Regularized Generalized Canonical Correlation Analysis and PLS Path Modeling

Arthur Tenenhaus

Regularized Generalized Canonical Correlation Analysis and PLS Path Modeling

Arthur Tenenhaus

Abstract Regularized Generalized Canonical Correlation Analysis (RGCCA) and Partial Least Squares Path Modeling (PLSPM) have been proposed for studying relationships between J sets of variables observed on the same set of individuals taking into account a graph of connection between blocks. The main goal of this this communication, is to compare the various options of PLS-PM and RGCCA. Actually, first comparisons show very close behavior of these two approaches.

Key words: Generalized Canonical Correlation Analysis, Regularization, PLS path modeling

1 Introduction

On the one hand, PLSPM is often used (Wold (1985), Tenenhaus et al. (2005)) for analyzing relationships between several set of variables observed on the same set of individuals. The PLSPM algorithm works very well in practice but, for some options, is lacking an analytical proof of convergence and suffer of lack of optimality condition (the criterion to which the PLSPM-mode A algorithm converges is unknown). On the other hand, Regularized Generalized Canonical Correlation Analysis (RGCCA) (Tenenhaus and Tenenhaus (2011)) is also used for analyzing relationship between several set of variables observed on the same set of individuals. However, RGCCA is based on a monotonically convergent iterative algorithm and has the distinct advantage to rely on an explicit optimization problem. The main goal of this communication is to compare RGCCA and PLS-PM. Actually, first comparisons show very close behavior of these two algorithms. This paper is organized in three sections. PLSPM is presented in section 1, RGCCA is discussed in section 2 and PLSPM and RGCCA are compared on simulated data in section 3.

Arthur Tenenhaus
Name, SUPELEC, e-mail: arthur.tenenhaus@supelec.fr

2 PLS Path Modeling (PLSPM)

PLSPM (Wold (1985); Tenenhaus et al., (2005)) is a framework for analyzing association between J set of variables observed on the same set of n individuals. Let us consider J blocks of variables $\mathbf{X}_1,\ldots,\mathbf{X}_J$. Each block $\mathbf{X}_J$ represents a set of p_j variables observed on the same set of n individuals. PLSPM aims at extracting the information which is shared by the J blocks of variables taking into account *a priori* graph of connections between blocks. In this framework, a design matrix $\mathbf{C} = \{c_{jk}\}$ defines this graph of connections: c_{jk} is equal to 1 if blocks $\mathbf{X}_j$ and $\mathbf{X}_k$ are connected, or is equal to 0 otherwise. Two algorithms for PLSPM were proposed: the Wold's algorithm (1985) and the Lohmöller's algorithm (1989). In this paper we only consider the Wolds PLSPM algorithm with the modes A and B and the centroid, factorial or Horst schemes. This algorithm is described in Algorithm 1.

Algorithm 1 Algorithm for PLS path modeling with $\tau_j = 0$ or $\tau_j = 1$

▷ **Step A. Initialisation:**
Choose arbitrary vectors $\tilde{\mathbf{a}}_j^0$, $j = 1,\ldots,J$ and normalize them such that $\text{var}(\mathbf{X}_j\tilde{\mathbf{a}}_j^0) = 1$

$$\mathbf{a}_j^0 = \left[\frac{1}{n}(\tilde{\mathbf{a}}_j^0)^t\mathbf{X}_j^t\mathbf{X}_j\tilde{\mathbf{a}}_j^0\right]^{-1/2}\tilde{\mathbf{a}}_j^0$$

repeat s = 1, 2, ...
  for $j = 1,2,\ldots,J$ **do**

▷ **Step B. Inner component for** $\mathbf{X}_j$

$$\mathbf{z}_j^s = \sum_{k=1}^{j-1} c_{jk}\text{w}(\text{cov}(\mathbf{X}_j\mathbf{a}_j^s,\mathbf{X}_k\mathbf{a}_k^{s+1}))\mathbf{X}_k\mathbf{a}_k^{s+1} + \sum_{k=j+1}^{J} c_{jk}\text{w}(\text{cov}(\mathbf{X}_j\mathbf{a}_j^s,\mathbf{X}_k\mathbf{a}_k^{s+1}))\mathbf{X}_k\mathbf{a}_k^s$$

where $\text{w}(x) = 1$ for the Horst scheme, x for the factorial scheme and $\text{sign}(x)$ for the centroid scheme and and

