hwang H. and Abdi H. (2008): Regularized Multiple-set Canonical Correlation Analysis, Psychometrika, 73 (4):753-775</td>
</tr>
<tr>
<td><strong>Carroll’s GCCA :</strong></td>
<td>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.</td>
</tr>
</tbody>
</table>
The two-block case: Regularized CCA

Maximize \( \text{Cov}(X_1a_1, X_2a_2) \)
subject to \( \tau_j \|a_j\| + (1 - \tau_j) \text{Var}(X_ja_j) = 1 \)

Special cases

<table>
<thead>
<tr>
<th>Method</th>
<th>Criterion</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS regression</td>
<td>Maximize ( \text{Cov}(X_1a_1, X_2a_2) )</td>
<td>( |a_1| = |a_2| = 1 )</td>
</tr>
<tr>
<td>Canonical Correlation Analysis</td>
<td>Maximize ( \text{Cor}(X_1a_1, X_2a_2) )</td>
<td>( \text{Var}(X_1a_1) = \text{Var}(X_2a_2) = 1 )</td>
</tr>
<tr>
<td>Redundancy analysis of ( X_1 ) with respect to ( X_2 )</td>
<td>Maximize ( \text{Cor}(X_1a_1, X_2a_2) \text{Var}(X_1a_1)^{1/2} )</td>
<td>( |a_1| = 1 ), ( \text{Var}(X_2a_2) = 1 )</td>
</tr>
</tbody>
</table>

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

Maximize $\text{Cov}(X_1a_1, X_2a_2)$

subject to $\tau_j \|a_j\| + (1 - \tau_j) \text{Var}(X_j a_j) = 1$

### Special cases

<table>
<thead>
<tr>
<th>Method</th>
<th>Criterion</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS regression</td>
<td>Maximize $\text{Cov}(X_1a_1, X_2a_2)$ $|a_1|=|a_2|=1$</td>
<td>Is favoring too much stability with respect to correlation</td>
</tr>
<tr>
<td>Canonical Correlation Analysis</td>
<td>Maximize $\text{Cor}(X_1a_1, X_2a_2)$</td>
<td>Is favoring too much correlation with respect to stability</td>
</tr>
</tbody>
</table>
Choice of the shrinkage constant $\tau_j$

Maximize $\text{Cov}(X_1a_1, X_2a_2)$

subject to $\tau_j \|a_j\| + (1 - \tau_j) \text{Var}(X_ja_j) = 1$

Schäfer and Strimmer (2005) give a formula for an optimal choice of $\tau_j$. 
Special cases of Regularized generalized CCA

**RGCCA and Multi-block data analysis**

<table>
<thead>
<tr>
<th>Method</th>
<th>Objective Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMCOR (Horst, 1961)</td>
<td>( \max_{\var{X}_j a_j, \var{X}<em>k a_k} \sum</em>{j, k, j \neq k} \text{Cor}(\var{X}_j a_j, \var{X}_k a_k) )</td>
</tr>
<tr>
<td>SSQCOR (Kettenring, 1971)</td>
<td>( \max_{\var{X}_j a_j, \var{X}<em>k a_k} \sum</em>{j, k, j \neq k} \text{Cor}^2(\var{X}_j a_j, \var{X}_k a_k) )</td>
</tr>
<tr>
<td>SABSCOR (Mathes, 1993, Hanafi, 2004)</td>
<td>( \max_{\var{X}_j a_j, \var{X}<em>k a_k} \sum</em>{j, k, j \neq k} \left</td>
</tr>
<tr>
<td>MAXDIFF (Van de Geer, 1984)</td>
<td>( \max_{|a_j|=1} \sum_{j, k, j \neq k} \text{Cov}(\var{X}_j a_j, \var{X}_k a_k) )</td>
</tr>
<tr>
<td>MAXDIFF B (Hanafi &amp; Kiers, 2006)</td>
<td>( \max_{|a_j|=1} \sum_{j, k, j \neq k} \text{Cov}^2(\var{X}_j a_j, \var{X}_k a_k) )</td>
</tr>
<tr>
<td>SABSCOV (Krämer, 2007)</td>
<td>( \max_{|a_j|=1} \sum_{j, k, j \neq k} \left</td>
</tr>
</tbody>
</table>
Special cases of Regularized generalized CCA

