An alternative algorithm to the PLS B problem

M. Hanafi*, E.M. Qannari

ENITIAA/INRA, Unité de Sensométrie et de Chimiométrie, La Géraudière, B.P. 82225, Nantes 44322, Cedex, France

Received 19 September 2003; received in revised form 19 September 2003

Abstract

The PLS B algorithm is used as a tool to investigate the relationships among several data sets. The first step of this algorithm entails an iterative procedure the convergence of which is not proved. In this paper, an alternative procedure is discussed. It achieves the same purpose and its convergence is proven.

© 2003 Elsevier B.V. All rights reserved.

Keywords: PLS path modelling; The PLS B algorithm; Latent variables

1. Introduction

Wold (1982, 1985) developed a general framework for data analysis which he called “soft modelling”. A soft model involves manifest variables which are directly observed and latent variables which are hidden variables that underlie the phenomenon under study and structure the network of relationships among sets of variables. The data are arranged into \( K \) sets \( \mathbf{X}_k(\mathbf{N}, p_k) \). An arrow scheme constitutes the conceptual representation of the data and the model. In this scheme, manifest variables are represented by squares and latent variables are represented by circles connected with arrows whenever these latent variables are assumed to be related. Moreover, the investigator should set up a conceptual specification in the arrow scheme by representing the hypothetical relationships and the expectations regarding whether the correlations are positive or negative between the latent variables and the manifest variables.

The estimation of the parameters of the model in the PLS algorithm consists in three stages. The first stage involves an iterative procedure in the course of which ordinary
least squares (OLS) is undergone. The two next stages consist in classical regression operations. These latter stages naturally have the well-known optimality property of OLS regressions: “Minimum distance property of orthogonal projections”, and do not have any difficulty as regards the computation.

We will focus on the first stage of the estimation algorithm. This stage involves, in particular, an algorithm called PLS B which entails an iterative procedure whose convergence is not proved as stated by Wold (1982, p. 2) “For multi-blocks soft models the convergence has not been proved”.

The development of a numerical and (hopefully) simple procedure to compute solutions to the PLS B problem is a main issue from both theoretical and practical standpoints. Indeed, in the PLS approach, the first stage is of paramount importance as the efficiency of the overall strategy depends on the results obtained from this stage. Moreover, by proposing a simple procedure, we hopefully aid insight more deeply into the method of analysis and shed some light on its connections with other approaches such as generalized canonical correlation analysis (Horst, 1961). The main focus of this paper is to discuss such a procedure.

2. The PLS B problem

In the following, the $K \geq 2$ sets of manifest variables, supposed throughout this paper to be mean centred, are denoted $X_1, X_2, \ldots, X_K$ and the latent variables are denoted $\xi_1, \xi_2, \ldots, \xi_K$. The specifications of the model as designed in the arrow scheme can be summed up by means of two $(K; K)$ matrices $\Theta = [\theta_{ij}], \Delta = [\delta_{ij}]$, and a vector $\Pi = [\pi_j]$, where $i,j = 1, 2, \ldots, K$, namely:

$$\theta_{ij} = \begin{cases} 
0 & \text{if } i = j \text{ or if the latent variables } \xi_i \text{ and } \xi_j \text{ are not linked,} \\
1 & \text{otherwise,}
\end{cases}$$

$$\delta_{ij} = \begin{cases} 
-1 & \text{if } i \neq j \text{ and if } \xi_i \text{ and } \xi_j \text{ are supposed to be negatively correlated,} \\
1 & \text{otherwise.}
\end{cases}$$

As to vector $\Pi$, it is defined as follows: $\pi_j = 1$, if the correlations between the manifest variables in $X_j$ and their associated latent variable $\xi_j$ are mainly positive as specified in the model, otherwise $\pi_j = -1$.