▷ **Step C. Outer weight for block** j

$$\mathbf{a}_j^{s+1} = \frac{\left[(1-\tau_j)\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I}\right]^{-1}\mathbf{X}_j^t\mathbf{z}_j^s}{\sqrt{(\mathbf{z}_j^s)^t\mathbf{X}_j\left[(1-\tau_j)\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I}\right]^{-1}\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j\left[(1-\tau_j)\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I}\right]^{-1}\mathbf{X}_j^t\mathbf{z}_j^s}}$$

where $\tau_j = 1$ if mode A is selected for block $\mathbf{X}_j$ and $\tau_j = 0$ if mode B is selected for block $\mathbf{X}_j$
  end for
until convergence

3 Regularized Generalized Canonical Correlation Analysis

Regularized Generalized Canonical Correlation Analysis (RGCCA) proposed in [Tenenhaus & Tenenhaus (2011)] tackles exactly the same problem than PLSPM. RGCCA is defined as the following optimization problem (1):

$$\begin{cases} \max_{\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_J} \sum_{j,k=1:j\neq k}^{J} c_{jk}\, g\left(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k\right) \\ \text{subject to} \quad \mathbf{a}_j^t(1-\tau_j)\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j + \tau_j\mathbf{I})\mathbf{a}_j = 1, \; j = 1,\ldots,J \end{cases} \tag{1}$$

In this optimization problem, g may be defined as $g(x) = x$ (Horst scheme proposed in (Kramer (2007)), $g(x) = |x$ (Centroid scheme proposed in (Wold, 1985) or $g(x) = x^2$ (Factorial scheme proposed in (Lohmöller (1989)). In this context, τ_j varies between 0 and 1. The vector $\mathbf{a}_j$ (resp. $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$) is referred to as an outer weight vector (resp. an outer component) and $\mathbf{z}_j$ is referered to as an inner component. The Horst scheme penalizes structural negative correlation between components while both Centroid and Factorial schemes can be viewed as attractive alternatives that enable two components to be negatively correlated. Optimization problem (1) is limited to these three schemes because they are the most used in the multiblock and Partial Least Squares literature.

From the viewpoint of optimization problem (1), the shrinkage parameters $\tau_j \in [0,1]$, $j = 1,\ldots,J$ interpolate smoothly between the maximisation of the covariance (all $\tau_j = 1$) and the maximisation of the correlation (all $\tau_j = 0$). The choice of the shrinkage parameters requires being clear on the goal pursued by the RGCCA analysis. A guideline is defined for the choice of the shrinkage constants by providing interpretations on the properties of the resulting block components.

- The covariance based model ($\tau_j = 1$ - a.k.a. RGCCA mode A) tends in first to find "stable" (large variance) block components $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$, $j = 1,\ldots,J$, while taking into account the correlations with neighboring components (second priority).

- The correlation based model ($\tau_j = 0$ - a.k.a. mode B) gives priority to the correlation between neighboring components and tends to find unstable block components $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$, $j = 1,\ldots,J$. It is worth noticing that RGCCA-Mode B gives exactly PLSPM-mode B.

- $0 < \tau_j < 1$ (a.k.a. mode ridge) yields a compromise between stability and correlation. This mode tends to find block components $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$, $j = 1,\ldots,J$ with large variance and at the same time are well correlated to its neighboring components. Mode B and RGCCA-Mode A, are unified towards a regularization parameter τ_j ($0 \leq \tau_j \leq 1$), which provides a connection between the two.

These two motivations (block components with large variance (PCA criteria) vs. correlation with their neighboring components) are opposed but the introduction of

the shrinkage parameters yields a compromise between these two objectives. An algorithm of resolution of optimization problem (1) is described in Algorithm 2.

Algorithm 2 Algorithm for Regularized Generalized Canonical Correlation Analysi with $0 \leq \tau_j \leq 1$

