Regularized generalized canonical correlation analysis

Arthur Tenenhaus, Michel Tenenhaus

Regularized generalized canonical correlation analysis

Abstract

Regularized generalized canonical correlation analysis (RGCCA) is a generalization of regularized canonical correlation analysis to three or more sets of variables. It constitutes a general framework for many multi-block data analysis methods. It combines the power of multi-block data analysis methods (maximization of well identified criteria) and the flexibility of PLS path modeling (the researcher decides which blocks are connected and which are not). Searching for a fixed point of the stationary equations related to RGCCA, a new monotonically convergent algorithm, very similar to the PLS algorithm proposed by Herman Wold, is obtained. Finally, a practical example is discussed.

Keywords: Generalized canonical correlation analysis, Multi-block data analysis, PLS path modeling, Regularized canonical correlation analysis.

I. Introduction

The main purpose of this paper is to combine the power of multi-block data analysis methods and the flexibility of the Partial Least Squares (PLS) path modeling algorithms proposed by Wold (1985), Lohmöller (1989) and Krämer (2007). These methods deal with the same kind of data and share the same objectives: how to relate several blocks of variables observed on the same set of individuals. The power of multi-block data analysis lies in the fact that it includes a great variety of methods with well identified criteria to be optimized. The great flexibility of PLS path modeling lies in the possibility of taking into account certain hypotheses on connections between blocks: the researcher decides which blocks are connected and which are not. Unfortunately, the criteria optimized by the various options of PLS path modeling algorithms are often unclear. In this paper, we propose a new method called Regularized Generalized Canonical Correlation Analysis (RGCCA). This is a generalization of regularized canonical correlation analysis (Vinod (1976), Leurgans, Moyeed and Silverman (1993)) to three or more sets of variables. RGCCA combines the power of multi-block data analysis methods and the flexibility of PLS path modeling.

RGCCA is a framework for modeling linear relationships between several blocks of variables observed on the same set of individuals. Considering a network of connections between these blocks, the objective of RGCCA is to find linear combinations of block variables (block components) such that (i) block components explain their own block well and/or (ii) block components that are assumed to be connected are highly correlated. In the literature, many methods exist with the objective of finding block components having these properties. Some of them are based on the maximization of a function of correlations: SUMCOR (sum of correlations method), SSQCOR (sum of squared correlations method), SABSCOR (sum of absolute values of the correlations method). Others are based on the maximization of a function of covariances: SUMCOV (sum of covariances method), SSQCOV (sum of squared covariances method), SABSCOV (sum of absolute values of the covariances method). Others are based on the maximization of a function of

both correlations and covariances. It appears that RGCCA constitutes a general framework for multi-block data analysis and includes all the previously listed methods as particular cases. Moreover, one of the important properties of RGCCA relies on the fact that blocks are not necessarily fully connected. This flexibility allows a large variety of hierarchical models such as Carroll's (1968) generalized canonical correlation analysis, Chessel & Hanafi's (1996) Multiple Co-Inertia Analysis and those given in Westerhuis, Kourti & MacGregor (1998) and Vivien & Sabatier (2003) to be included in the RGCCA framework. RGCCA also contains the PLS approach of Wold-Lohmöller-Krämer as a special case, but only when the option "mode B for all blocks" is selected. The full comparison between PLS path modeling and RGCCA is, however, beyond the scope of this paper and will be discussed in a separate paper.

Moreover, in a high-dimensional block setting or in the presence of multicollinearity within blocks (ill-conditioned data blocks), correlation-based methods lead to spurious relationships between blocks. This gives an impression of links between blocks that are invalid when objectively examined. RGCCA constitutes a regularized version of various correlation-based methods and makes a stable analysis of ill-conditioned data blocks possible. It allows for the introduction of a continuum between correlation-based criteria and covariance-based criteria.

There is no analytical solution to RGCCA, so we propose a monotonically convergent algorithm based on a modification of Wold's PLS algorithm (1985). RGCCA is obtained by following several steps. Firstly, a new generalized canonical correlation analysis (GCCA) method, which takes into account the hypotheses on the connections between sets of variables, is defined at the population level. Secondly, using shrinkage estimates for block covariance matrices, stationary equations related to population GCCA are written at the sample level. Thirdly, these stationary equations are related to a new optimization problem called RGCCA. We have obtained a new monotonically convergent algorithm when searching for a fixed point of these stationary equations.

Finally, we will conclude this paper with a detailed analysis of a practical example where many of the RGCCA possibilities are explored. We will also show how the PLS path modeling drawing conventions can be extended to multi-block data analysis and RGCCA.

This paper is organized as follows:

(1) Definition of population generalized canonical correlation analysis. Construction of the stationary equations at the population level. Searching for a fixed point of the population level stationary equations: construction of a PLS algorithm for population GCCA. Convergence properties of the PLS algorithm for population GCCA.

(2) Construction of the stationary equations at the sample level, using shrinkage estimates of the block covariance matrices.

(3) Definition of the regularized generalized canonical correlation analysis (RGCCA).

(4) Searching for a fixed point of the sample level stationary equations: construction of a PLS algorithm for RGCCA. Convergence properties of the PLS algorithm for RGCCA.

(5) The Russett example.

II Population generalized canonical correlation analysis

Canonical correlation analysis of two sets of random variables can be extended to three or more sets of variables in many ways. For instance, Kettenring (1971) studied five specific methods: (i) the sum of correlations method (SUMCOR), (ii) the maximum variance method (MAXVAR), (iii) the sum of squared correlations method (SSQCOR), (iv) the minimum variance method (MINVAR), (v) the generalized variance method (GENVAR). A sixth method, the sum of absolute value correlations method (SABSCOR), is also considered in this paper. This last criterion plays a central role in the PLS approach of Wold (1985).

We propose a modification of the SUMCOR, SSQCOR and SABSCOR methods which takes into account some hypotheses on the connections between the sets of variables. Let's consider J

random p_j - dimensional zero mean column vectors $\boldsymbol{x}_j = (x_{j1},...,x_{jp_j})^t$ defined on the same population and J non-random p_j - dimensional column vectors $\boldsymbol{\alpha}_j = (\alpha_{j1},...,\alpha_{jp_j})^t$. We also consider a network of connections between the random vectors by defining a design matrix $\mathbf{C} = \{c_{jk}\}$: $c_{jk} = 1$ if $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ are connected and 0 otherwise.

Consider two linear components $\eta_j = \sum_h \alpha_{jh} x_{jh} = \boldsymbol{\alpha}_j^t \boldsymbol{x}_j$ and $\eta_k = \sum_h \alpha_{kh} x_{kh} = \boldsymbol{\alpha}_k^t \boldsymbol{x}_k$. The correlation between the two random variables η_j and η_k is

$$\rho(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k) = \frac{\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k}{(\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jj} \boldsymbol{\alpha}_j)^{1/2} (\boldsymbol{\alpha}_k^t \boldsymbol{\Sigma}_{kk} \boldsymbol{\alpha}_k)^{1/2}} \tag{1}$$

where $\boldsymbol{\Sigma}_{jj} = E(\boldsymbol{x}_j \boldsymbol{x}_j^t)$, $\boldsymbol{\Sigma}_{kk} = E(\boldsymbol{x}_k \boldsymbol{x}_k^t)$ and $\boldsymbol{\Sigma}_{jk} = E(\boldsymbol{x}_j \boldsymbol{x}_k^t)$. All $\boldsymbol{\Sigma}_{jj}$ are supposed to be of full rank.

II.1 Definition of Population GCCA

We define "population generalized canonical correlation analysis" as the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_J}{\text{Maximize}} \quad \sum_{j,k=1,\, j \neq k}^{J} c_{jk}\, g(\rho(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)) \\ &\text{subject to the constraints } \operatorname{Var}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j) = 1, \;\; j = 1,...,J \end{aligned} \tag{2}$$

where g is the identity, the absolute value or the square function. As the algorithm proposed in this paper for solving optimization problem (2) is close to Wold's PLS algorithm (1985), we feel that it would be useful and necessary to introduce the PLS terminology here. The centroid scheme introduced by Wold (1985) is related to g being equal to the absolute value. The factorial scheme proposed by Lohmöller (1989) is related to g being equal to the square. The Horst scheme proposed by Krämer (2007) is related to g being equal to the identity. We have limited the optimization problem (2) to these three schemes because they are the most used in the multi-block and PLS path modeling literature. The centroid and factorial schemes usually give close results in practical applications (Noonan and Wold, 1982). The Horst scheme can be very useful when the researcher is

looking for positively correlated components. Due to this constraint, the Horst scheme can yield quite different solutions from the two other schemes.

Problem (2) is equivalent to the following optimization problem:

$$\text{(3)} \qquad \begin{array}{l} \underset{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_J}{\text{Maximize}} \sum_{j,k=1,\, j\neq k}^{J} c_{jk}\, g(\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k) \\ \text{subject to the constraints } \boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jj} \boldsymbol{\alpha}_j = 1, \;\; j = 1,\ldots,J \end{array}$$

The following Lagrangian function of optimization problem (3) is then considered:

$$\text{(4)} \qquad F(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_J,\lambda_1,\ldots,\lambda_J) = \sum_{j,k=1,\, j\neq k}^{J} c_{jk}\, g(\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k) - \varphi \sum_{j=1}^{J} \frac{\lambda_j}{2} (\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jj} \boldsymbol{\alpha}_j - 1)$$

where $\varphi = 1$ when g is the identity or the absolute value and 2 when g is the square function. We may suppose that $\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k$ is different from 0, because if it were not the case, we would just set the design coefficient c_{jk} to zero. Therefore, we may also consider the derivative g' when g is the absolute value. Cancelling the derivatives of the Lagrangian function with respect to $\boldsymbol{\alpha}_j$ and λ_j yields the following stationary equations for population GCCA:

$$\text{(5)} \qquad \frac{1}{\varphi} \boldsymbol{\Sigma}_{jj}^{-1} \sum_{k=1,k\neq j}^{J} c_{jk}\, g'(\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k) \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k = \lambda_j \boldsymbol{\alpha}_j \quad , \quad j = 1,\ldots,J$$

with the normalization constraints

$$\text{(6)} \qquad \boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jj} \boldsymbol{\alpha}_j = 1, \;\; j = 1,\ldots,J$$

These stationary equations have no analytical solution, but they can be used to build a monotonically convergent algorithm for optimization problem (3). This new algorithm, to be described in the next section, appears to be very similar to the Wold's PLS algorithm (1985).

II.2 A PLS algorithm for population GCCA

In the sequel to this paper, we use the expression $\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)$ instead of $\boldsymbol{\alpha}_j^t \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k$ for easier mathematical treatments and the function $w(x) = \frac{1}{\varphi} g'(x)$ for easier readability. The function w(x) is equal to 1 for the Horst scheme, to x for the factorial scheme and to sign(x) for the centroid

scheme. In PLS terminology, $\boldsymbol{\alpha}_j$ is a vector of outer weights, $\eta_j = \boldsymbol{\alpha}_j^t \boldsymbol{x}_j$ is called an outer component and an inner component ν_j is defined as follows:

$$\nu_j = \sum_{k=1,k\neq j}^{J} c_{jk} w(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)) \boldsymbol{\alpha}_k^t \boldsymbol{x}_k \tag{7}$$

For the various schemes, we get the following inner components:

- Horst (g = identity): $\nu_j = \sum_{k=1,k\neq j}^{J} c_{jk} \boldsymbol{\alpha}_k^t \boldsymbol{x}_k$

- Factorial (g = square): $\nu_j = \sum_{k=1,k\neq j}^{J} c_{jk} \mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k) \boldsymbol{\alpha}_k^t \boldsymbol{x}_k$

- Centroid (g = absolute value): $\nu_j = \sum_{k=1,k\neq j}^{J} c_{jk} \mathrm{sign}\left[\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)\right] \boldsymbol{\alpha}_k^t \boldsymbol{x}_k$

The inner components ν_j are useful to simplify the stationary equation for population GCCA (5). Using the following expression

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{x}_j, \nu_j) &= \mathrm{E}(\boldsymbol{x}_j \nu_j) = \mathrm{E}(\boldsymbol{x}_j \sum_{k=1,k\neq j}^{J} c_{jk} w(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)) \boldsymbol{\alpha}_k^t \boldsymbol{x}_k) \\ &= \sum_{k=1,k\neq j}^{J} c_{jk} w(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k) \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k = \frac{1}{\varphi} \sum_{k=1,k\neq j}^{J} c_{jk} g'(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k) \boldsymbol{\Sigma}_{jk} \boldsymbol{\alpha}_k \end{aligned} \tag{8}$$

and the normalization constraints (6), the stationary equations (5) become

$$\boldsymbol{\alpha}_j = \left[\mathrm{Cov}(\boldsymbol{x}_j, \nu_j)^t \boldsymbol{\Sigma}_{jj}^{-1} \mathrm{Cov}(\boldsymbol{x}_j, \nu_j)\right]^{-1/2} \boldsymbol{\Sigma}_{jj}^{-1} \mathrm{Cov}(\boldsymbol{x}_j, \nu_j) \quad , \quad j = 1, ..., J \tag{9}$$

It is worth noting that the stationary equations (9) are also obtained by considering the minimum of $\mathrm{E}\left[(\nu_j - \boldsymbol{\beta}_j^t \boldsymbol{x}_j)^2\right]$ with respect to $\boldsymbol{\beta}_j$, subject to the normalization constraint $\boldsymbol{\beta}_j^t \boldsymbol{\Sigma}_{jj} \boldsymbol{\beta}_j = 1$ (or equivalently, the maximum of $\mathrm{Cov}(\boldsymbol{\beta}_j^t \boldsymbol{x}_j, \nu_j)$, subject to the same normalisation constraint). This result is central to the proof of Proposition 2 below.

