

Non-Metric Partial Least Squares

Giorgio Russolillo

▶ **To cite this version:**

Giorgio Russolillo. Non-Metric Partial Least Squares. Electronic Journal of Statistics , 2012, 6, 10.1214/12-EJS724 . hal-03988559

HAL Id: hal-03988559
https://cnam.hal.science/hal-03988559

Submitted on 14 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Electronic Journal of Statistics
Vol. 6 (2012) 1641–1669
ISSN: 1935-7524
DOI: 10.1214/12-EJS724

Non-Metric Partial Least Squares

Giorgio Russolillo

Laboratoire CEDRIC,
Conservatoire National des Arts et Métiers
292 rue Saint Martin 75141 Paris - France
e-mail: giorgio.russolillo@cnam.fr

Abstract: In this paper I review covariance-based Partial Least Squares (PLS) methods, focusing on common features of their respective algorithms and optimization criteria. I then show how these algorithms can be adjusted for use as optimal scaling tools. Three new PLS-type algorithms are proposed for the analysis of one, two or several blocks of variables: the Non-Metric NIPALS, the Non-Metric PLS Regression and the Non-Metric PLS Path Modeling, respectively. These algorithms extend the applicability of PLS methods to data measured on different measurement scales, as well as to variables linked by non-linear relationships.

AMS 2000 subject classifications: Primary 62-07, 62H25.
Keywords and phrases: Optimal Scaling, NIPALS, PLS Regression, PLS Path Modeling, non-linearity.

Received April 2012.

Contents

1 Introduction 1642
2 A brief overview of PLS algorithms 1642
2.1 PLS approach to PCA: the NIPALS algorithm 1643
2.2 PLS approach to multivariate regression: the PLS Regression . 1644
2.3 PLS approach to Structural Equation Models: the PLS Path Modeling 1645
2.4 A bridge between the three PLS approaches 1648
3 Handling non-metric data in PLS: the statement of the problem . . 1649
4 Non-Metric PLS approach 1650
5 Non-Metric NIPALS 1653
5.1 Links with other non-metric approaches to PCA 1655
6 Non-Metric PLS Regression 1655
7 Non-Metric PLS Path Modeling 1659
7.1 The optimizing criterion of Non-Metric PLS Path Modeling . . 1659
8 Missing data 1661
9 An application to macroeconomic data 1662
10 Discussion 1665
Acknowledgements 1666
References 1666

1. Introduction

Partial Least Squares (PLS) methods encompass a suite of data analysis techniques based on algorithms belonging to the PLS family. These algorithms consist of various extensions of the Nonlinear estimation by Iterative PArtial Least Squares (NIPALS) algorithm, which was proposed by Herman Wold [35] as an alternative algorithm for implementing a Principal Component Analysis (PCA). Wold also proposed using NIPALS to analyze relationships between several blocks of variables [36, 38]: this is the PLS approach to Structural Equation Modeling, later called PLS Path Modeling (PLS-PM) [17, 29, 4]. Svante Wold, Herman's son, noticed that the PLS approach could be used in order to implement a component-based regularized regression, called PLS-Regression (PLS-R) [39].

PLS techniques were originally devised to handle data sets forming metric spaces, where variables embedded in the analysis are observed along interval or ratio scales. In this article, variables measured at ratio or interval scale level will be referred to as *numeric* or *metric* variables. Unfortunately, in many fields where PLS methods are applied (e.g. genomics, sensory analysis, consumer analysis, marketing) researchers are also interested in analyzing sets of variables measured on a non-metric scale, i.e. ordinal and nominal variables. This article focuses on new methodological proposals to enable PLS techniques able to handle both metric and non-metric variables at once. The next section begins by presenting NIPALS, PLS-R and PLS-PM, before integrating their algorithmic flows and correspondent criteria into a unitary framework. Section 3 explains why PLS methods cannot handle non-metric variables and reviews the remedies for this drawback proposed in the literature to date. Subsequent sections show that PLS iteration can indeed be modified to give PLS-type algorithms, called Non-Metric PLS (NM-PLS), which also work as optimal scaling methods.

2. A brief overview of PLS algorithms

NIPALS, PLS Regression and PLS Path Modeling (when *Mode A* option in the estimation process is used, see section 2.3) algorithms share one important feature. Their iterative loops converge to criteria strongly related to the covariance. Given one set of variables, NIPALS loop maximizes the covariance of a weighted aggregate of the variables (called component, or latent variable) with itself (i.e. its variance); given two sets of variables, PLS-R loop maximizes the covariance of a component in a set with another component belonging to the other set; finally, given several sets of variables, PLS-PM loop maximizes the covariance between components belonging to the different sets. The choice to focus on non-metric extensions of PLS algorithms oriented to the maximization of covariance between components, in contrast to PLS algorithms that work on correlation-based criteria, relies on two main reasons. First of all, non-metric extension of a method necessarily requires a parametric expansion that may affect the variability of the estimates. This drawback is greatly reduced by working on

covariance-based criteria, which yield more robust estimates. The second reason, as will be shown in the rest of the paper, is that in these frameworks optimal scaling parameter estimates are clearly interpretable.