Hierarchical models

(a) One second order block

(b) Several second order blocks

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 2\textsuperscript{nd} order block

<table>
<thead>
<tr>
<th>Method</th>
<th>Criterion</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical PLS regression</td>
<td>Maximize $\sum_{j=1}^{J} g(\text{Cov}(X_j a_j, X_{J+1} a_{J+1}))$</td>
<td>$|a_j| = 1, j = 1, \ldots, J + 1$</td>
</tr>
<tr>
<td>Hierarchical Canonical Correlation Analysis</td>
<td>Maximize $\sum_{j=1}^{J} g(\text{Cor}(X_j a_j, X_{J+1} a_{J+1}))$</td>
<td>$\text{Var}(X_j a_j) = 1, j = 1, \ldots, J + 1$</td>
</tr>
<tr>
<td>Hierarchical Redundancy analysis of the $X_j$’s with respect to $X_{J+1}$</td>
<td>Maximize $\sum_{j=1}^{J} g(\text{Cor}(X_j a_j, X_{J+1} a_{J+1})\text{Var}(X_j a_j)^{1/2})$</td>
<td>$|a_j| = 1, j = 1, \ldots, J$; $\text{Var}(X_{J+1} a_{J+1}) = 1$</td>
</tr>
<tr>
<td>Hierarchical Redundancy analysis of $X_{J+1}$ with respect to the $X_j$’s</td>
<td>Maximize $\sum_{j=1}^{J} g(\text{Cor}(X_j a_j, X_{J+1} a_{J+1})\text{Var}(X_{J+1} a_{J+1})^{1/2})$</td>
<td>$\text{Var}(X_j a_j) = 1, j = 1, \ldots, J$; $|a_{J+1}| = 1$</td>
</tr>
</tbody>
</table>

$g = \text{identity, square or absolute value}$
Special cases of Regularized generalized CCA

Hierarchical model: one 2\textsuperscript{nd} order block
Factorial scheme: \( g = \text{square function} \)

Concordance analysis (Hanafi & Lafosse, 2001)

Maximize \( \sum_{j=1}^{J} \text{Cov}^2(X_j M_j b_j, X_{J+1} M_{J+1} b_{J+1}) \)
subject to \( b_j^t M_j b_j = 1, \ j = 1, \ldots, J + 1 \)

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

\[ X_{J+1} = [X_1, \ldots, X_J] \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Criterion</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMCOR (Horst, 1961)</td>
<td>Maximize [ \sum_{j=1}^{J} \text{Cor}(X_j a_j, X_{J+1} a_{J+1}) ] or Maximize [ \sum_{j=1}^{J}</td>
<td>\text{Cor}(X_j a_j, X_{J+1} a_{J+1})</td>
</tr>
<tr>
<td>Generalized CCA (Carroll, 1968a,b)</td>
<td>Maximize [ \sum_{j=1}^{J_1} \text{Cor}^2(X_j a_j, X_{J+1} a_{J+1}) ] + [ \sum_{j=J_1+1}^{J} \text{Cov}^2(X_j a_j, X_{J+1} a_{J+1}) ]</td>
<td>[ \text{Var}(X_j a_j) = 1, \ j = 1, \ldots, J_1, J + 1 ] [ |a_j| = 1, \ j = J_1 + 1, \ldots, J ]</td>
</tr>
</tbody>
</table>
Special cases of Regularized generalized CCA

Hierarchical model: one 2nd order block

\[ X_{J+1} = [X_1, \ldots, X_J] \]

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

Maximize \[ \sum_{j=1}^{J} \text{Cov}^2(X_j a_j, X_{J+1} a_{J+1}) \]
subject to \[ \|a_j\| = 1, j = 1, \ldots, J, \ \text{Var}(X_{J+1} a_{J+1}) = 1 \]