While the notion of inner component is not usually found in the multi-block literature, it plays a central role in the PLS approach and allows us to design a very efficient algorithm for

population GCCA. The following proposition specifies the role of the inner components in the criterion to be maximized.

Proposition 1

For g equal to the identity, equal to the square or equal to the absolute value, we obtain the following result:

$$(10) \qquad \sum_{j,k=1,k\neq j}^{J} c_{jk} g\left[\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)\right] = \sum_{j=1}^{J} \mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \nu_j)$$

Proof

Equality (10) is obtained from the identity $w(x)x = g(x)$ for g equal to the identity, to the square or to the absolute value:

$$\sum_{j=1}^{J} \mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \nu_j) = \sum_{j=1}^{J} \mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \sum_{k=1,k\neq j}^{J} c_{jk} w(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)) \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)$$

$$= \sum_{j,k=1,k\neq j}^{J} c_{jk} w(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)) \mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k) = \sum_{j,k=1,k\neq j}^{J} c_{jk} g(\mathrm{Cov}(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k))$$

□

It is possible to construct a monotonically convergent algorithm related to optimization problem (3). That is to say, that the bounded criterion to be maximized in (3) is increasing at each step of the proposed iterative procedure. Stationary equations (9) and proposition 1 suggest an iterative algorithm for optimization problem (3):

- Begin with arbitrary normalized outer weights $\boldsymbol{\alpha}_j,\ j = 1,\ldots,J$
- Compute the inner components $\nu_j,\ j = 1,\ldots,J$, according to formula (7)
- Compute new normalized outer weights using formula (9)
- Iterate this procedure

To obtain a monotonically convergent algorithm, it is necessary to use a sequence of operations similar to the ones used by Wold (1985) and Hanafi (2007) for PLS path modeling. This PLS algorithm for population GCCA is described in Figure 1.

Insert Figure 1 approximately here

The procedure begins by an arbitrary choice of initial normalized outer weight vectors $\boldsymbol{\alpha}_1^0, \boldsymbol{\alpha}_2^0, ..., \boldsymbol{\alpha}_J^0$ (step A in Figure 1). Suppose outer weight vectors $\boldsymbol{\alpha}_1^{s+1}, \boldsymbol{\alpha}_2^{s+1}, ..., \boldsymbol{\alpha}_{j-1}^{s+1}$ have been constructed. The outer weight vector $\boldsymbol{\alpha}_j^{s+1}$ is computed by considering the inner component v_j^s given in step B in Figure 1, and the formula given in step C in Figure 1. The procedure is iterated until convergence of the bounded criterion, which is due to the following proposition:

Proposition 2: Let $\boldsymbol{\alpha}_j^s$, $j = 1, ..., J$, $s = 0, 1, 2, ...$, be a sequence of outer weight vectors generated by the PLS algorithm for population GCCA. The following function is defined for outer weight vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_J$:

$$(11) \qquad f(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_J) = \sum_{j,k=1, j\neq k}^{J} c_{jk} g\left[\rho(\boldsymbol{\alpha}_j^t \boldsymbol{x}_j, \boldsymbol{\alpha}_k^t \boldsymbol{x}_k)\right]$$

The following inequalities hold:

$$(12) \qquad \forall\ s \quad f(\boldsymbol{\alpha}_1^s, \boldsymbol{\alpha}_2^s, ..., \boldsymbol{\alpha}_J^s) \leq f(\boldsymbol{\alpha}_1^{s+1}, \boldsymbol{\alpha}_2^{s+1}, ..., \boldsymbol{\alpha}_J^{s+1})$$

Proof: See Appendix 1.

The essential feature of this algorithm is that each replacement is optimal, and sequential, that is to say that $\boldsymbol{\alpha}_j^s$ must be replaced by $\boldsymbol{\alpha}_j^{s+1}$ before replacing $\boldsymbol{\alpha}_{j+1}^s$. This is the essence of the Gauss-Seidel algorithm for solving a system of linear equations and of several other iterative algorithms such as the Wold and Hanafi PLS algorithms, the MAXDIFF algorithm (Ten Berge, 1988) and the multivariate eigenvalue problem (Chu & Watterson, 1993). This sequential approach leads to the monotonic convergence of these algorithms.

Because the main objective is to estimate the outer weights $\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_J$ from a finite sample, it is now necessary to write the population stationary equations at the sample level.

III. Stationary equations at the sample level using shrinkage estimates of the covariance matrices $\boldsymbol{\Sigma}_{jj}$

Let's consider J blocks $\mathbf{X}_1,...,\mathbf{X}_J$ of centered variables measured on a set of n individuals. A row of $\mathbf{X}_j$ represents a realization of the row-random vector $\boldsymbol{x}_j^t$, a column $\mathbf{x}_{jh}$ of $\mathbf{X}_j$ is considered as a variable observed on n individuals, x_{jhi} is the value of variable $\mathbf{x}_{jh}$ for individual *i*. Now, $\mathbf{C} = \{c_{jk}\}$ is a design matrix describing a network of relationships between blocks: $c_{jk} = 1$ for two connected blocks, and 0 otherwise.

The sample covariance matrices are $\mathbf{S}_{jj} = \frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j$ and $\mathbf{S}_{jk} = \frac{1}{n}\mathbf{X}_j^t\mathbf{X}_k$. In case of high multicollinearity or when the number of individuals is smaller than the number of variables, the sample covariance matrix $\mathbf{S}_{jj}$ is a poor estimate of the true covariance matrix $\boldsymbol{\Sigma}_{jj}$ for a very simple reason pointed out by Schäfer and Strimmer (2005): the number of eigenvalues close or equal to zero is much higher in the sample covariance matrix than in the true covariance matrix. One suggestion of Ledoit and Wolf (2004) for finding a better estimate of the true covariance matrix $\boldsymbol{\Sigma}_{jj}$ is to consider the class of linear combinations $\left\{\hat{\mathbf{S}}_{jj} = \tau_j\mathbf{I} + (1-\tau_j)\mathbf{S}_{jj},\ \ 0 \le \tau_j \le 1\right\}$ of the identity matrix $\mathbf{I}$ and the sample covariance matrix $\mathbf{S}_{jj}$. The matrix $\hat{\mathbf{S}}_{jj}$ is called a shrinkage estimate of $\boldsymbol{\Sigma}_{jj}$ and τ_j is the shrinkage constant. Furthermore, Schäfer and Strimmer, using the unbiased estimate $\mathbf{S}'_{jj} = \frac{n}{n-1}\mathbf{S}_{jj}$ of $\boldsymbol{\Sigma}_{jj}$ instead of $\mathbf{S}_{jj}$, give a formula for finding an optimal shrinkage constant, i.e., minimizing the mean square error $\mathrm{MSE} = \mathrm{E}\left[\left\|\hat{\mathbf{S}}_{jj} - \boldsymbol{\Sigma}_{jj}\right\|^2\right]$, where $\|\cdot\|$ is the Frobenius norm (formula 6, p. 8 of Schäfer and Strimmer (2005)). In the same reference, they also give a formula for estimating this optimal shrinkage constant from the data (formula for Target A, p.11 and appendix A):

$$\hat{\tau}_j^* = \frac{\sum_{k\neq l=1}^{p_j} \widehat{\mathrm{Var}}(s'_{j,kl}) + \sum_{k=1}^{p_j} \widehat{\mathrm{Var}}(s'_{j,kk})}{\sum_{k\neq l=1}^{p_j} (s'_{j,kl})^2 + \sum_{k=1}^{p_j} (s'_{j,kl} - 1)^2}$$

where $s'_{j,kl}$ is an entry of $\mathbf{S}'_{jj}$ and, setting $w_{jkli} = (x_{jki} - \bar{x}_{jk})(x_{jli} - \bar{x}_{jl})$, $\widehat{\mathrm{Var}}(s'_{j,kl}) = \frac{n}{(n-1)^3} \sum_{i=1}^{n} \left(w_{jkli} - \bar{w}_{jkl} \right)^2$ is the empirical unbiased variance of $s'_{j,kl}$.

We also introduce for each block an outer weight vector $\mathbf{a}_j$, an outer component $\mathbf{y}_j = \mathbf{X}_j \mathbf{a}_j$ and an inner component $\mathbf{z}_j$ defined as

$$\mathbf{z}_j = \sum_{k=1, k\neq j}^{J} c_{jk} w\left[\mathrm{Cov}(\mathbf{y}_j, \mathbf{y}_k)\right] \mathbf{y}_k \tag{13}$$

These vectors $\mathbf{a}_j, \mathbf{y}_j, \mathbf{z}_j$ are the sample versions of the previously defined $\boldsymbol{\alpha}_j, \eta_j, \nu_j$. For the Horst scheme, the inner component $\mathbf{z}_j = \sum_{k=1, k\neq j}^{J} c_{jk} \mathbf{y}_k$, for the factorial one $\mathbf{z}_j = \sum_{k=1, k\neq j}^{J} c_{jk} \left[\mathrm{Cov}(\mathbf{y}_j, \mathbf{y}_k)\right] \mathbf{y}_k$ and for the centroid one $\mathbf{z}_j = \sum_{k=1, k\neq j}^{J} c_{jk} \mathrm{sign}\left[\mathrm{Cov}(\mathbf{y}_j, \mathbf{y}_k)\right] \mathbf{y}_k$. For easier readability, we prefer to keep the same notation for the population covariance and the sample covariance.

In this section, we consider a sample version of the stationary equations (9) where $\boldsymbol{\Sigma}_{jj}$ is replaced by $\hat{\mathbf{S}}_{jj} = \tau_j \mathbf{I} + (1-\tau_j)\mathbf{S}_{jj}$. This leads to the following sample stationary equations

$$\mathbf{a}_j = \left[\mathbf{z}_j^t \mathbf{X}_j \left[\tau_j \mathbf{I} + (1-\tau_j) \frac{1}{n} \mathbf{X}_j^t \mathbf{X}_j \right]^{-1} \mathbf{X}_j^t \mathbf{z}_j \right]^{-1/2} \left[\tau_j \mathbf{I} + (1-\tau_j) \frac{1}{n} \mathbf{X}_j^t \mathbf{X}_j \right]^{-1} \mathbf{X}_j^t \mathbf{z}_j, \quad j = 1, \ldots, J \tag{14}$$

and to normalization constraints on the outer weight vectors $\mathbf{a}_j$ which depend upon the values of the shrinkage constants τ_j:

$$\mathbf{a}_j^t \left[\tau_j \mathbf{I} + (1-\tau_j) \frac{1}{n} \mathbf{X}_j^t \mathbf{X}_j \right] \mathbf{a}_j = 1, \; j = 1, \ldots, J \tag{15}$$

Let's introduce some useful terminology inspired by the PLS approach: the situation corresponding to $\tau_j = 0$ is called "mode B", the one corresponding to $\tau_j = 1$ is called "new mode A" and the case $0 < \tau_j < 1$ is called "mode Ridge". Let's have a closer look at these three modes.

Mode B ($\tau_j = 0$)

For $\tau_j = 0$, the normalization constraint is $\mathrm{Var}(\mathbf{X}_j \mathbf{a}_j) = 1$ and the stationary equation (14) becomes

$$\mathbf{a}_j = n^{1/2} \left[\mathbf{z}_j^t \mathbf{X}_j \left(\mathbf{X}_j^t \mathbf{X}_j \right)^{-1} \mathbf{X}_j^t \mathbf{z}_j \right]^{-1/2} \left(\mathbf{X}_j^t \mathbf{X}_j \right)^{-1} \mathbf{X}_j^t \mathbf{z}_j \tag{16}$$

This vector of outer weights $\mathbf{a}_j$ is proportional to the vector of the regression coefficients in the multiple regression of $\mathbf{z}_j$ on $\mathbf{X}_j$. It is worth pointing out that due to the inversion of the intra-block covariance matrices, this way of computing the outer weight vector cannot be applied to an ill-conditioned block. This way of computing the outer weights is the usual mode B of the PLS approach.

New mode A ($\tau_j = 1$)

For $\tau_j = 1$, the normalization constraint becomes $\|\mathbf{a}_j\| = 1$ and the stationary equation (14) is written as

$$\mathbf{a}_j = \mathbf{X}_j^t \mathbf{z}_j \,/\, \|\mathbf{X}_j^t \mathbf{z}_j\| \tag{17}$$

We may note that the outer component $\mathbf{y}_j = \mathbf{X}_j \mathbf{a}_j$ is the first PLS component in the PLS regression (Wold, Martens & Wold, 1983) of the inner component $\mathbf{z}_j$ on block $\mathbf{X}_j$. In this paper, this way of computing the outer weights is called "new mode A". In the original mode A of the PLS approach, the outer weights are computed in the same way as formula (17), but normalized so that the outer component $\mathbf{y}_j = \mathbf{X}_j \mathbf{a}_j$ is standardized. This "new mode A" shrinks the intra-block covariance

matrix to the identity. This shrinkage is probably too strong, but is useful for very high dimensional data because it avoids the inversion of the intra-block covariance matrix.