NIPALS, PLS Regression and PLS Path Modeling algorithms are described in detail below. Since PLS algorithms work on centered variables, variables are assumed to be centered throughout the rest of this section.

2.1. PLS approach to PCA: the NIPALS algorithm

The NIPALS algorithm performs a PCA on a set $\boldsymbol{X} = [\boldsymbol{x}_1 \dots \boldsymbol{x}_p \dots \boldsymbol{x}_P]$ of variables observed on N individuals. The peculiarity of this algorithm is that it calculates principal components by means of iterated simple ordinary least squares (OLS) regressions. This feature overcomes computational problems caused by missing data and landscape data matrices, i.e. matrices with more columns than rows.

The NIPALS algorithm starts choosing an arbitrary centered vector $\boldsymbol{t}_{(1)}^{(0)}$, that is a first approximation of the first principal component. Next, a loop which at the s-th iteration ($s = 0, 1, \dots$) alternately computes a weight vector $\boldsymbol{w}_{(1)}^{(s)}$ and a score vector $\boldsymbol{t}_{(1)}^{(s+1)}$ is repeated. The loop converges [19] to the first principal component $\boldsymbol{t}_{(1)}$ and to vector $\boldsymbol{w}_{(1)}$ of weights (or loadings), which maximizes the PCA criterion

$$\underset{||\boldsymbol{w}_{(1)}||=1}{\arg\max}\ \mathrm{var}(\boldsymbol{X}\boldsymbol{w}_{(1)}). \tag{1}$$

It is well known that the eigenvector associated to the dominant eigenvalue of the covariance matrix $(1/N)\boldsymbol{X}'\boldsymbol{X}$ is the solution to this problem. Since the dominant eigenvalue of this matrix equals the sum of the squared covariances between the variables and the first principal component, when variables are standardized, criterion (1) can be rewritten as:

$$\underset{||\boldsymbol{w}_{(1)}||=1}{\arg\max} \sum_p \mathrm{cor}^2(\boldsymbol{x}_p, \boldsymbol{X}\boldsymbol{w}_{(1)}). \tag{2}$$

The second component $\boldsymbol{t}_{(2)}$ is extracted by iterating the same loop on matrix $\boldsymbol{E}_1$, which is obtained by deflating $\boldsymbol{X}$. The deflation step consists of calculating $\boldsymbol{E}_1$ as the residual matrix of the regression of $\boldsymbol{X}$-variables on $\boldsymbol{t}_{(1)}$. In general, for each h ($1 \leq h \leq H \leq rank(\boldsymbol{X})$), the NIPALS algorithm maximizes the criterion

$$\underset{||\boldsymbol{w}_{(h)}||=1}{\arg\max}\ \mathrm{var}(\boldsymbol{E}_{(h-1)}\boldsymbol{w}_{(h)}) \tag{3}$$

by implementing the loop described in Step 1 of Algorithm 1 on $\boldsymbol{E}_{(h-1)}$ (with $\boldsymbol{E}_{(0)} = \boldsymbol{X}$).

Algorithm 1 NIPALS algorithm

Input: $\boldsymbol{E}_{(0)} = \boldsymbol{X}$
Output: $\boldsymbol{W}=[\boldsymbol{w}_{(1)},\ldots,\boldsymbol{w}_{(H)}]$, $\boldsymbol{T}=[\boldsymbol{t}_{(1)},\ldots,\boldsymbol{t}_{(H)}]$
for all $h = 1,\ldots,H$ **do**
  Step 0: **Initialize** $\boldsymbol{t}_{(h)}^{(0)}$
  Step 1:
  repeat
    Step 1.1: $\boldsymbol{w}_{(h)}^{(s)} = \boldsymbol{E}_{(h-1)}^{'}\boldsymbol{t}_{(h)}^{(s)}/(\boldsymbol{t}_{(h)}^{(s)'}\boldsymbol{t}_{(h)}^{(s)})$
    Step 1.2: $\boldsymbol{w}_{(h)}^{(s)} = \boldsymbol{w}_{(h)}^{(s)}/\|\boldsymbol{w}_{(h)}^{(s)}\|$
    Step 1.3: $\boldsymbol{t}_{(h)}^{(s+1)} = \boldsymbol{E}_{h-1}\boldsymbol{w}_{(h)}^{(s)}$
  until convergence of $\boldsymbol{w}_{(h)}$
  Step 2: $\boldsymbol{E}_{(h)} = \boldsymbol{E}_{(h-1)} - \boldsymbol{t}_{(h)}\boldsymbol{w}_{(h)}^{'}$
end for

2.2. PLS approach to multivariate regression: the PLS Regression

Generally speaking, PLS-R involves a set $\boldsymbol{X}_2 = [\boldsymbol{x}_{12}\ldots\boldsymbol{x}_{p2}\ldots\boldsymbol{x}_{P_2 2}]$ of response variables and a set $\boldsymbol{X}_1 = [\boldsymbol{x}_{11}\ldots\boldsymbol{x}_{p1}\ldots\boldsymbol{x}_{P_1 1}]$ of predictors. PLS-R extracts a suite of orthogonal components in the predictor space intended both to explain predictors and to predict response variables.