Special case of Carroll’s GCCA
Special cases of Regularized generalized CCA

Hierarchical model: several 2\textsuperscript{nd} order blocks

\[ c_{jk} = 1 \quad \text{if response block} \ X_k \text{ is connected to predictor block} \ X_j, \]
\[ = 0 \quad \text{otherwise} \]
Maximize \[ \sum_{j=1}^{J_1} \sum_{k=J_1+1}^{J} c_{jk} g(\text{Cov}(X_j a_j , X_k a_k)) \]
subject to the constraints: \( \tau_j \|a_j\|^2 + (1 - \tau_j) \text{Var}(X_j a_j) = 1, \quad j = 1, \ldots, J \)

\( g = \text{identity, square or absolute value} \)
Special cases of Regularized generalized CCA

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

Maximize \[ \sum_{j=1}^{J_1} \sum_{k=J_1+1}^{J} \text{Cov}(X_j a_j, X_k a_k) \]
subject to the constraints: \[ \|a_j\| = 1, \quad j = 1, \ldots, J \]
RGCCA applied to the Russett data

**Agricultural inequality** $(X_1)$

- GINI
- FARM
- RENT
- GNPR
- LABO

**Industrial development** $(X_2)$

**Political instability** $(X_3)$

Maximize \( g(Cov(X_1a_1, X_3a_3)) + g(Cov(X_2a_2, X_3a_3)) \)

subject to the constraints \( \tau_j \|a_j\|^2 + (1-\tau_j)Var(X_ja_j) = 1, \quad j = 1, 2, 3 \)

\( 0 \leq \tau_j \leq 1, \quad g = \text{identity, square or absolute value} \)
Monotone convergence of the algorithm
**RGCCA: all $\tau_i = 1$ - centroid scheme**

$$\text{Max} \left[ |\text{Cov}(X_1 w_1, X_3 w_3)| + |\text{Cov}(X_2 w_2, X_3 w_3)| \right] = 2.69$$

### 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

$$\text{Max} \left[ \text{Cov}^2 (X_1w_1, X_3w_3) + \text{Cov}^2 (X_2w_2, X_3w_3) \right] = 3.86$$
RGCCA: $\tau_i = \tau^*_i$ - centroid scheme

\[ y_1 = X_1w_1 \]
\[ y_2 = X_2w_2 \]
\[ y_3 = X_3w_3 \]

\[ \tau_1 = 0.14 \]
\[ \tau_2 = 0.07 \]
\[ \tau_3 = 0.12 \]

Correlation:
- $y_1$ with $y_2$: cor = 0.92
- $y_1$ with $y_3$: cor = 0.54
- $y_2$ with $y_3$: cor = -0.78
Generalized Barker & Rayens PLS-DA
\[ \tau_1 = \tau_2 = 1 \text{ and } \tau_3 = 0 \] - Factorial scheme

\[ \text{Max}\left[ \text{Cor}^2(X_1w_1, X_3w_3) \times \text{Var}(X_1w_1) + \text{Cor}^2(X_2w_2, X_3w_3) \times \text{Var}(X_2w_2) \right] = 1.39 \]
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)

\[ X_{11} \text{ with block } X_2 = [X_{21}, X_{22}] \]  
\[ X_{12} \text{ with block } X_2 = [X_{21}, X_{22}] \]
RGCCA results

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

\[ y_1 = X_1 a_1 \]

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 \((X_1)\)

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gene 1</th>
<th>Gene 2</th>
<th>...</th>
<th>Gene 27982</th>
<th>CGH1</th>
<th>...</th>
<th>CGH 3268</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>0.18</td>
<td>-0.21</td>
<td></td>
<td>-0.73</td>
<td>0.00</td>
<td></td>
<td>-0.55</td>
<td>(X)</td>
</tr>
<tr>
<td>Patient 2</td>
<td>1.15</td>
<td>-0.45</td>
<td></td>
<td>0.27</td>
<td>-0.30</td>
<td></td>
<td>0.00</td>
<td>(X)</td>
</tr>
<tr>
<td>Patient 3</td>
<td>1.35</td>
<td>0.17</td>
<td></td>
<td>0.22</td>
<td>0.33</td>
<td></td>
<td>0.64</td>
<td>(Y)</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>Patient 36</td>
<td>1.39</td>
<td>0.18</td>
<td></td>
<td>-0.17</td>
<td>0.00</td>
<td></td>
<td>0.43</td>
<td>(Z)</td>
</tr>
</tbody>
</table>