Mode Ridge ($0 < \tau_j < 1$)

For $0 < \tau_j < 1$, (14) may also be written as:

$$(18) \qquad \mathbf{a}_j = \frac{n}{1-\tau_j}\left[\mathbf{z}_j^t\mathbf{X}_j\left[\tau_j\mathbf{I} + (1-\tau_j)\frac{1}{n}\mathbf{X}_j^t\mathbf{X}_j\right]^{-1}\mathbf{X}_j^t\mathbf{z}_j\right]^{-1/2}\left[(\mathbf{X}_j^t\mathbf{X}_j + \frac{n\tau_j}{1-\tau_j}\mathbf{I})\right]^{-1}\mathbf{X}_j^t\mathbf{z}_j$$

The vector of outer weights $\mathbf{a}_j$ is proportional to the vector of the regression coefficients in the ridge regression of $\mathbf{z}_j$ on $\mathbf{X}_j$ with a ridge constant equal to $n\tau_j/(1-\tau_j)$. We call this new way of computing the outer weights "mode Ridge". This "mode Ridge" allows a gradual shrinkage of the intra-block covariance matrix towards the identity.

Three interesting properties of the outer component $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$ are established in Qannari & Hanafi (2005):

a) $\mathrm{Var}(\mathbf{X}_j\mathbf{a}_j / \|\mathbf{a}_j\|)$ is an increasing function of τ_j.

b) $\mathrm{Cov}(\mathbf{X}_j\mathbf{a}_j / \|\mathbf{a}_j\|, \mathbf{z}_j)$ is an increasing function of τ_j.

c) $\mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{z}_j)$ is a decreasing function of τ_j.

These properties serve as a useful guide to choosing the shrinkage constant τ_j. They show that, for outer weights $\mathbf{a}_j$ defined in (14), an increase of $\mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{z}_j)$ is balanced by a decrease of $\mathrm{Var}(\mathbf{X}_j\mathbf{a}_j / \|\mathbf{a}_j\|)$. Furthermore, property (b) shows that the variance term dominates over the correlation term. Therefore, if the user is favoring stability (high variance) compared to correlation, $\tau_j = 1$ is the natural choice. If the user wants to give priority to the correlation between $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$ and its neighboring components, property (c) shows that $\tau_j = 0$ is the best choice. For a compromise between variance and correlation, the shrinkage constant τ_j can be determined by using the Schäfer & Strimmer's formula. This automatic estimation of the shrinkage constant allows one

to come closer to the correlation criterion, even in the case of high multicollinearity or when the number of individuals is smaller than the number of variables. In PLS path modeling, Fornell and Bookstein (1982) propose to favor the variance when a block is reflective (block variables reflect some underlying unobservable variable) and to favor the correlation when a block is formative (block variables are used to produce a score).

IV Regularized generalized canonical correlation analysis (RGCCA)

By considering the J data blocks $\mathbf{X}_1,...,\mathbf{X}_J$, the design matrix $\mathbf{C}=\{c_{jk}\}$, the function g and the shrinkage constants $\tau_1,...,\tau_J$ described in section III, we define regularized generalized canonical correlation analysis (RGCCA) as the following optimization problem:

$$\text{(19)} \qquad \begin{aligned} &\underset{\mathbf{a}_1,...,\mathbf{a}_J}{\text{Maximize}} \sum_{j,k=1,\, j\neq k}^{J} c_{jk}\, g(\text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)) \\ &\text{subject to the constraints } \tau_j \|\mathbf{a}_j\|^2 + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1, \;\; j=1,...,J \end{aligned}$$

The stationary equations obtained by cancelling the derivatives of the Lagrangian function related to optimization problem (19) are exactly the stationary equations (14).

The PLS algorithm for RGCCA

There is no known analytical solution to optimization problem (19). However, writing the PLS algorithm for population GCCA at the sample level straightforwardly yields a monotonically convergent algorithm. This algorithm is described in Figure 2.

Insert Figure 2 approximately here

The procedure begins by an arbitrary choice of initial values $\mathbf{a}_1^0, \mathbf{a}_2^0, ..., \mathbf{a}_J^0$ (step A in Figure 2). Suppose outer weight vectors $\mathbf{a}_1^{s+1}, \mathbf{a}_2^{s+1}, ..., \mathbf{a}_{j-1}^{s+1}$ are constructed for blocks $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{j-1}$. The outer weight vector $\mathbf{a}_j^{s+1}$ is computed by considering the inner component $\mathbf{z}_j^s$ for block $\mathbf{X}_j$ given in step B in Figure 2, and the formula given in step C in Figure 2. The procedure is iterated until convergence of the bounded criterion, which is due to proposition 3 given below.

Proposition 3: Let $\mathbf{a}_j^s$, $j = 1,...,J$, $s = 0,1,2,...$, be a sequence of outer weight vectors generated by the PLS algorithm for RGCCA. The following function is defined:

$$h(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_J) = \sum_{j,k=1, j\neq k}^{J} c_{jk} g\left[\mathrm{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)\right] \quad (20)$$

for vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_J$ verifying normalization constraints (15). The following inequalities hold:

$$\forall\ s \quad h(\mathbf{a}_1^s, \mathbf{a}_2^s, ..., \mathbf{a}_J^s) \leq h(\mathbf{a}_1^{s+1}, \mathbf{a}_2^{s+1}, ..., \mathbf{a}_J^{s+1}) \quad (21)$$

Proof: The proof is similar to that of proposition 2.

The proposed algorithm has two limitations:

1) There is no proof that the algorithm converges towards a fixed point of the stationary equations, although it has always been the case in our simulations.
2) There is no guarantee that the algorithm converges towards a global optimum of the criterion. Krämer (2007) has given an example of convergence to a local optimum.

V Special cases of RGCCA

The great power and flexibility of RGCCA allow a large spectrum of methods to be recovered as special cases. In the following, we list the main ones.

V.1. Multi-block data analysis

In multi-block data analysis, all blocks $\mathbf{X}_j$, $j = 1,...,J$ are connected and many criteria exist with the objective of finding components $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$ with useful properties. Many of them are listed in Hanafi & Kiers (2006) and Tenenhaus & Hanafi (2010). In this section, we only consider methods related to optimization problem (19) with all $c_{jk} = 1$ for $j \neq k$. In Table 1 various methods which are all special cases of RGCCA are listed.

Insert Table 1 approximately here

The SUMCOR criterion has been proposed by Horst (1961) and the SSQCOR criterion by Kettenring (1971). The SUMCOV criterion is in fact a special case of the MAXDIFF criterion

proposed by Van de Geer (1984) and the SSQCOV criterion is a special case of the MAXDIFF B criterion proposed by Hanafi and Kiers (2006). The SABSCOR and SABSCOV criteria are not common in the multi-block data analysis literature and are simply a special case of RGCCA. From the point of view of optimization problem (19), the first three criteria correspond to shrinkage constants τ_j equal to 0 for all blocks and the last three criteria correspond to shrinkage constants equal to 1 for all blocks.

We may also consider the situation where the shrinkage constants are 0 for some blocks and 1 for others. Therefore, optimization problem (19), with all $c_{jk} = 1$ for $j \neq k$, becomes

$$\begin{aligned} &\underset{\mathbf{a}_1,\ldots,\mathbf{a}_J}{\text{Maximize}} \quad \sum_{j,k=1,\, j\neq k}^{J} g\Big[\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)(\text{Var}(\mathbf{X}_j\mathbf{a}_j))^{\tau_j/2}(\text{Var}(\mathbf{X}_k\mathbf{a}_k))^{\tau_k/2} \Big] \\ &\text{subject to the constraints } \tau_j \left\|\mathbf{a}_j\right\|^2 + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1 \text{ with } \tau_j = 0 \text{ or } 1,\ j = 1,\ldots,J \end{aligned} \tag{22}$$

If $\tau_j = 0$ (mode B), the main objective in the construction of the outer component $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$ is to maximize its correlation with its neighboring components. Conversely, if $\tau_j = 1$ (new mode A), the objective is to construct a component $\mathbf{y}_j = \mathbf{X}_j\mathbf{a}_j$ which well explains its own block $\mathbf{X}_j$ (first priority), and at the same time, is well correlated to its neighboring components (second priority).

V.2 Regularized Canonical Correlation Analysis

For the two block situation, optimization problem (19) becomes:

$$\begin{aligned} &\underset{\mathbf{a}_1,\mathbf{a}_2}{\text{Maximize}} \ \text{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)) \\ &\text{subject to the constraints } \tau_j \left\|\mathbf{a}_j\right\|^2 + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1,\ j = 1,2 \end{aligned} \tag{23}$$

This problem has been introduced under the name of "Regularized Canonical Correlation Analysis" for handling high dimensional data or highly correlated data in order to stabilize the solution (see Vinod, 1976, Leurgans, Moyeed and Silverman, 1993; Shawe-Taylor and Cristianini, 2004). When one block is reduced to only one variable, optimization problem (23) is equivalent to the simple continuum regression approach proposed by Qannari and Hanafi (2005). Using the stationary equations (14) we get:

- $\mathbf{a}_1$ is the eigenvector of the matrix $[\tau_1\mathbf{I}+(1-\tau_1)\mathbf{S}_{11}]^{-1}\mathbf{S}_{12}[\tau_2\mathbf{I}+(1-\tau_2)\mathbf{S}_{22}]^{-1}\mathbf{S}_{21}$ related to the largest eigenvalue λ_1^2,
- $\mathbf{a}_2$ is the eigenvector of the matrix $[\tau_2\mathbf{I}+(1-\tau_2)\mathbf{S}_{22}]^{-1}\mathbf{S}_{21}[\tau_1\mathbf{I}+(1-\tau_1)\mathbf{S}_{11}]^{-1}\mathbf{S}_{12}$ related to the largest eigenvalue λ_1^2.

This problem covers a situation which goes from inter-battery factor analysis (or equivalently PLS regression with one component) to canonical correlation analysis while passing through redundancy analysis. The special case $0 \le \tau_1 \le 1$ and $\tau_2 = 1$ which corresponds to a regularized version of redundancy analysis has been studied by Takane and Hwang (2007) and by Bougeard, Hanafi and Qannari (2008) under the name "Continuum redundancy-PLS regression". For various extreme cases $\tau_1 = 0$ or 1 and $\tau_2 = 0$ or 1, we get the methods detailed in Table 2, which corresponds exactly to the framework proposed by Burnham, Viveros & MacGregor (1996).

Insert Table 2 approximately here

V.3 Hierarchical models

The models described in Figure 3 have been called hierarchical models by Wold (1982) and are very common in the chemometric literature. In this paper, the left blocks are called first order blocks and are often considered as predictor blocks. The right blocks are called second order blocks and are often considered as response blocks. We consider J_1 first order blocks $\mathbf{X}_1,...,\mathbf{X}_{J_1}$ and $J-J_1$ second order block $\mathbf{X}_{J_1+1},...,\mathbf{X}_J$. We will discuss two cases: the one response block case (Figure 3a) and the several response blocks case (Figure 3b).

Insert Figure 3 approximately here

V.3.1 Hierarchical model with one second order block

The following optimization problem

$$\text{(24)} \qquad \begin{aligned} &\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \ \sum_{j=1}^{J} \text{g}(\text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})) \\ &\text{subject to the constraints: } \tau_j \left\|\mathbf{a}_j\right\|^2 + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1, \ \ j = 1,\ldots,J+1 \end{aligned}$$

includes many useful methods that are described in Table 3.

Insert Table 3 approximately here

To the best of our knowledge, the methods corresponding to g being equal to the identity (Horst scheme) or to the absolute value (Centroid scheme) are new. When g is equal to the square function (factorial scheme), optimization problem (24) is similar to "concordance analysis" proposed by Hanafi & Lafosse (2001). An application of concordance analysis using the shrinkage constants $\tau_j = 0$ for $j = 1,\ldots,J$ and $\tau_{J+1} = 1$ (line 4 of Table 3 with g being equal to the square function) has been proposed by Bougeard, Hanafi and Qannari (2007). These authors note that, due to the choice of the shrinkage constants, this method is oriented towards the construction of predictor block components correlated as much as possible with a stable (large variance) response block component. If the objective is to build good prediction of the response block with stable predictor block components, then the methods corresponding to line 3 of Table 3 would be more adapted. An illustration of the latter method is shown in the application section.

When the second order block $\mathbf{X}_{J+1}$ is equal to the concatenation of the first order blocks $\mathbf{X}_1,\ldots,\mathbf{X}_J$ ($\mathbf{X}_{J+1} = [\mathbf{X}_1,\ldots,\mathbf{X}_J]$ and is called the super-block) this hierarchical model leads to the Horst's and Carroll's GCCA (see Table 4).