▷ **Step A. Initialisation:**
Choose arbitrary vectors $\tilde{\mathbf{a}}_j^0$ such that (**??**) holds:

$$\mathbf{a}_j^0 = \left[(\tilde{\mathbf{a}}_j^0)^t \left[\tau_j \mathbf{I} + (1-\tau_j) \frac{1}{n} \mathbf{X}_j^t \mathbf{X}_j \right] \tilde{\mathbf{a}}_j^0 \right]^{-1/2} \tilde{\mathbf{a}}_j^0$$

repeat s = 1, 2, ...
  for $j = 1, 2, \ldots, J$ **do**

▷ **Step B. Inner component for $\mathbf{X}_j$**

$$\mathbf{z}_j^s = \sum_{k=1}^{j-1} c_{jk} \mathrm{w}(\mathrm{cov}(\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^{s+1})) \mathbf{X}_k \mathbf{a}_k^{s+1} + \sum_{k=j+1}^{J} c_{jk} \mathrm{w}(\mathrm{cov}(\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^{s+1})) \mathbf{X}_k \mathbf{a}_k^s$$

where $\mathrm{w}(x) = 1$ for the Horst scheme, x for the factorial scheme and $\mathrm{sign}(x)$ for the centroid scheme and and

▷ **Step C. Outer weight for block j:**

$$\mathbf{a}_j^{s+1} = \left[(\mathbf{z}_j^s)^t \mathbf{X}_j \left[\tau_j \mathbf{I} + (1-\tau_j) \frac{1}{n} \mathbf{X}_j^t \mathbf{X}_j \right]^{-1} \mathbf{X}_j^t \mathbf{z}_j^s]^{-1/2} \left[\tau_j \mathbf{I} + (1-\tau_j) \frac{1}{n} \mathbf{X}_j^t \mathbf{X}_j \right]^{-1} \mathbf{X}_j^t \mathbf{z}_j^s$$

  end for
until convergence

To obtain a monotonically convergent algorithm (i.e. the bounded criteria to be maximized increases at each iteration), we use a sequence of operations similar to the one used in the Wold's algorithm for PLSPM. The procedure begins by an arbitrary choice of initial normalized $\mathbf{a}_1^0, \ldots, \mathbf{a}_J^0$ (Step A in Algorithm 2). Suppose outer weight vectors $\mathbf{a}_1^{s+1}, \mathbf{a}_2^{s+1}, \ldots, \mathbf{a}_{j-1}^{s+1}$ are constructed for $\mathbf{X}_1, \ldots, \mathbf{X}_{j-1}$. The outer weight vector $\mathbf{a}_j^{s+1}$ is computed by considering the inner component $\mathbf{z}_j^s$ associated with $\mathbf{X}_j$ given in Step B in Algorithm 2, and the formula given in Step C in Algorithm 2. The procedure is iterated until convergence of the bounded criterion.

4 Comparison between RGCCA and PLS path modeling

First, PLSPM Algorithm (Algorithm 1) and RGCCA algorithm (Algorithm 2) are equivalent when $\tau_j = 0$ for all blocks. Therefore, in this comparison, PLSPM and RGCCA are only compared for $\tau_j = 1$.
As discussed in the introduction, PLSPM-mode A algorithm works very well in

practice but is lacking an analytical proof of convergence and suffers of lack of optimality property (the criterion to which the PLSPM-mode A algorithm converges is unknown). On the other hand, RGCCA is based on a monotonically convergent iterative algorithm and has the distinct advantage to rely on an explicit optimization problem. Moreover, Algorithm 1 and Algorithm 2 are very similar. The only difference between those two algorithms relies on the normalisation of the outer weight vectors. In PLSPM-mode A, the outer components are normalised to unit variance. In RGCCA-mode A, the outer weights are normalized to unit norm. Intuitively, the benefit to standardize the outer components in PLSPM is that each block has the same weight (independently of its number of variables) in the computation of the inner component. In others words, the number of variables in each block has limited influence on the computation of the inner components. At first glance, when inspecting the RGCCA new mode A algorithm, one may think that RGCCA-mode A gives more importance to high dimensional blocks than to ones characterized by few variables. One of the main objective of this comparison is to look at this question.

4.1 Simulation settings

In this simulation RGCCA-mode A and PLSPM-mode A are compared on a simple simulated example

Simulation 1. In this simulation, we consider $J = 3$ blocks, where the $n \times p$ block $\mathbf{X}_j$, $j = 1, \ldots, 3$ is generated according to the following model.