CGH data \((X_2)\)

Outcome \((X_3)\)
RGCCA in action in a very high dimensional blocks settings

• Requires to invert $J$ matrices

$$M_j = (1 - \tau_j) \frac{1}{n} X_j^t X_j + \tau_j I_{p_j} \quad j = 1, \ldots, J$$

where $M_j \in \mathbb{R}^{p_j \times p_j} \quad j = 1, \ldots, J$

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

\[
\Phi_j(x) = k_j(., x) \quad j = 1, \ldots, J
\]

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 $\Phi_1(x_1^{(i)})$ on $f_1$ and $y_{2i}$ of $\Phi_2(x_2^{(i)})$ on $f_2$ by:

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

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:

$$\arg\max_{f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2} \text{cov}(y_1, y_2)$$

s.t. $$(1 - \tau_j) \text{var}(y_j) + \tau_j \|f_j\|_{\mathcal{H}_j}^2 = 1, \quad j = 1, 2$$
Kernel GCCA (two-block case)

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

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

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

Let us note $K_j$ the $n \times n$ kernel matrix defined by:

$$(K_j)_{ki} = k (x_j^{(k)}, x_j^{(i)}) = \langle \Phi_j (x_j^{(k)}), \Phi_j (x_j^{(i)}) \rangle$$

We deduce that $y_j$ is defined by:

$$y_j = K_j \alpha_j$$
Kernel GCCA (two-block case)

Then the previous optimization problem:

\[
\arg\max_{f_1 \in H_1, f_2 \in H_2} \text{cov}(y_1, y_2) \\
\text{s.c. } (1 - \tau_j) \text{var}(y_j) + \tau_j \|f_j\|_F^2 = 1, \quad j = 1, 2
\]

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

\[
\arg\max_{\alpha_1, \alpha_2} \frac{1}{n} \alpha_1^t K_1 K_2 \alpha_2 \\
\text{s.c. } (1 - \tau_j) \frac{1}{n} \alpha_j^t K_j^2 \alpha_j + \tau_j \alpha_j^t K_j \alpha_j = 1 \quad j = 1, 2
\]
Kernel KGCCA (J-block case)

\[
\begin{align*}
\text{argmax} \quad & \sum_{j \neq k}^{J} c_{jk} \ g \left( \frac{1}{n} \alpha_j^t K_j K_k \alpha_k \right) \\
\text{s.c.} \quad & (1 - \tau_j) \frac{1}{n} \alpha_j^t K_j^2 \alpha_j + \tau_j \alpha_j^t K_j \alpha_j = 1 \quad j = 1, \ldots, J
\end{align*}
\]

Requires to invert \( J \) matrices

\[
N_j = (1 - \tau_j) \frac{1}{n} K_j^2 + \tau_j K_j \quad j = 1, \ldots, J
\]

where \( N_j \in \mathbb{R}^{n \times n} \quad j = 1, \ldots, J \)

Problem: \( K_j \) (and thus \( N_j \)) 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 $\mathbf{R}_j \in \mathbb{R}^{\text{rank } (\mathbf{K}_j) \times n}$

$$
\arg\max_{\alpha_1, \alpha_2, \ldots, \alpha_J} \sum_{j \neq k}^J c_{j,k} \cdot g\left( \frac{1}{n} \alpha_j^t \mathbf{R}_j^t \mathbf{R}_j \mathbf{R}_k^t \mathbf{R}_k \alpha_k \right)
$$

s.c. $\left(1 - \tau_j\right) \frac{1}{n} \alpha_j^t \mathbf{R}_j^t \mathbf{R}_j \mathbf{R}_j^t \mathbf{R}_j \alpha_j + \tau_j \alpha_j^t \mathbf{R}_j^t \mathbf{R}_j \alpha_j = 1, \quad j = 1, \ldots, J$
The KGCCA optimization problem

\[
\arg\max_{w_1, w_2, \ldots, w_J} \sum_{j,k} c_{jk} \ g \left( \frac{1}{n} w_j^t R_j R_k^t w_k \right)
\]

s.c. \( (1 - \tau_j) \frac{1}{n} w_j^t R_j R_j^t w_j + \tau_j w_j^t w_j = 1, \ j = 1, \ldots, J \)

Apply the initial algorithm of RGCCA on \( R_1^t, \ldots, R_J^t \) to obtain the Latent variables outer estimation.