Insert Table 4 approximately here

The proofs that the optimization problems given in the second column of Table 4 yield the same optimal solutions $\mathbf{a}_1,\ldots,\mathbf{a}_J$ (and, for Carroll's GCCA, the same global component $\mathbf{X}_{J+1}\mathbf{a}_{J+1}$) as the methods given in the first column of Table 4 are given in Appendix 2. The interest of these methods

compared to the previous multi-block data analysis methods, lies in the computation of a global component related to the super-block. Furthermore, one might notice that "Multiple Co-inertia Analysis" proposed by Chessel & Hanafi (1996) is in fact a special case of Carroll's GCCA with only the covariance criterion being used ($J_1 = 0$ in Table 4). At last, it's worth mentioning that Dahl & Naes (2006) and Takane, Hwang & Abdi (2008) incorporated other ridge type regularization into Carroll's GCCA.

V.3.2 Hierarchical model with several second order blocks

All second order blocks are not necessarily connected to all first order blocks. Therefore, in this section, we need to introduce the design coefficient c_{jk} equal to 1 if the second order block $\mathbf{X}_k$ is related to the first order block $\mathbf{X}_j$ and 0 otherwise. The following optimization problem

$$\begin{aligned} &\underset{\mathbf{a}_1,\ldots,\mathbf{a}_J}{\text{Maximize}} \sum_{j=1}^{J_1}\sum_{k=J_1+1}^{J} c_{jk}\, g(\text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)) \\ &\text{subject to the constraints: } \tau_j \left\|\mathbf{a}_j\right\|^2 + (1-\tau_j)\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1, \;\; j = 1,\ldots,J \end{aligned} \tag{25}$$

corresponds to this kind of design. It contains "generalized orthogonal multiple co-inertia analysis" proposed by Vivien & Sabatier (2003) as a particular case. The first step of this method consists in the following optimization problem

$$\begin{aligned} &\underset{\mathbf{a}_1,\ldots,\mathbf{a}_J}{\text{Maximize}} \sum_{j=1}^{J_1}\sum_{k=J_1+1}^{J} \text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k) \\ &\text{subject to the constraints: } \left\|\mathbf{a}_j\right\| = 1, \;\; j = 1,\ldots,J \end{aligned} \tag{26}$$

Therefore, this method is a special case of RGCCA applied to hierarchical data, using the Horst scheme and new mode A for all blocks. The other steps of this method consist in applying (26) to deflated blocks obtained through various strategies described in Vivien & Sabatier (2003).

V.3.3 PLS path modeling

Wold (1982) proposed the PLS approach to structural equation modeling (SEM) as a component-based alternative to the covariance-based SEM proposed by Jöreskog (1970). The final version of the PLS approach (also called PLS path modeling) is described in detail in Wold (1985).

This algorithm has been modified and extended by Lohmöller (1989) and by Krämer (2007). The most recent review of the PLS approach can be found in Tenenhaus, Esposito Vinzi, Chatelin, Lauro (2005). In PLS path modeling, we consider that each block $\mathbf{X}_j$ is the expression of an unobservable latent variable (LV) and that structural relations (i.e. multiple regression equations) exist between the latent variables. We consider that two blocks $\mathbf{X}_j$ and $\mathbf{X}_k$ are connected if the associated latent variables are related: LV($\mathbf{X}_j$) explains LV($\mathbf{X}_k$) or vice-versa. Let's define the design matrix $\mathbf{C}$: c_{jk} = 1 if blocks $\mathbf{X}_j$ and $\mathbf{X}_k$ are connected in the structural equation model and = 0 otherwise. The PLS algorithm relies on two modes for the latent variable outer estimations (A or B) and four schemes for the latent variable inner estimations (Horst, centroid, factorial and structural). When mode B and the Horst, centroid or factorial schemes are selected, the inner and outer latent variable estimations of PLS path modeling are identical to the inner and outer components of RGCCA. Mode A of PLS path modeling and new mode A of RGCCA leads, however, to different outer components. Both approaches will be compared in further studies.

VI. The Russett example

RGCCA is somewhere between multi-block data analysis and component-based structural equation modelling (Tenenhaus, 2008, Tenenhaus and Hanafi, 2010), and therefore uses ingredients of both approaches. This double filiation is illustrated on the Russet data (Russett, 1964) previously studied in Gifi (1990).

VI.1 Data description

Three blocks of variables have been defined for 47 countries. The first block $\mathbf{X}_1$ = {GINI, FARM, RENT} is related to "Agricultural Inequality": GINI = Inequality of land distribution, FARM = % farmers that own half of the land (> 50), RENT = % farmers that rent all their land. The second block $\mathbf{X}_2$ = {GNPR, LABO} describes "Industrial Development": GNPR = Gross national product per capita ($ 1955), LABO = % of labor force employed in agriculture. The third one $\mathbf{X}_3$ =

{INST, ECKS, DEAT} measures "Political Instability": INST = Instability of Executive (45-61), ECKS = Number of violent internal war incidents (46-61), DEAT = Number of people killed as a result of civic group violence (50-62). An additional variable DEMO describes the political regime: stable democracy, unstable democracy, dictatorship. These data have been used in Gifi (1990) to illustrate nonlinear canonical correlation analysis. This method has been applied on the two blocks {GINI, FARM, RENT, GNPR, LABO} and {INST, ECKS, DEAT, DEMO}. Optimal transformations of these variables are given in Gifi, page 230. We have approximated these optimal transformations. GINI and FARM have not been transformed. The following parametric transformations have been used for the other variables: Ln(RENT+1), Ln(GNPR), Ln(LABO), Exp(INST−16.3), Ln(ECKS+1) and Ln(DEAT+1). The missing data of variable Ln(RENT+1) have been estimated (using multiple regression) for Australia, Nicaragua and Peru by the values 3.27, 2.39 and 2.61. Gifi had some regrets about his analysis. Here is what he wrote on page 226: "*We have chosen to use canonical correlation analysis for the two sets of variables GINI, FARM, RENT, GNPR, LABO versus INST, ECKS, DEAT, DEMO. This is not necessarily the most rational choice. Both Russett's discussion and the definition of the variables suggest that it may be preferable to use three-set canonical correlation analysis.*" Therefore, we have the opportunity in this paper to both repair Gifi's regrets and illustrate RGCCA.

VI.2 Drawing conventions

In RGCCA, we use drawing conventions similar to the PLS path modeling ones. Each block $\mathbf{X}_j$ is represented by an ellipse, each variable $\mathbf{x}_{jh}$ by a rectangle. In the ellipses, we put the block names and in the rectangles the variable names. Each variable is connected to its block by an arrow. For new mode A ($\tau = 1$), the arrow goes from the ellipse to the rectangle in order to symbolize that each outer weight is computed by a simple regression of the manifest variable on the block inner component. For mode B ($\tau = 0$), the arrow goes from the rectangle to the ellipse in order to symbolize that the vector of outer weights for a block is computed by multiple regression of the

block inner component on the block manifest variables. Finally, for mode ridge, we use double headed arrows in order to symbolize the continuum between new mode A and mode B. Two connected blocks are linked with a line. The resulting figure is called a model. Two sub-models are also considered: the outer model is concerned with the relations between the block variables and their block component and the inner model is concerned with the relations between the block components. The model studied for the Russett data is shown in Figure 4 where the transformed variables are labeled using their original names. In this example, we have decided to connect Agricultural Inequality to Political Instability ($c_{13} = 1$), Industrial Develoment to Political Instability ($c_{23} = 1$) and to not connect Agricultural Inequality to Industrial Development ($c_{12} = 0$). The dummy variable "unstable democracy" has been left out because of redundancy. All the manifest variables have been standardized.

Insert Figure 4 approximately here

In section VI.3, both centroid and factorial schemes are compared using mode B for all blocks (full correlation criterion) or new mode A for all blocks (full covariance criterion). We conclude that there is very little difference between these two schemes. In section VI.4, the various modes new A, B and Ridge are compared. Some useful recommendations in selecting the most adequate mode will emerge from these comparisons. Finally in section VI.5, we present a hierarchical version of Barker and Rayens PLS discriminant analysis in order to show the ability of RGCCA to straightforwardly produce a new method.

VI.3 Comparison between centroid and factorial schemes

In this section, we want to compare the centroid and factorial schemes when all blocks are fixed to mode B or to new mode A.

Using mode B

An R-code (R Development Core Team, 2009), available upon request, has been written for the PLS algorithm for RGCCA described in Figure 2. It has been used for analyzing the model

shown in Figure 4 using the factorial scheme. The mode B option is shown by building arrows going from the rectangles to the ellipses. For this model and this scheme, the outer components $\mathbf{y}_1 = \mathbf{X}_1\mathbf{a}_1$, $\mathbf{y}_2 = \mathbf{X}_2\mathbf{a}_2$ and $\mathbf{y}_3 = \mathbf{X}_3\mathbf{a}_3$ maximize the criterion

$$\text{Cor}^2(\mathbf{y}_1, \mathbf{y}_3) + \text{Cor}^2(\mathbf{y}_2, \mathbf{y}_3) \tag{27}$$

To ensure that the global optimum was reached, we tried 50 000 different random initial weights. They all led to the same maximum. If the centroid scheme is used instead of the factorial scheme, then the following criterion is maximized:

$$\left|\text{Cor}(\mathbf{y}_1, \mathbf{y}_3)\right| + \left|\text{Cor}(\mathbf{y}_2, \mathbf{y}_3)\right| \tag{28}$$

Here too, to check for global optimality, we tried 50 000 different random initial weights and they all led to the same maximum. The Russett model analyzed with the centroid scheme and Mode B is not shown. In Table 5, the factorial and centroid schemes with mode B are compared.

Insert Table 5 approximately here

Practice supports theory: the square value criterion is larger for "Mode B + Factorial scheme" than for "Mode B + Centroid scheme" as it should be, and we see the opposite case for the absolute value criterion. The very small difference between the criterion values in both situations confirms a well known result (Noonan & Wold, 1982): the values of the outer components are not very sensitive to the choice of the factorial or centroid schemes.

Using mode A

In Figures 5, we report new mode A results for the factorial scheme. The new mode A option is shown by building arrows going from the ellipses to the rectangles.

Insert Figure 5 approximately here

Factorial and centroid schemes are compared in Table 6. We also checked that the maximum was reached for both criteria by trying 50 000 different random initial weights. Here too, practice supports theory and the values of the outer components are not very sensitive to the choice of the factorial or centroid schemes.

Insert Table 6 approximately here

VI.4 Comparison between modes

In this section, we choose to compare the various modes when the factorial scheme is used. Using the idea of average variance explained (AVE), the following indicators of model quality are defined:

- For one block $\mathbf{X}_j$: $\text{AVE}(\mathbf{X}_j) = (1/\text{p}_j)\sum_{h=1}^{p_j} \text{Cor}^2(\mathbf{x}_{jh}, \mathbf{y}_j)$
- For all blocks: $\text{AVE(outer model)} = (1/\sum_j \text{p}_j)\sum_j \text{p}_j \text{AVE}(\mathbf{X}_j)$
- For the inner model: $\text{AVE(inner model)} = (1/\sum_{j<k} \text{c}_{jk})\sum_{j<k} \text{c}_{jk}\text{Cor}^2(\mathbf{y}_j, \mathbf{y}_k)$

Using these AVEs, new mode A, mode B and mode Ridge associated with the factorial scheme are compared in Table 7.

Insert Table 7 approximately here

As, in the Russett example, all the manifest variables are standardized, the block AVEs and the AVE(outer model) are maximum when the block outer components $\mathbf{y}_j$ are the first principal components of each block $\mathbf{X}_j$. These maximal AVEs are given in Table 7 in the column labeled "mode PCA".

New mode A

For new mode A + factorial scheme, the maximized criterion is here written as

$$\text{Cor}^2(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_3\mathbf{a}_3)\text{Var}(\mathbf{X}_1\mathbf{a}_1)\text{Var}(\mathbf{X}_3\mathbf{a}_3) + \text{Cor}^2(\mathbf{X}_2\mathbf{a}_2, \mathbf{X}_3\mathbf{a}_3)\text{Var}(\mathbf{X}_2\mathbf{a}_2)\text{Var}(\mathbf{X}_3\mathbf{a}_3)$$

with the constraints $\|\mathbf{a}_1\| = \|\mathbf{a}_2\| = \|\mathbf{a}_3\| = 1$. Therefore, this optimization problem leads to a situation between mode PCA and mode B. However, as we note in Table 7, column new mode A is very close to column mode PCA. In fact the block outer components computed using the new mode A are very close to the block first principal components. This is generally the result when the blocks are fairly unidimensional. The reason for this result has been mentioned in section III: when new

mode A is used, the variance terms dominate over the correlation terms. This result is also true for the usual mode A of PLS path modelling (see Tenenhaus, 2008).

Mode B

The AVE(inner model) is maximum for mode B as this is exactly the criterion maximized by mode B + factorial scheme. However, when mode B is used, the value of AVE(Agricultural Inequality) is small: the block component fails to explain its block.