$$\mathbf{X}_j = \mathbf{u}\mathbf{a}_j^t + \mathbf{E}_j, j = 1, \ldots, 3 \tag{2}$$

where $\mathbf{u} \in \mathbb{R}^{50}$ are drawn from a normal distribution with zero mean and unit variance, $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3$ is a vector of ones $\in \mathbb{R}^{10}$ and $\mathbf{E}_j$ is an $n \times p$ noisy matrix where each element is drawn from a normal distribution with zero mean and unit variance. Therefore, for this first simulation $n = 50$, $p_1 = 10$, $p_2 = 10$ and $p_3 = 10$. Each block is strongly unidimensional with associated latent variable $\mathbf{u}$. All blocks are supposed to be connected which means that $c_{13} = c_{12} = c_{23} = 1$. RGCCA-mode A and PLSPM-mode A with centroid scheme (g(x)= $|x|$) are applied on this simulated dataset. Table 1 reports correlations between outer components obtained with RGCCA-mode A and PLSPM-mode A. On this simple simulation, RGCCA-mode

PLS-PM (mode A) vs. RGCCA new mode A	correlation between $\mathbf{y}_j^{PLS}$ and $\mathbf{y}_j^{RGCCA}$ with $\mathbf{u}$
$\text{cor}(\mathbf{y}_1^{\text{PLS}}, \mathbf{y}_1^{\text{RGCCA}}) = 1$	$\text{cor}(\mathbf{y}_1^{\text{PLS}}, \mathbf{u}) = \text{cor}(\mathbf{y}_1^{\text{RGCCA}}, \mathbf{u}) = 1$
$\text{cor}(\mathbf{y}_2^{\text{PLS}}, \mathbf{y}_2^{\text{RGCCA}}) = 1$	$\text{cor}(\mathbf{y}_2^{\text{PLS}}, \mathbf{u}) = \text{cor}(\mathbf{y}_2^{\text{RGCCA}}, \mathbf{u}) = 1$
$\text{cor}(\mathbf{y}_3^{\text{PLS}}, \mathbf{y}_3^{\text{RGCCA}}) = 1$	$\text{cor}(\mathbf{y}_3^{\text{PLS}}, \mathbf{u}) = \text{cor}(\mathbf{y}_3^{\text{RGCCA}}, \mathbf{u}) = 1$

Table 1 Simulation 1. Correlations between outer components obtained from RGCCA new mode A and PLS path modeling mode A.

A and PLSPM-mode A give the same outer components.

Simulation 2. The next simulation was designed to evaluate the sensitivity of both RGCCA-mode A and PLSPM-mode A to unbalanced dimensional block settings. For that purpose, the following simulated example is considered

$$\mathbf{X}_1^* = [\mathbf{X}_1; \mathbf{H}_1 + \mathbf{F}_1] \tag{3}$$

where $\mathbf{X}_1 = \mathbf{u}\mathbf{a}_1^t + \mathbf{E}_1$, the kth column of $\mathbf{H}_1$ is defined by $\mathbf{h}_{1k} = (1-\alpha)\mathbf{u} + \alpha\mathbf{v}$ with $\mathbf{v}$ is a vector orthgonal to each column of $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ and $\mathbf{F}_1$ is a noisy matrix. We are interested in studying relationships between $\mathbf{X}_1^*$ and unchanged $\mathbf{X}_2$ and $\mathbf{X}_3$. as a function of the number of additional variable and the value of α.

- For $\alpha = 0$ additional variable $\mathbf{h}_{1k}$ are closed to $\mathbf{u}$ which implies that no additional information is added. The first principal component of $\mathbf{X}_1^*$ is very close to $\mathbf{u}$. Results for RGCCA-mode A and PLS-PM mode A is expected to be unchanged whatever the number of additional variables.
- For $\alpha = 1$ additional variable $\mathbf{h}_{1k}$ are orthogonal to each column of $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$. $\mathbf{h}_{1k}$ does not influence the construction of the first outer component of each block. We may however mention that for a large number of addtional variables, the first principal component of $\mathbf{X}_1^*$ is very close to $\mathbf{v}$. Results of RGCCA-mode A and PLSPM-mode A are expected to be unchanged because $\mathbf{v}$ is orthogonal to each column of $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ and thus strongly orthogonal to $\mathbf{u}$.
- For $0 < \alpha < 1$ additional variable $\mathbf{h}_{1k}$ are a mix between $\mathbf{u}$ and $\mathbf{v}$ and may attract the first latent variable in term of correlation and variance. Due to the criteria on which RGCCA-mode A is based, RGCCA is maybe more attracted by the additional variables when α is close from 0.5 than PLSPM-mode A.