In the glioma Cancer Data:

\[
X_1 \in \mathbb{R}^{36 \times 27982} \rightarrow R_1^t \in \mathbb{R}^{36 \times 35}
\]
\[
X_2 \in \mathbb{R}^{36 \times 3268} \rightarrow R_2^t \in \mathbb{R}^{36 \times 34}
\]
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 RGCCA procedure to solve the stationary equations.
• This procedure is monotonically convergent: the criterion increases at each step of the algorithm.
The Kernel GCCA algorithm

Outer Estimation (explains the block)

\[ w'_j \left( (1 - \tau_j) \frac{1}{n} R_j R'_j + \tau_j I_{\text{rank}(K_j)} \right) w_j = 1 \]

Inner Estimation (explains relation between block)

\[ z_j = \sum_{k \neq j} e_{jk} y_k \]

Initial step

Iterate until convergence of the criterion

\[ w_j = \frac{\left( (1 - \tau_j) \frac{1}{n} R_j R'_j + \tau_j I_{\text{rank}(K_j)} \right)^{-1} R_j z_j}{\sqrt{z'_j R'_j \left( (1 - \tau_j) \frac{1}{n} R_j R'_j + \tau_j I_{\text{rank}(K_j)} \right)^{-1} R_j z_j}} \]

Choice of weights \( e_{jh} \):

- Horst: \( e_{jk} = c_{jk} \)
- Centroid: \( e_{jk} = c_{jk} \text{sign} \left( \text{cor}(y_j, y_k) \right) \)
- Factorial: \( e_{jk} = c_{jk} \text{cov}(y_j, y_k) \)

\( c_{jk} = 1 \) 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

\[ y_1 = R_1^t w_1 \]

\[ y_2 = R_2^t w_2 \]
RGCCA results (factorial scheme, $\tau_j = 1$)
KGCCA results (factorial scheme, $\tau_j = 1$, 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 \( (X_1) \)

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
<th>…</th>
<th>Gene 27982</th>
<th>CGH1</th>
<th>…</th>
<th>CGH 3268</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>0.18</td>
<td>-0.21</td>
<td>-0.73</td>
<td>0.00</td>
<td>-0.55</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Patient 2</td>
<td>1.15</td>
<td>-0.45</td>
<td>0.27</td>
<td>-0.30</td>
<td>0.00</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Patient 3</td>
<td>1.35</td>
<td>0.17</td>
<td>0.22</td>
<td>0.33</td>
<td>0.64</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Patient 36</td>
<td>1.39</td>
<td>0.18</td>
<td>-0.17</td>
<td>0.00</td>
<td>0.43</td>
<td>Z</td>
<td></td>
</tr>
</tbody>
</table>

CGH data \( (X_2) \)

outcome \( (X_3) \)
Glioma Cancer: from a RGCCA point of view

Transcriptomic data ($X_1$)

Gene1
...
Gene27982

CGH data ($X_2$)

CGH1
...
CGH3962

outcome ($X_3$)

$C_{13} = 1$

$C_{12} = 1$

$C_{23} = 1$
Glioma Cancer: from a KGCCA point of view

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

**Transcriptomic data** ($R^t_1$)

$V_{11}$

$\ldots$

$V_{1,35}$

$\xi_1$

$C_{13} = 1$

$C_{12} = 1$

**Outcome** ($R^t_3$)

$\xi_2$

$C_{23} = 1$

$X$

$Y$

**CGH data** ($R^t_2$)

$(\tau_1 = 1)$

$(\tau_2 = 1)$

$V_{21}$

$\ldots$

$V_{2,34}$
Monotone convergence of the KGCCA algorithm
$y_1$ versus $y_2$ (learning phase)
$y_1$ versus $y_2$ (leave one out phase)
As conclusion: special cases of KGCCA


Generalized orthogonal multiple co-inertia Analysis:


PLS regression\(^{(1)}\):


---


\(^{(5)}\) 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.