Mode Ridge

For mode Ridge, using the Schäfer & Strimmer's formula, we have computed the optimal shrinkage constants $\tau_1 = 0.1355$, $\tau_2 = 0.0739$ and $\tau_3 = 0.1242$. Results for "Mode Ridge + Factorial scheme" are shown in Figure 6 and Table 7. To figure out the mode Ridge, double-headed arrows are used for connecting rectangles and ellipses.

Insert Figure 6 approximately here

In mode Ridge, the AVEs are located between those of new mode A and mode B. New mode A is favoring the outer model and mode B is favoring the inner model. Mode Ridge appears to be a useful compromise between new mode A and mode B. As compared to mode B, mode Ridge yields components with higher block AVEs, specifically for Agricultural Inequality, and comparable AVE(inner model).

Recommendations

New mode A is recommended when the user wants a stable component (large variance) while taking into account the correlations with connected blocks. The user must, however, be aware that variance dominates over correlation. Mode B is recommended when the user wants to maximize correlations between connected components. This option can yield unstable solutions in case of multicollinearity and cannot be used when a data block is not of full rank. Mode Ridge is a good compromise between new mode A and mode B: the block component is simultaneously stable and as well correlated as possible with its connected components. Mode Ridge can be used when the data

block is not of full rank. The shrinkage constants can be determined by using the Schäfer & Strimmer's formula.

VI.5 Hierarchical Barker & Rayens PLS-DA

It is possible to mix the modes of the various blocks. A special case of this problem is the PLS method for discrimination proposed by Barker & Rayens (2003). They consider a block **X** of explanatory variables and a block **Y** of dummy variables describing a categorical variable. They are looking for a component **Xa** (with **a** normalized) and a standardized component **Yb** maximizing the criterion

$$\text{Cov}(\mathbf{Xa}, \mathbf{Yb}) = \text{Cor}(\mathbf{Xa}, \mathbf{Yb}) \times \text{Var}(\mathbf{Xa})^{1/2} \tag{29}$$

The solution of the Barker & Rayens PLS-DA is therefore obtained by running a redundancy analysis of block **X** with respect to block **Y**. Thus, Barker & Rayens PLS-DA is an intermediate solution between discriminant analysis (Maximize $\text{Cor}(\mathbf{Xa}, \mathbf{Yb})$ subject to **Xa** and **Yb** standardized) and PLS-DA (Maximize $\text{Cov}(\mathbf{Xa}, \mathbf{Yb}) = \text{Cor}(\mathbf{Xa}, \mathbf{Yb}) \times \text{Var}(\mathbf{Xa})^{1/2} \times \text{Var}(\mathbf{Yb})^{1/2}$ subject to **a** and **b** normalized). Suppose now that several blocks $\mathbf{X}_1, ..., \mathbf{X}_J$ of explanatory variables are used to predict the block **Y** of dummy variables describing the categorical variable. It is then rather natural to extend the Barker & Rayens PLS-DA to a new method called hierarchical Barker & Rayens PLS-DA. Therefore, we look for components $\mathbf{X}_1\mathbf{a}_1, ..., \mathbf{X}_J\mathbf{a}_J$ (with $\|\mathbf{a}_j\| = 1,\ j = 1, ..., J$) and for a standardized component **Yb** maximizing the criterion

$$\sum_{j=1}^{J} \text{g}[\text{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{Yb})] = \sum_{j=1}^{J} \text{g}[\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{Yb}) \text{Var}(\mathbf{X}_j\mathbf{a}_j)^{1/2}] \tag{30}$$

where g is the identity (Horst scheme), the square (factorial scheme) or the absolute value (centroid scheme). In this problem, new mode A is used for the explanatory blocks $\mathbf{X}_1, ..., \mathbf{X}_J$ and mode B is used for the response block **Y**.

We may illustrate the hierarchical Barker & Rayens PLS discriminant analysis by relating the block formed by the Agricultural Inequality ($\mathbf{X}_1$) and the block formed by Industrial Development ($\mathbf{X}_2$) to the **Y** block formed by the political regimes "stable democracy" and "dictatorship". Here too, the dummy variable "unstable democracy" is left out because of redundancy. RGCCA is applied to the three blocks with $\mathbf{X}_1$, $\mathbf{X}_2$ as first order blocks and **Y** as the second order block. When the new mode A is used for all blocks, hierarchical PLS-DA is obtained. When mode B is used for all blocks, hierarchical factorial discriminant analysis is obtained. At last, when the new mode A is used for the **X** blocks and mode B for **Y**, we obtain hierarchical redundancy analysis of $\mathbf{X}_1$, $\mathbf{X}_2$ with respect to **Y** (i.e. hierarchical Barker & Rayens PLS-DA). Using the factorial scheme, this last method leads to maximizing the criterion

(31) $$\mathrm{Cov}^2(\mathbf{X}_1\mathbf{a}_1, \mathbf{Yb}) + \mathrm{Cov}^2(\mathbf{X}_2\mathbf{a}_2, \mathbf{Yb})$$

subject to the constraints:

(32) $$\|\mathbf{a}_1\| = \|\mathbf{a}_2\| = \mathrm{Var}(\mathbf{Yb}) = 1$$

This is equivalent to maximizing the criterion

(33) $$\mathrm{Cor}^2(\mathbf{X}_1\mathbf{a}_1, \mathbf{Yb})\mathrm{Var}(\mathbf{X}_1\mathbf{a}_1) + \mathrm{Cor}^2(\mathbf{X}_2\mathbf{a}_2, \mathbf{Yb})\mathrm{Var}(\mathbf{X}_2\mathbf{a}_2)$$

subject to the constraints:

(34) $$\|\mathbf{a}_1\| = \|\mathbf{a}_2\| = 1$$

We get the components from Figure 7 (* signifies standardized)

$$\mathbf{X}_1\mathbf{a}_1 = .62\mathrm{Gini}^* + .75\mathrm{Farm}^* - .22\mathrm{Rent}^*$$

$$\mathbf{X}_2\mathbf{a}_2 = .67\mathrm{GNPR}^* - .74\mathrm{Labo}^*$$

$$\mathbf{Yb} = -.72(\text{Stable Democracy})^* + .39(\text{Dictatorship})^*$$

Insert Figure 7 approximately here

The graphical display of the countries obtained by crossing $\mathbf{X}_1\mathbf{a}_1$ = Agricultural Inequality and $\mathbf{X}_2\mathbf{a}_2$ = Industrial Development and marked with their political regime in 1960 is shown in Figure 8. It may be noted that the upper left quadrant concentrates on dictatorships. It is difficult for a country to escape dictatorship when its industrial development is below-average and its agricultural inequality is above-average. It is worth pointing out that some unstable democracies located in this quadrant (or close to it) became dictatorships for a period of time after 1960: Greece (1967-1974), Brazil (1964-1985), Chili (1973-1990), Argentina (1966-1973).

Insert Figure 8 approximately here

VII. Conclusion

Regularized Generalized Canonical Correlation Analysis is a very attractive and general framework for multi-block data analysis with good generalization in the case where each block is connected to only a subset of blocks. By defining the design matrix, the shrinkage constants and the function g, RGGCA includes a remarkably large spectrum of methods as particular cases. In this paper, we have also tried to define a guideline for the choice of the shrinkage constants by providing interpretations on the properties of the resulting block components. If one wants to give priority to the correlation between a component and its neighboring components, mode B ($\tau_j = 0$) is the natural choice. Conversely, if one wants a component which explains its own block well, and remains, however, correlated to its neighboring components, new mode A ($\tau_j = 1$) is the natural choice. For a better compromise between variance and correlation, mode Ridge is proposed in this paper. For each block, the determination of the shrinkage constant can be made fully automatic by using the analytical formula proposed by Schäfer and Strimmer. In contrast to other common approaches that are based on cross-validation or bootstrap strategies, the computational effort of this technique is very attractive. Moreover, the regularization aspect of RGCCA allows one to come closer to mode B even in case of high multicollinearity within blocks or when the number of individuals is smaller than the number of variables.

Several points have been eluded in this paper but will be investigated in future research. (i) Computing several components for each block: In this paper, we suppose that each block is summarized by one block component only. Using a deflation strategy, it is possible to compute additional components for each block which are orthogonal to the previous ones. (ii) Model assessment by bootstrap strategy: RGCCA can be viewed as an estimation procedure for population GCCA. Therefore, using a bootstrap strategy, it is possible to compute confidence intervals for all the elements computed in RGCCA. (iii) Comparison with PLS path modeling: Wold (1982) mentions a drawback of the outer model estimation step of PLS path modeling: "*Most nonlinear iterative techniques of estimation are lacking an analytic proof of convergence. The proof of the pudding is in the eating*". This outer model estimation step can be replaced by RGCCA, a monotonically convergent algorithm. Simulations will be carried out in further studies to compare PLS path modeling (mode A) and RGCCA (new mode A) from a practical point of view. Nevertheless, this paper already suggests that the PLS approach of Herman Wold is much more powerful and founded on much more solid theoretical bases than it seemed to be at first. (iv) Nonlinear relationships between blocks: Finally, RGCCA captures only linear relations between blocks. To assess nonlinear relations, a kernel extension of RGCCA is currently under investigation.

Acknowledgments

The authors are very grateful to Mohamed Hanafi, Yoshio Takane, the editor, the associate editor and three anonymous reviewers for their constructive comments that contributed to improving the quality of the paper. The authors also thank Monika Jelizarow for her contribution in the programming.

References

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

Bougeard S., Hanafi M. and Qannari E.M. (2007): ACPVI Multibloc. Application en épidémiologie animale. *Journal de la Société Française de Statistique*, 148, 4, 77-94.

Bougeard S., Hanafi M. and Qannari E.M. (2008): Continuum redundancy-PLS regression: a simple continuum approach. *Computational Statistics and Data Analysis*, 52, 3686-3696.

Burnham A.J., Viveros R. & MacGregor J.F. (1996): Frameworks for latent variable multivariate regression, *Journal of Chemometrics*, 10, 31-45.

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

Carroll, J.D. (1968b): Equations and tables for a generalization of canonical correlation analysis to three or more sets of variables. *Unpublished companion paper to Carroll, J.D. (1968a)*.

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

Chu M.T. and Watterson J.L. (1993): On a multivariate eigenvalue problem: I. Algebraic theory and power method, *SIAM J. Sci. Comput.*, 14, 1089-1106.

Dahl T. and Næs T. (2006): A bridge between Tucker-1 and Carroll's generalized canonical analysis. *Computational Statistics and Data Analysis*, 50, 3086-3098.

Fornell C. and Bookstein F.L. (1982): Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 19, 440-452.

Gifi A. (1990): *Nonlinear multivariate analysis*, John Wiley & Sons, Chichester, UK.

Hanafi M. (2007): PLS Path modelling: computation of latent variables with the estimation mode B, *Computational Statistics*, 22, 275-292.

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.

Hanafi M. & Lafosse R. (2001): Généralisations de la régression simple pour analyser la dépendance de K ensembles de variables avec un K+1ème. *Revue de Statistique Appliquée*, 49, 1, 5-30.

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

Jöreskog K.G. (1970): A general method for the analysis of covariance structure, *Biometrika*, 57, 239-251.

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

Krämer N. (2007): Analysis of high-dimensional data with partial least squares and boosting. Doctoral dissertation, Technischen Universität Berlin.

Ledoit O. & Wolf M. (2004): A well conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88, 365-411.

Leurgans S. E., Moyeed R. A. and Silverman, B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society*. Series B. 55 725–740.

Lohmöller J.-B. (1989): *Latent Variables Path Modeling with Partial Least Squares*, Physica-Verlag, Heildelberg.

Noonan, R., Wold, H. (1982). PLS path modeling with indirectly observed variables: a comparison of alternative estimates for the latent variable. in *Systems under indirect observation, Part 2*, K.G. Jöreskog & H. Wold (Eds), North-Holland, Amsterdam, pp. 75–94.

Qannari E.M. and Hanafi M. (2005): A simple continuum regression approach, *Journal of Chemometrics*, 19, 387-392.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Russett B.M. (1964): Inequality and Instability: The Relation of Land Tenure to Politics. *World Politics* 16:3, 442-54.

Shawe-Taylor J. and Cristianini N. (2004): *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, NY, USA.

Schäfer J and Strimmer K. (2005): A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical applications in genetics and molecular biology* 4, 1, Article 32.

Takane Y. and Hwang H. (2007): Regularized linear and kernel redundancy analysis. *Computational Statistics and Data Analysis*, 52:394-405.

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

Ten Berge J.M.F. (1988): Generalized approaches to the MAXBET problem and the MAXDIFF problem, with applications to canonical correlations, *Psychometrika,* 53, 487-494.

Tenenhaus M. (2008): Component-based structural equation modelling. *Total Quality Management & Business Excellence,* vol. 19, n° 7, pp. 871-886.

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

Tenenhaus M & Hanafi M. (2010): A bridge between PLS path modeling and multi-block data analysis. In *Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications*, V. Esposito Vinzi, J. Henseler, W. Chin, H. Wang, Eds, Springer Verlag, pp. 99-123.

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

Van de Geer J. P. (1984): Linear relations among *k* sets of variables. *Psychometrika, 4*9, 70-94.