The PLS B problem concerns the estimation of the case values (loadings) $v_k$ associated with latent variables, $\zeta_k = X_k v_k$ ($k = 1, 2, \ldots, K$), with the constraints that these latent variables are of unit variance. By denoting

$$s_k = \sum_{j=1, j \neq k}^{K} \theta_{kj} \delta_{kj} \pi_j X_j v_j, \quad k = 1, 2, \ldots, K, \quad (1)$$

the vectors $v_k$ ($k = 1, 2, \ldots, K$) are solutions to the following system:

$$(X_k' X_k)^{-1} X_k' s_k = \lambda_k \pi_k v_k, \quad k = 1, 2, \ldots, K, \quad (2)$$
with \( \lambda_k = \sqrt{b_k^Tb_k/N} \), \( b_k = X_k(X_k'X_k)^{-1}X_k's_k \), \( k = 1, 2, \ldots, K \), and \( N \) is the number of observations (rows of \( X_k \)).

An explicit solution to Eq. (2) is actually unknown, therefore an iterative procedure appears to be necessary. Wold (1982, p. 2) proposed an iterative algorithm to solve this system but its convergence is not granted. We propose a convergent procedure which provides solutions to the PLS B problem. It consists in two stages. The first stage entails a new transformation of the PLS B problem, the second stage consists in the introduction of an optimization problem which leads to solutions to the problem. More precisely, we show that finding solutions to the PLS B problem is equivalent to finding solutions to an optimization problem which consists in maximizing an objective function. A solution to this optimization problem is given by means of an iterative algorithm in the course of which the objective function is monotonically increased.

3. Equivalent statement to the PLS B problem

In order to take account of the constraints which impose to the latent variables to be of unit variance, we suggest that the above Eq. (2) can be expressed in terms of vectors \( u_j \) defined as

\[
 u_j = \left( \frac{X_j'X_j}{N} \right)^{1/2} v_j, \quad j = 1, 2, \ldots, K.
\]

It is easy to check that vectors \( u_j \) satisfy the constraints

\[
 u_j'u_j = 1, \quad j = 1, 2, \ldots, K.
\]

From Eq. (1), \( s_k \) can be expressed in terms of \( u_j \) as follows:

\[
 s_k = \frac{\sqrt{N}}{K} \sum_{j=1,j\neq k}^{K} \pi_j \theta_{kj} \delta_{kj} P_j u_j, \quad k = 1, 2, \ldots, K,
\]

where \( P_j = X_j(X_j'X_j)^{-1/2} \), \( j = 1, 2, \ldots, K \).

Multiplying Eq. (2) by \( \pi_k (X_k'X_k/N)^{1/2} \) and recalling that \( \pi_k^2 = 1 \), it follows

\[
 \frac{\pi_k P_j' s_k}{\sqrt{N}} = \lambda_k u_k, \quad k = 1, 2, \ldots, K.
\]

By replacing \( s_k \) by its expression given in (3), we are led to system (S):

\[
 \begin{align*}
 (S) & \left\{ \begin{array}{l}
 \sum_{j=1,j\neq k}^{K} \pi_k \pi_j \theta_{kj} \delta_{kj} P_j' P_k u_j = \lambda_k u_k, \\
 u_k'u_k = 1.
 \end{array} \right. \\
 & \quad k = 1, 2, \ldots, K.
\end{align*}
\]

In the following section we discuss a procedure of finding solutions to system (S) which will naturally lead to solutions to the PLS B problem.
4. Optimization problem associated with the PLS B problem

Let \( A \) be the matrix formed of blocks \( A_{kj}, \ k, j = 1, 2, \ldots, K \), defined by
\[
A_{kj} = \begin{cases} 
(\pi_k \pi_j \delta_{kj})P_k' P_j, & k \neq j, \\
0, & k = j.
\end{cases}
\]

The following property constitutes the backbone of the procedure which will be proposed to solve system (S). Let us consider the vector \( u' = [u'_1 u'_2 \ldots u'_K] \) with \( u'_k u_k = 1, \ k = 1, 2, \ldots, K \). Then a vector which is solution to the following optimization problem is also solution to system (S)
\[
\text{Maximize} \quad u' A u \quad \text{subject to} \quad u'_k u_k = 1, \quad k = 1, 2, \ldots, K.
\]

To prove this property consider the Lagrange multipliers associated with the constraints \( u'_k u_k = 1, \ k = 1, 2, \ldots, K \). Then, it is easy to see that developing the Lagrange equations associated with the above optimization problem exactly leads to system (S).