Results for RGCCA-mode A and PLSPM-mode A are reported in Figure 1 and 2. From figure 1 is observed that RGCCA-mode A and PLSPM-mode A yield very close results and seem to be sensitive to additional variables weakly correlated to the latent variable. Moreover, from figure 2 we can note that the outer components computed with RGCCA-mode A and PLSPM-mode A are "attracted" by the first principal components which means that the variance terms weight to much compared to the term of correlation.

5 Conclusion

On these two simulations and on all the real datasets we analyzed, we have failed to find situations where RGCCA-mode A and PLSPM-mode A exhibit different behaviors. Empirically we have noted that both RGCCA-mode A outer components and PLSPM-mode A outer components are very close to the first principal components which means that the variance terms weight to much compared to the term of correlation. If a compromise (block components with large variance (PCA criteria)

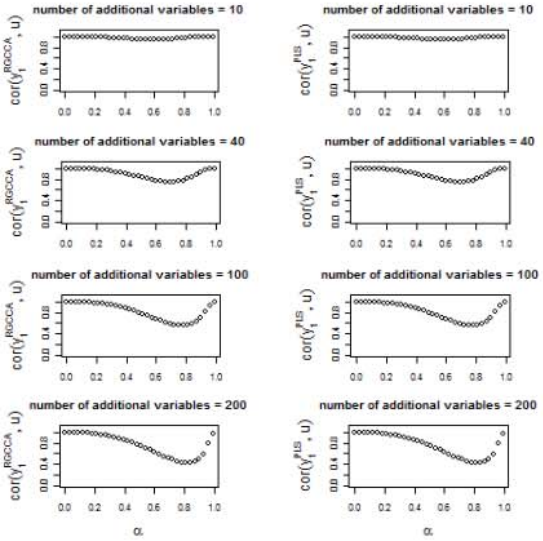

Fig. 1 The number of additional variables vary from 10 to 200. Correlations between $\mathbf{y}_1^{\text{PLS}}$ and $\mathbf{y}_1^{\text{RGCCA}}$ with $\mathbf{u}$ as a function of α

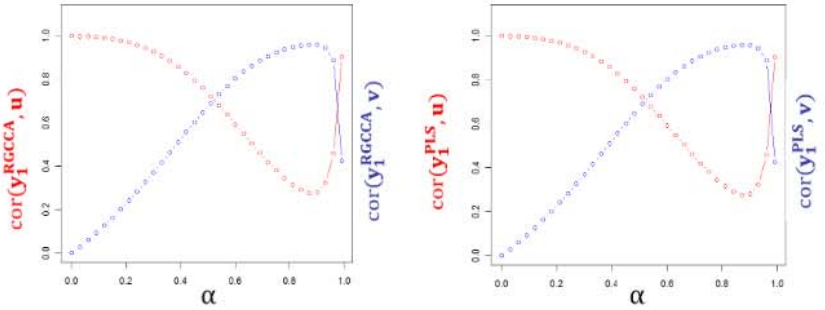

Fig. 2 The number of additional variables is set to 500. Correlation between $\mathbf{y}_1^{\text{PLS}}$ and $\mathbf{y}_1^{\text{RGCCA}}$ with $\mathbf{u}$ and $\mathbf{y}_1^{\text{PLS}}$ and $\mathbf{y}_1^{\text{RGCCA}}$ with $\mathbf{v}$ as a function of α

vs. correlation with their neighboring components) is desired the RGCCA-mode ridge ($0 \leq \tau_j \leq 1$) is an attractive solution.

References

1. Krämer N. (2007) : Analysis of high-dimensional data with partial least squares and boosting. Doctoral dissertation, Technischen Universitt Berlin.
2. Tenenhaus M., Esposito Vinzi V., Chatelin Y.-M., Lauro C. (2005) : PLS path modeling. Computational Statistics & Data Analysis, 48, 159-205.
3. Tenenhaus A. and Tenenhaus M. (2011) : Regularized generalized canonical correlation analysis. Psychometrika (paratre).
4. Wold H. (1982) : Soft Modeling: The Basic Design and Some Extensions, in Systems under indirect observation, Part 2, K.G. Jreskog & H. Wold (Eds), North-Holland, Amsterdam, pp. 1-54.
5. Wold H. (1985) : Partial Least Squares, in Encyclopedia of Statistical Sciences, vol. 6, Kotz, S & Johnson, N.L. (Eds), John Wiley & Sons, New York, pp. 581-591.
6. Lohmöller, J.B., Latent Variables Path Modeling with Partial Least Squares. Physica-Verlag, Heildelberg, 1989.