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

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

Westerhuis J.A., Kourti T. & MacGregor J.F. (1998): Analysis of multiblock and hierarchical PCA and PLS models. Journal of Chemometrics, 12, 301-321.

Wold H. (1982): Soft Modeling: The Basic Design and Some Extensions, in *Systems under indirect observation, Part 2*, K.G. Jöreskog & H. Wold (Eds), North-Holland, Amsterdam, pp. 1-54.

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.

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.

APPENDIX 1

Monotonic convergence of the PLS algorithm for population GCCA

Hanafi (2007) has proven the monotonic convergence of the PLS algorithm of Wold (1985) when the options mode B for all blocks and the centroid or factorial schemes are used. In this appendix, we adapt the proof of Hanafi to the PLS algorithm for population GCCA described in Figure 1. The proof of Proposition 2 relies on the following lemma, where the outer weight vectors $\boldsymbol{\alpha}_j^s$ and the inner components ν_j^s are defined in Figure 1:

Lemma 1: For $j = 1,...,J$, $s = 0,1,2,...$, let f_j^s be the function defined by

$$\mathrm{f}_j^s(\boldsymbol{\alpha}_j) = \sum_{k<j} \mathrm{c}_{jk}\mathrm{g}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k)\right] + \sum_{k>j} \mathrm{c}_{jk}\mathrm{g}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s})^t \boldsymbol{x}_k)\right]$$

Then the following property holds

$$\forall\ s \quad \mathrm{f}_j^s(\boldsymbol{\alpha}_j^s) \le \mathrm{f}_j^s(\boldsymbol{\alpha}_j^{s+1}) \tag{35}$$

Proof of Lemma 1:

The function $\mathrm{f}_j^s(\boldsymbol{\alpha}_j^s)$ may be written as:

$$\begin{aligned}
\mathrm{f}_j^s(\boldsymbol{\alpha}_j^s) &= \sum_{k<j} \mathrm{c}_{jk}\mathrm{g}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k)\right] + \sum_{k>j} \mathrm{c}_{jk}\mathrm{g}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s})^t \boldsymbol{x}_k)\right] \\
&= \sum_{k<j} \mathrm{c}_{jk}\mathrm{w}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k)\right]\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) \\
&\qquad + \sum_{k>j} \mathrm{c}_{jk}\mathrm{w}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s})^t \boldsymbol{x}_k)\right]\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s})^t \boldsymbol{x}_k) \\
&= \mathrm{Cov}\left\{(\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, \sum_{k<j} \mathrm{c}_{jk}\mathrm{w}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k)\right](\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k \right. \\
&\qquad \left. + \sum_{k>j} \mathrm{c}_{jk}\mathrm{w}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s})^t \boldsymbol{x}_k)\right](\boldsymbol{\alpha}_k^{s})^t \boldsymbol{x}_k \right\} \\
&= \mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, \nu_j^s)
\end{aligned} \tag{36}$$

Using the definitions of ν_j^s and $\boldsymbol{\alpha}_j^{s+1}$, the following inequality holds

$$\mathrm{f}_j^s(\boldsymbol{\alpha}_j^s) = \mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, \nu_j^s) \le \mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, \nu_j^s) \tag{37}$$

and we get the equality

$$\text{(38)}\quad \begin{aligned}\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,\nu_j^s) &= \sum_{k<j} c_{jk}\, w\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\right]\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\\ &\quad+\sum_{k>j} c_{jk}\, w\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\right]\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\end{aligned}$$

Then, we have to consider the three schemes separately.

Horst scheme (g = identity and w(x) = 1)

For the Horst scheme, equality (38) becomes

$$\text{(39)}\quad \mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,\nu_j^s) = \sum_{k<j} c_{jk}\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k) + \sum_{k>j} c_{jk}\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k) = f_j^s(\boldsymbol{\alpha}_j^{s+1})$$

and therefore $f_j^s(\boldsymbol{\alpha}_j^s) \le f_j^s(\boldsymbol{\alpha}_j^{s+1})$.

Centroid scheme (g = absolute value and w(x) = sign(x))

For the centroid scheme

$$\text{(40)}\quad f_j^s(\boldsymbol{\alpha}_j^s) = \sum_{k<j} c_{jk}\left|\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\right| + \sum_{k>j} c_{jk}\left|\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\right|$$

and

$$\text{(41)}\quad \begin{aligned}\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,\nu_j^s) &= \sum_{k<j} c_{jk}\,\mathrm{sign}(\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\right])\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\\ &\quad+\sum_{k>j} c_{jk}\,\mathrm{sign}(\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\right])\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\end{aligned}$$

Therefore, we get

$$\text{(42)}\quad \begin{aligned}&f_j^s(\boldsymbol{\alpha}_j^s) = \left|f_j^s(\boldsymbol{\alpha}_j^s)\right| \le \mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,\nu_j^s) = \left|\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,\nu_j^s)\right|\\ &\le \sum_{k<j} c_{jk}\left|\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\right| + \sum_{k>j} c_{jk}\left|\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\right| = f_j^s(\boldsymbol{\alpha}_j^{s+1})\end{aligned}$$

Factorial scheme (g = square and w(x) = x)

For the factorial scheme

$$\text{(43)}\quad f_j^s(\boldsymbol{\alpha}_j^s) = \sum_{k<j} c_{jk}\mathrm{Cov}^2((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k) + \sum_{k>j} c_{jk}\mathrm{Cov}^2((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)$$

and

$$\text{(44)}\quad \begin{aligned}\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,\nu_j^s) &= \sum_{k<j} c_{jk}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\right]\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s+1})^t\boldsymbol{x}_k)\\ &\quad+\sum_{k>j} c_{jk}\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\right]\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t\boldsymbol{x}_j,(\boldsymbol{\alpha}_k^{s})^t\boldsymbol{x}_k)\end{aligned}$$

The right term of (44) is the scalar product between two vectors. Using the Cauchy-Schwartz inequality and the fact that $c_{jk}^2 = c_{jk}$, we get

$$\begin{aligned}(45)\qquad \mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, \nu_j^s) \le & [\sum_{k<j} c_{jk} \mathrm{Cov}^2((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) + \sum_{k>j} c_{jk} \mathrm{Cov}^2((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k)]^{1/2} \\ & \times [\sum_{k<j} c_{jk} \mathrm{Cov}^2((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) + \sum_{k>j} c_{jk} \mathrm{Cov}^2((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k)]^{1/2}\end{aligned}$$

From that we deduce

$$(46)\qquad f_j^s(\boldsymbol{\alpha}_j^s) \le [f_j^s(\boldsymbol{\alpha}_j^s)]^{1/2} [f_j^s(\boldsymbol{\alpha}_j^{s+1})]^{1/2}$$

and therefore

$$(47)\qquad f_j^s(\boldsymbol{\alpha}_j^s) \le f_j^s(\boldsymbol{\alpha}_j^{s+1})$$

Proof of Proposition 2

Proposition 2 is deduced from lemma 1 and the following equality:

$$\begin{aligned}\sum_{j=1}^J \left[f_j^s(\boldsymbol{\alpha}_j^{s+1}) - f_j^s(\boldsymbol{\alpha}_j^s) \right] = & \sum_{k<j} c_{jk} g\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) \right] + \sum_{k>j} c_{jk} g\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k) \right] \\ & - \sum_{k<j} c_{jk} g\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) \right] - \sum_{k>j} c_{jk} g\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k) \right] \\ = & \frac{1}{2} \sum_{j,k=1, j\neq k}^J c_{jk} g\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^{s+1})^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) \right] - \frac{1}{2} \sum_{j,k=1, j\neq k}^J c_{jk} g\left[\mathrm{Cov}((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k) \right] \\ = & \frac{1}{2} \left[f(\boldsymbol{\alpha}_1^{s+1}, \ldots, \boldsymbol{\alpha}_J^{s+1}) - f(\boldsymbol{\alpha}_1^s, \ldots, \boldsymbol{\alpha}_J^s) \right]\end{aligned}$$

APPENDIX 2

Proof of the equivalences mentioned in Table 4

1) SUMCOR

The stationary equations for the SUMCOR method (see first line of Table 1) are obtained by using equation (13) with all $c_{jk} = 1$ for $j \neq k$, with the Horst scheme and equation (14) with all $\tau_j = 0$. Pre-multiplying these equations by $\mathbf{X}_j$, they become

(48) $$\mathbf{y}_j = \mathbf{X}_j \mathbf{a}_j \propto \mathbf{P}_j (\sum_{k \neq j} \mathbf{y}_k)$$

where $\propto$ means that the left side is the standardized version of the right side and where $\mathbf{P}_j = \mathbf{X}_j (\mathbf{X}_j^t \mathbf{X}_j)^{-1} \mathbf{X}_j^t$ is the orthogonal projection operator onto the $\mathbf{X}_j - \text{space}$. We now consider the stationary equations of the following optimization problem (mode B for all blocks and Horst scheme)

(49) $$\underset{\mathbf{a}_1,...,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} \text{Cor}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_{J+1} \mathbf{a}_{J+1})$$
$$\text{subject to the constraints: } \text{Var}(\mathbf{X}_j \mathbf{a}_j) = 1, \; j = 1,..., J+1$$

where $\mathbf{X}_{J+1} = [\mathbf{X}_1,..., \mathbf{X}_J]$ is the super-block. In this situation $c_{j,J+1} = 1$ for $j = 1,..., J$ and $c_{jk} = 0$ otherwise. The stationary equations for this optimization problem are

(50) $$\mathbf{y}_j = \mathbf{X}_j \mathbf{a}_j \propto \mathbf{P}_j (\mathbf{X}_{J+1} \mathbf{a}_{J+1})$$

(51) $$\mathbf{y}_{J+1} = \mathbf{X}_{J+1} \mathbf{a}_{J+1} \propto \mathbf{P}_{J+1} (\sum_{j=1}^{J} \mathbf{X}_j \mathbf{a}_j)$$

However, as $\sum_{j=1}^{J} \mathbf{X}_j \mathbf{a}_j$ belongs to the space generated by the super-block $\mathbf{X}_{J+1}$, we get

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

Therefore, (50) becomes

(53) $$\mathbf{y}_j \propto \mathbf{P}_j(\sum_{k=1}^{J}\mathbf{y}_k) = \mathbf{y}_j + \mathbf{P}_j(\sum_{k\neq j}\mathbf{y}_k)$$

and equation (48) is found again. The equivalence mentioned in Table 4 is also valid if the centroid scheme is used because the sign of $\mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})$ is equal to the sign of $\mathbf{y}_{J+1}^t\mathbf{P}_j\mathbf{y}_{J+1}$ and therefore always positive.

2) Carroll's Generalized CCA

In his 1968b paper, Carroll considers the "Mixed" correlation and covariance criterion which consists in the following optimization problem

(54) $$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_J,\mathbf{z}}{\text{Maximize}} \ \sum_{j=1}^{J_1}\mathrm{Cor}^2(\mathbf{X}_j\mathbf{a}_j,\mathbf{z}) + \sum_{j=J_1+1}^{J}\mathrm{Cov}^2(\mathbf{X}_j\mathbf{a}_j,\mathbf{z})$$
$$\text{subject to the constraints: } \mathrm{Var}(\mathbf{X}_j\mathbf{a}_j) = \mathrm{Var}(\mathbf{z}) = 1,\ j = 1,\ldots,J_1 \text{ and } \|a_j\| = 1,\ j = J_1+1,\ldots,J$$

The solution of this optimization problem is obtained for the eigenvector **z** associated with the largest eigenvalue of the matrix $\mathbf{Q} = \sum_{j=1}^{J_1}\mathbf{X}_j\left(\mathbf{X}_j^t\mathbf{X}_j\right)^{-1}\mathbf{X}_j^t + \sum_{j=J_1+1}^{J}\mathbf{X}_j\mathbf{X}_j^t$. As the eigenvector **z** belongs to the space generated by the columns of $\mathbf{X}_{J+1}$, optimization problem (54) is equivalent to the following optimization problem

(55) $$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \ \sum_{j=1}^{J_1}\mathrm{Cor}^2(\mathbf{X}_j\mathbf{a}_j,\mathbf{X}_{J+1}\mathbf{a}_{J+1}) + \sum_{j=J_1+1}^{J}\mathrm{Cov}^2(\mathbf{X}_j\mathbf{a}_j,\mathbf{X}_{J+1}\mathbf{a}_{J+1})$$
$$\text{subject to the constraints: } \mathrm{Var}(\mathbf{X}_j\mathbf{a}_j) = 1,\ j = 1,\ldots,J_1,J+1 \text{ and } \|a_j\| = 1,\ j = J_1+1,\ldots,J$$

And this shows the equivalence announced in Table 4.