Clearly, matrix \( A \) is symmetric but not necessarily positive definite. However, it can be seen that the optimization problem remains unchanged if we add to matrix \( A \) a matrix \( mI \) to make it positive definite, where \( m \) is a positive scalar and \( I \), the identity matrix. Indeed, there is an equivalence between the maximization of the two objective functions \( u' A u \) and \( u' (A + mI) u \) under the constraints \( u'_k u_k = 1, \ k = 1, 2, \ldots, K \) which imply that \( u' u = K \). In the following, matrix \( A \) will be considered as positive definite.

5. Convergent procedure for finding solutions to the PLS B problem

The principle of the algorithm used to find solutions to the PLS B problem consists in starting with an initial solution for vector \( u' = [u'_1 u'_2 \ldots u'_K] \) subject to the constraints \( u'_k u_k = 1, \ k = 1, 2, \ldots, K \) and iteratively updating this vector in such a way as to increase the objective function in the optimization problem, namely \( u' A u \). At the current stage of this algorithm, we consider vector \( \tilde{u}' = [\tilde{u}'_1 \tilde{u}'_2 \ldots \tilde{u}'_K] \) subject to the constraints \( \tilde{u}'_k \tilde{u}_k = 1, \ k = 1, 2, \ldots, K \). We have
\[
\tilde{u}' A \tilde{u} = \sum_{k=1}^{K} \tilde{u}'_k \sum_{j=1}^{K} A_{kj} \tilde{u}_j.
\]

Let us consider vector \( \tilde{u}' = [\tilde{u}'_1 \tilde{u}'_2 \ldots \tilde{u}'_K] \) given by
\[
\tilde{u}'_k = \frac{\sum_{j=1}^{K} A_{kj} \tilde{u}_j}{\| \sum_{j=1}^{K} A_{kj} \tilde{u}_j \|}, \quad k = 1, 2, \ldots, K.
\]

By applying the Cauchy–Schwartz inequality, it follows readily from the expression of \( \tilde{u}' A \tilde{u} \) given above that
\[
\tilde{u}' A \tilde{u} \leq \sum_{k=1}^{K} \tilde{u}'_k \sum_{j=1}^{K} A_{kj} \tilde{u}_j = \tilde{u}' A \tilde{u}.
\]
By applying the Cauchy–Schwartz inequality again, we have
\[ \bar{u}^T A \bar{u} \preceq \tilde{u}^T A \tilde{u} \leq \sqrt{\bar{u}^T A \bar{u}} \sqrt{\tilde{u}^T A \tilde{u}}. \]
Therefore,
\[ \bar{u}^T A \bar{u} \preceq \tilde{u}^T A \tilde{u}. \]
This proves that using vector \( \tilde{u} \) in place of vector \( \bar{u} \) results in an increase in the objective function.

In the course of the iterative algorithm which emerges from this property, the objective function is monotonically increased until convergence is achieved at a stationary point. The procedure can be summed up as follows:

**Step 0:** Start with a vector \( u' = [u'_1, u'_2, \ldots, u'_K] \) with \( u'_k u_k = 1, k = 1, 2, \ldots, K \).

**Step 1:** Set,
\[ t_k = \sum_{j=1}^{K} A_{kj} u_j, \quad k = 1, 2, \ldots, K; \]
\[ \mu_k = \sqrt{t_k^T t_k}, \quad k = 1, 2, \ldots, K. \]
**Step 2:** Update vector \( u_k \):
\[ u_k = \frac{t_k}{\mu_k}, \quad k = 1, 2, \ldots, K. \]
The process starting from step 1 is repeated until convergence is reached.

6. Conclusion

The procedure discussed in this paper provides a solution to the PLS B problem. The advantage of this solution over Wold’s procedure is three fold: (i) the convergence of the proposed algorithm is granted, (ii) it consists in a simple procedure based on an optimization criterion, (iii) it makes it possible to assess the overall goodness of fit of the model, thus allowing the comparison of several models.

A simulation study was undertaken and it showed that the convergence of the algorithm was very quick.

More investigation is needed in order to establish the connections of the method of analysis related to PLS B algorithm with other approaches such as generalized canonical analysis. We believe that we paved the way for such an undertaking by expressing the problem on the basis of an objective function to be maximized.

References