Figure 1: A PLS algorithm for population GCCA

A. ***Initialisation***

A1. Choose J arbitrary vectors $\tilde{\boldsymbol{\alpha}}_1^0, \tilde{\boldsymbol{\alpha}}_2^0, ..., \tilde{\boldsymbol{\alpha}}_J^0$

A2. Compute normalized outer weight vectors $\boldsymbol{\alpha}_1^0, \boldsymbol{\alpha}_2^0, ..., \boldsymbol{\alpha}_J^0$ as:

$$\boldsymbol{\alpha}_j^0 = [(\tilde{\boldsymbol{\alpha}}_j^0)^t \boldsymbol{\Sigma}_{jj}^{-1} \tilde{\boldsymbol{\alpha}}_j^0]^{-1/2} \boldsymbol{\Sigma}_{jj}^{-1} \tilde{\boldsymbol{\alpha}}_j^0$$

For *s* = 0,1,… (until convergence)

For *j* = 1,2,…,*J*

B. ***Computing the inner component*** v_j^s

Compute the inner component according to the selected scheme:

$$v_j^s = \sum_{k<j} c_{jk} w\left[Cov((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k) \right] (\boldsymbol{\alpha}_k^{s+1})^t \boldsymbol{x}_k + \sum_{k>j} c_{jk} w\left[Cov((\boldsymbol{\alpha}_j^s)^t \boldsymbol{x}_j, (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k) \right] (\boldsymbol{\alpha}_k^s)^t \boldsymbol{x}_k$$

where w(x) = 1 for the Horst scheme, x for the factorial scheme and sign(x) for the centroid scheme.

C. ***Computing the outer weight vector*** $\boldsymbol{\alpha}_j^{s+1}$

Compute the outer weight vector

$$\boldsymbol{\alpha}_j^{s+1} = [Cov(\boldsymbol{x}_j, v_j^s)^t \boldsymbol{\Sigma}_{jj}^{-1} Cov(\boldsymbol{x}_j, v_j^s)]^{-1/2} \boldsymbol{\Sigma}_{jj}^{-1} Cov(\boldsymbol{x}_j, v_j^s)$$

End

End

Figure 2: A PLS algorithm for RGCCA

A. Initialisation

A1. Choose J arbitrary vectors $\tilde{\mathbf{a}}_1^0, \tilde{\mathbf{a}}_2^0, ..., \tilde{\mathbf{a}}_J^0$

A2. Compute normalized outer weight vectors $\mathbf{a}_1^0, \mathbf{a}_2^0, ..., \mathbf{a}_J^0$ as:

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

For *s* = 0,1,…(until convergence)

For *j* = 1,2,…,*J*

B. Computing the inner component for block $\mathbf{X}_j$

Compute the inner component according to the selected scheme:

$$\mathbf{z}_j^s = \sum_{k<j} \mathrm{c}_{jk} \mathrm{w}\left[\mathrm{Cov}(\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^{s+1})\right] \mathbf{X}_k \mathbf{a}_k^{s+1} + \sum_{k>j} \mathrm{c}_{jk} \mathrm{w}\left[\mathrm{Cov}(\mathbf{X}_j \mathbf{a}_j^s, \mathbf{X}_k \mathbf{a}_k^s)\right] \mathbf{X}_k \mathbf{a}_k^s$$

where w(x) = 1 for the Horst scheme, x for the factorial scheme and sign(x) for the centroid scheme.

C. Computing the outer weight vector for block $\mathbf{X}_j$

Compute the outer weight vector

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

End

End

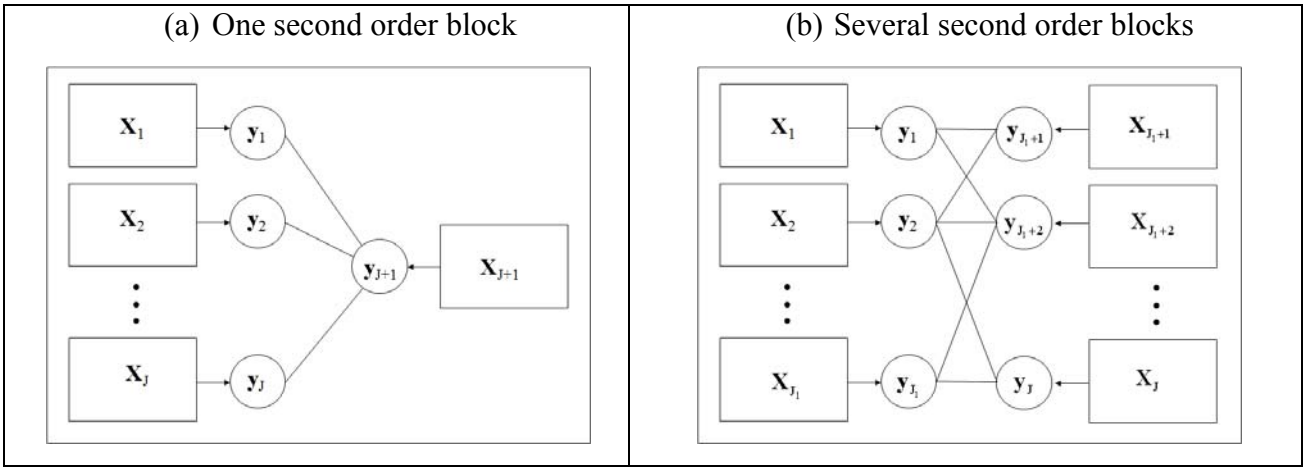

Figure 3: Hierarchical models

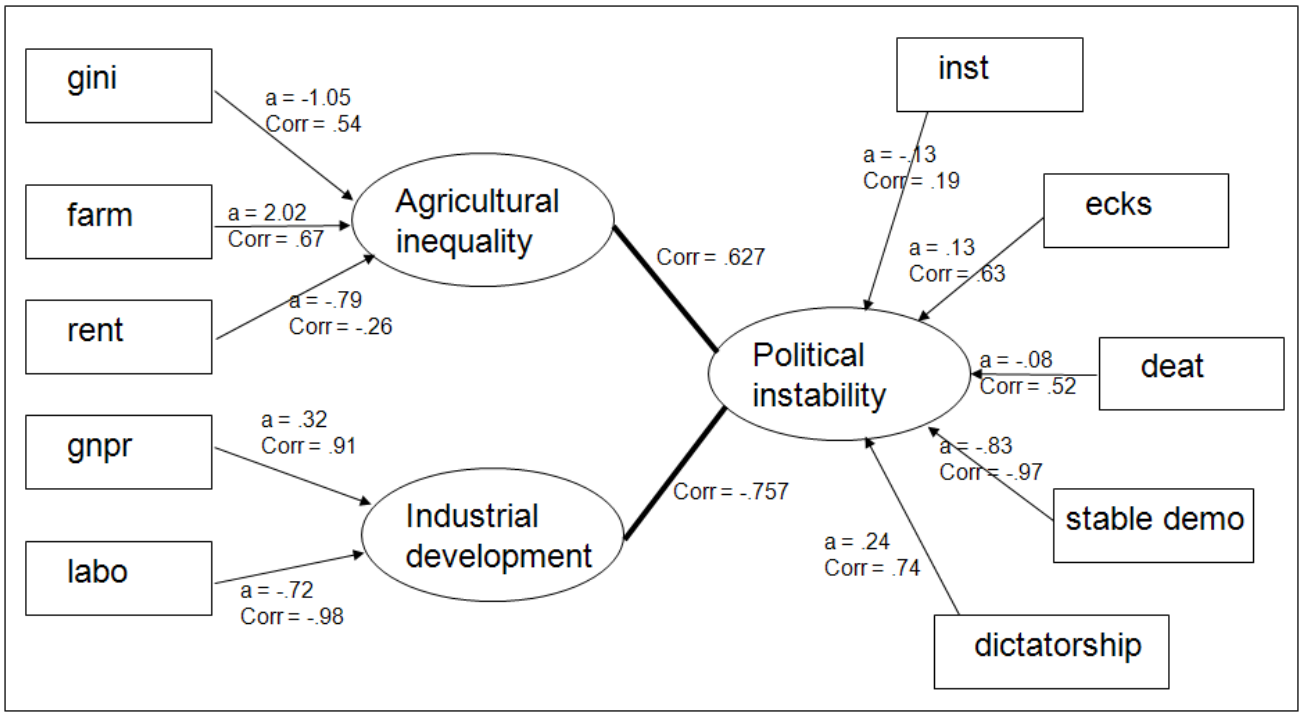

Figure 4: Russett data: Factorial scheme + Mode B

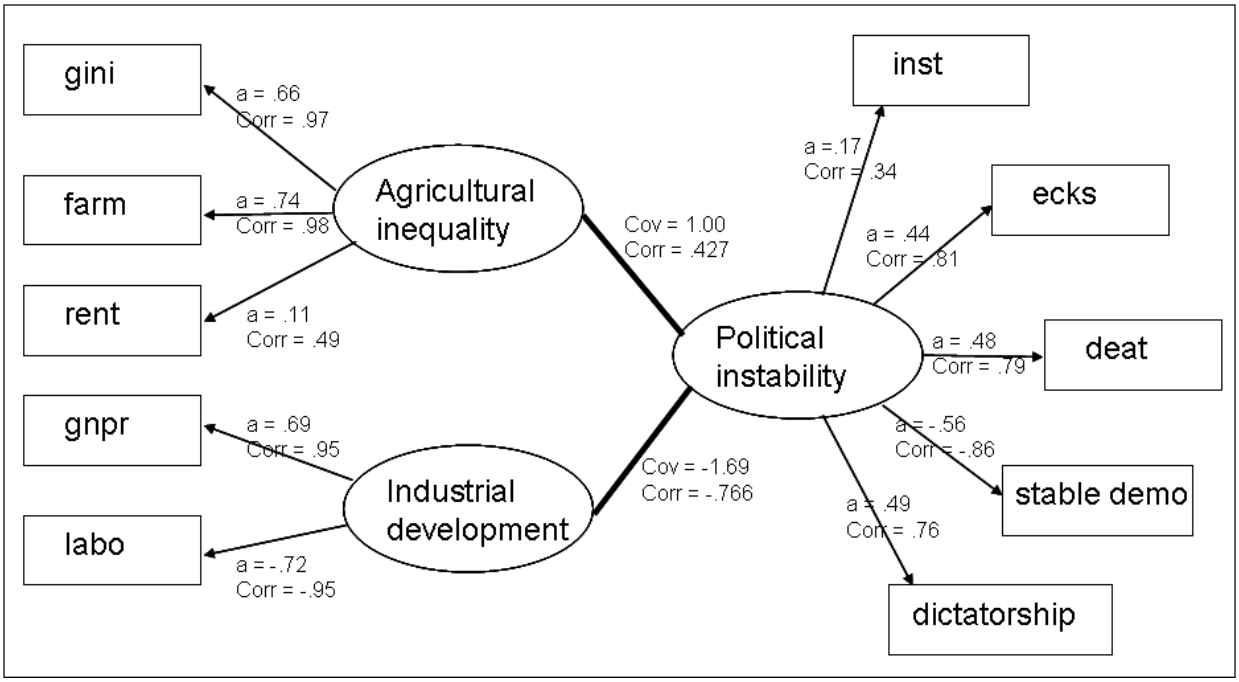

Figure 5: Russett data: Factorial scheme + New mode A

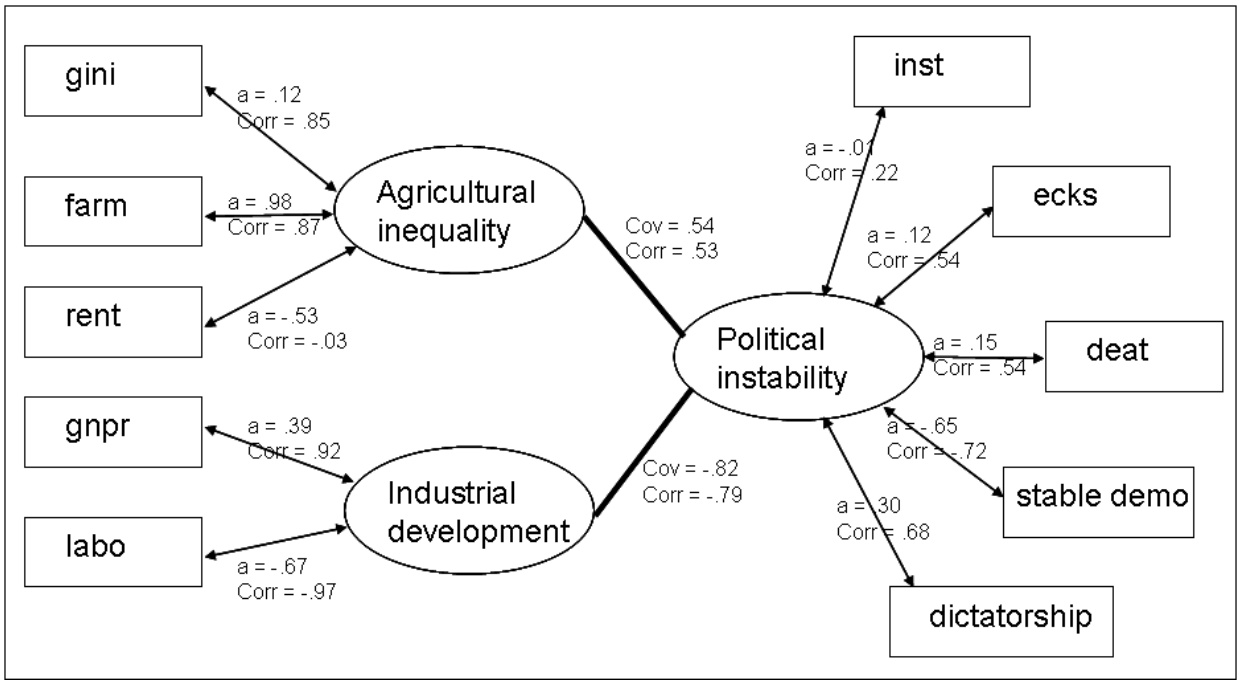

Figure 6: Russett data: Factorial scheme + Mode Ridge (Optimal shrinkage)

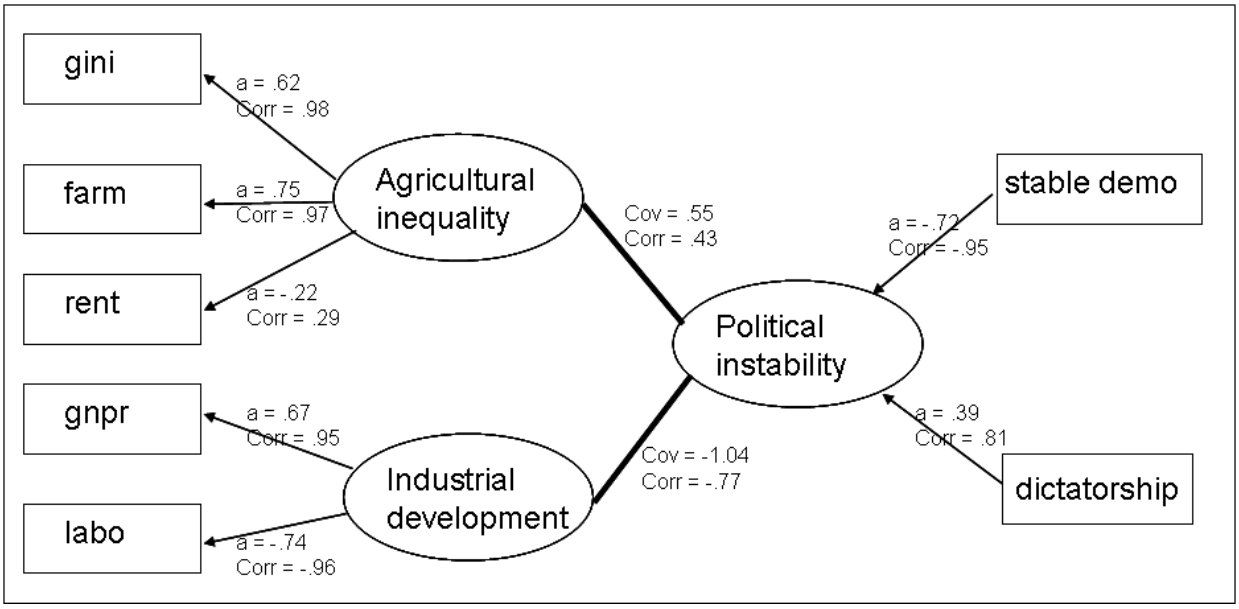

Figure 7: Hierarchical Barker & Rayens PLS Discriminant Analysis

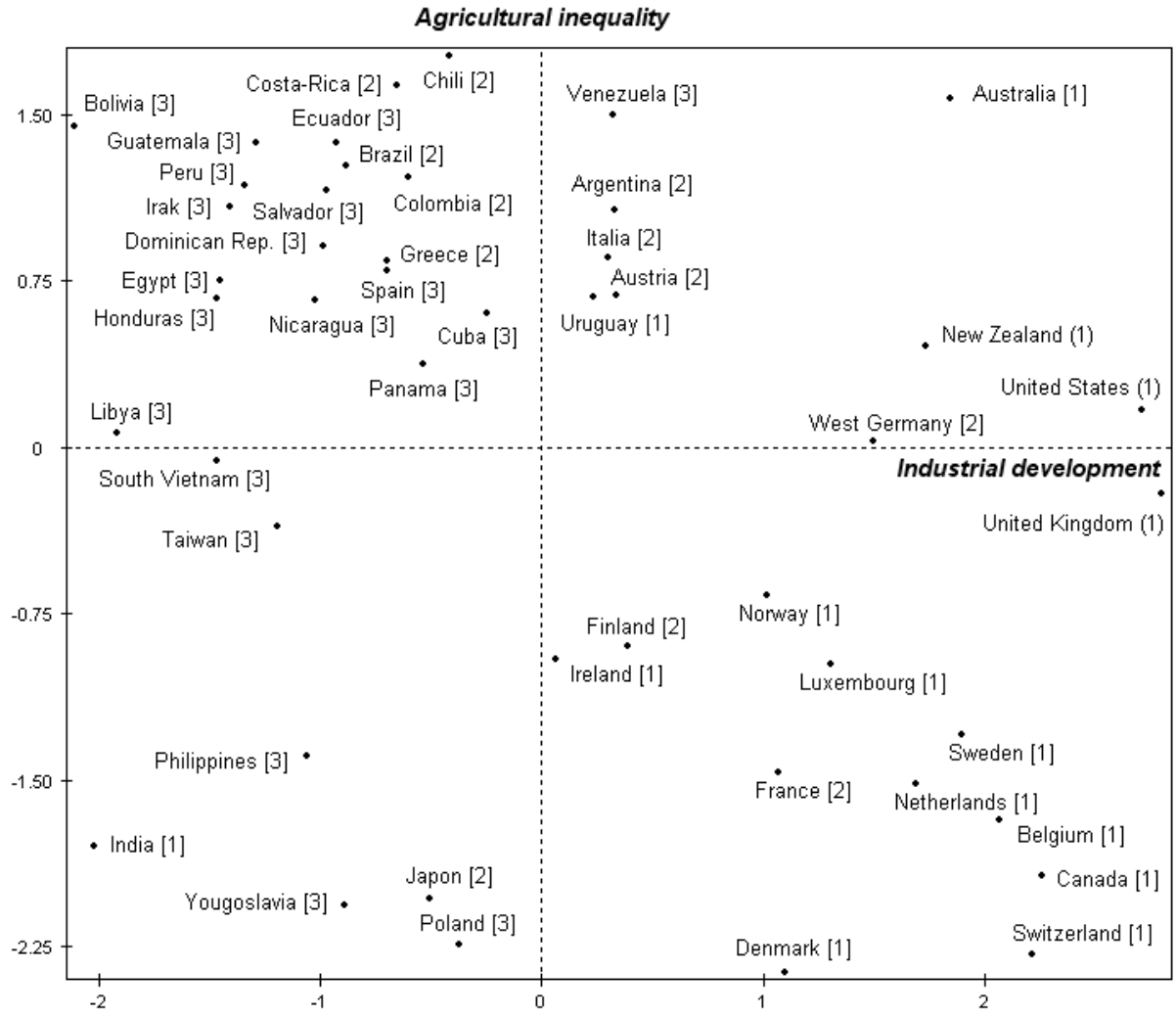

Figure 8: Hierarchical Barker & Rayens PLS Discriminant Analysis
Graphical display of countries marked with their political regime in 1960
(1 = stable democracy, 2 = unstable democracy, 3 = dictatorship)

Table 1: Special cases of RGCCA for multi-block data analysis

Method	**Criterion to be maximized**	**Scheme**	**Normalization**	**Shrinkage constants**
SUMCOR	$\sum_{j,k,j\neq k} \mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$	Horst	$\mathrm{Var}(\mathbf{X}_j\mathbf{a}_j)=1, \quad j=1,...,J$	$\tau_j=0,\ j=1,...,J$
SSQCOR	$\sum_{j,k,j\neq k} \mathrm{Cor}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$	Factorial	$\mathrm{Var}(\mathbf{X}_j\mathbf{a}_j)=1, \quad j=1,...,J$	$\tau_j=0,\ j=1,...,J$
SABSCOR	$\sum_{j,k,j\neq k} \lvert\mathrm{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)\rvert$	Centroid	$\mathrm{Var}(\mathbf{X}_j\mathbf{a}_j)=1, \quad j=1,...,J$	$\tau_j=0,\ j=1,...,J$
SUMCOV	$\sum_{j,k,j\neq k} \mathrm{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$	Horst	$\lVert\mathbf{a}_j\rVert=1, \quad j=1,...,J$	$\tau_j=1,\ j=1,...,J$
SSQCOV	$\sum_{j,k,j\neq k} \mathrm{Cov}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)$	Factorial	$\lVert\mathbf{a}_j\rVert=1, \quad j=1,...,J$	$\tau_j=1,\ j=1,...,J$
SABSCOV	$\sum_{j,k,j\neq k} \lvert\mathrm{Cov}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k)\rvert$	Centroid	$\lVert\mathbf{a}_j\rVert=1, \quad j=1,...,J$	$\tau_j=1,\ j=1,...,J$

Table 2: Special cases of RGCCA for the two block situation

Method	Criterion	Constraints	Values of τ_1 and τ_2
Inter-Battery Factor Analysis (≈ PLS regression)	Maximize $\mathrm{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$	$\lVert\mathbf{a}_1\rVert=1$ $\lVert\mathbf{a}_2\rVert=1$	$\tau_1=1$ $\tau_2=1$
Canonical Correlation Analysis	Maximize $\mathrm{Cor}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)$	$\mathrm{Var}(\mathbf{X}_1\mathbf{a}_1)=1$ $\mathrm{Var}(\mathbf{X}_2\mathbf{a}_2)=1$	$\tau_1=0$ $\tau_2=0$
Redundancy analysis of $\mathbf{X}_1$ with respect to $\mathbf{X}_2$	Maximize $\mathrm{Cor}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)\mathrm{Var}(\mathbf{X}_1\mathbf{a}_1)^{1/2}$	$\lVert\mathbf{a}_1\rVert=1$ $\mathrm{Var}(\mathbf{X}_2\mathbf{a}_2)=1$	$\tau_1=1$ $\tau_2=0$
Redundancy analysis of $\mathbf{X}_2$ with respect to $\mathbf{X}_1$	Maximize $\mathrm{Cor}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_2\mathbf{a}_2)\mathrm{Var}(\mathbf{X}_2\mathbf{a}_2)^{1/2}$	$\mathrm{Var}(\mathbf{X}_1\mathbf{a}_1)=1$ $\lVert\mathbf{a}_2\rVert=1$	$\tau_1=0$ $\tau_2=1$

Table 3: Special cases of RGCCA for the hierarchical model with one second order block

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

Table 4: Special cases of RGCCA for hierarchical multi-block data analysis

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 $\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J} \lvert\text{Cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})\rvert$	$\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1,\ \ j = 1,\ldots, J+1$
Generalized CCA (Carroll, 1968a,b)	$\underset{\mathbf{a}_1,\ldots,\mathbf{a}_{J+1}}{\text{Maximize}} \sum_{j=1}^{J_1} \text{Cor}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1}) + \sum_{j=J_1+1}^{J} \text{Cov}^2(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_{J+1}\mathbf{a}_{J+1})$	$\text{Var}(\mathbf{X}_j\mathbf{a}_j) = 1,\ j = 1,\ldots, J_1, J+1$ $\|\mathbf{a}_j\| = 1,\ j = J_1+1,\ldots, J$

Table 5: Comparison of factorial and centroid schemes for mode B

Criterion to be maximized	Mode B + Factorial scheme	Mode B + Centroid scheme
$\text{Cor}^2(\mathbf{y}_1, \mathbf{y}_3) + \text{Cor}^2(\mathbf{y}_2, \mathbf{y}_3)$	**0.967**	0.966
$\lvert\text{Cor}(\mathbf{y}_1, \mathbf{y}_3)\rvert + \lvert\text{Cor}(\mathbf{y}_2, \mathbf{y}_3)\rvert$	1.384	**1.386**

Table 6: Comparison of factorial and centroid schemes for new mode A

Criterion to be maximized	New mode A + Factorial scheme	New mode A + Centroid scheme
$\text{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_3\mathbf{a}_3)^2 + \text{Cov}(\mathbf{X}_2\mathbf{a}_2, \mathbf{X}_3\mathbf{a}_3)^2$ *	**3.8711**	3.8676
$\lvert\text{Cov}(\mathbf{X}_1\mathbf{a}_1, \mathbf{X}_3\mathbf{a}_3)\rvert + \lvert\text{Cov}(\mathbf{X}_2\mathbf{a}_2, \mathbf{X}_3\mathbf{a}_3)\rvert$ *	2.6952	**2.6964**

(*) with $\|\mathbf{a}_1\| = \|\mathbf{a}_2\| = \|\mathbf{a}_3\| = 1$

Table 7: Russett data: Results summary for the factorial scheme

	Mode PCA	New Mode A	Mode Ridge (optimal shrinkage)	Mode B
AVE (Agricultural Inequality)	.7432	.7225	.4954	.2696
AVE (Industrial Development)	.9075	.9074	.9017	.8956
AVE (Political Instability)	.5423	.5412	.5056	.4387
AVE (outer model)	.6756	.6688	.5818	.4793
AVE (inner model)	.3603	.3851	.4507	.4834