In the PLS-R algorithm a NIPALS-type loop is iterated to calculate model parameters for each component. Two slightly different algorithms implementing the PLS-R are found in the literature, described in two works by Tenenhaus [28] and Höskuldsson [8] respectively; here Höskuldsson's approach is considered. To compute the first component, the following procedure is repeated upon convergence: proxies of $\boldsymbol{X}_2$-scores ($\boldsymbol{t}_{2(1)}$), $\boldsymbol{X}_2$-weights ($\boldsymbol{w}_{2(1)}$), $\boldsymbol{X}_1$-scores ($\boldsymbol{t}_{1(1)}$) and $\boldsymbol{X}_1$-weights ($\boldsymbol{w}_{1(1)}$) are sequentially calculated each one being a function of the previous proxy. The sequences of proxies $\boldsymbol{w}_{1(1)}^{(s)}$ and $\boldsymbol{w}_{2(1)}^{(s)}$ obtained at the s-th iteration converge to the dominant eigenvectors of matrices $(1/N^2)\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{X}_1'\boldsymbol{X}_1$ and $(1/N^2)\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{X}_1'\boldsymbol{X}_2$, respectively [8]. Therefore, when convergence is reached, vectors $\boldsymbol{w}_{1(1)}$ and $\boldsymbol{w}_{2(1)}$ satisfy the criterion

$$\underset{||\boldsymbol{w}_{1(h)}||=||\boldsymbol{w}_{2(1)}||=1}{\arg\max} \operatorname{cov}^2(\boldsymbol{X}_2\boldsymbol{w}_{2(1)}, \boldsymbol{X}_1\boldsymbol{w}_{1(1)}) \tag{4}$$

The dominant eigenvectors of $(1/N^2)\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{X}_1'\boldsymbol{X}_2$ and $(1/N^2)\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{X}_1'\boldsymbol{X}_2$ coincide respectively with the dominant right and left singular vectors of matrix $(1/N)\boldsymbol{X}_2'\boldsymbol{X}_1$. Therefore, vectors $\boldsymbol{w}_{1(1)}$ and $\boldsymbol{w}_{2(1)}$ are also solutions of the criterion

$$\underset{||\boldsymbol{w}_{1(1)}||=||\boldsymbol{w}_{2(1)}||=1}{\arg\max} \operatorname{cov}(\boldsymbol{X}_2\boldsymbol{w}_{2(1)}, \boldsymbol{X}_1\boldsymbol{w}_{1(1)}) \tag{5}$$

This criterion is common to a number of PLS approaches to two-block cross-covariance analysis, such as SIMPLS [9], PLS canonical analysis [38, 28] and various approaches, cited in the literature under various names [26, 21, 34,

32], which may be traced back to Inter-Battery Analysis [33]. However, these approaches differ between themselves and from the PLS-R in their successive components, which are obtained using different types of deflation.

In PLS-R, both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are deflated as a function of $\boldsymbol{t}_{1(1)}$, as shown in detail in Steps 2-5 of Algorithm 2, in order to use the second component to explain the portion of variability in responses that is not explained by the first component. Each generic successive component of order $h > 1$ ($1 \leq h \leq H \leq rank(\boldsymbol{X}_1)$) is then calculated by iterating the same loop (Steps 1 of Algorithm 2) on the deflated matrices $\boldsymbol{E}_{(h-1)}$ and $\boldsymbol{F}_{(h-1)}$ respectively (with $\boldsymbol{E}_0 = \boldsymbol{X}_1$, $\boldsymbol{F}_0 = \boldsymbol{X}_2$).

After extracting a suitable number of components, chosen by cross-validation, a standard regression equation can be calculated in order to explain the responses as a linear combination of the original predictor variables.

Algorithm 2 PLS-R algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}_1, \boldsymbol{F}_0 = \boldsymbol{X}_2$
Output: $\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{T}_1, \boldsymbol{T}_2, \boldsymbol{P}, \boldsymbol{Q}$
for all $h = 1, \ldots, H$ **do**
  Step 0: **Initialize** $\boldsymbol{t}_{2(h)} = \boldsymbol{t}_{2(h)}^{(0)}$
  Step 1:
  repeat
    Step 1.1: $\boldsymbol{w}_{1(h)}^{(s)} = \boldsymbol{E}'_{(h-1)}\boldsymbol{t}_{2(h)}^{(s)} / \|\boldsymbol{E}'_{(h-1)}\boldsymbol{t}_{2(h)}^{(s)}\|$
    Step 1.2: $\boldsymbol{t}_{1(h)}^{(s)} = \boldsymbol{E}_{(h-1)}\boldsymbol{w}_{1(h)}^{(s)}$
    Step 1.3: $\boldsymbol{w}_{2(h)}^{(s)} = \boldsymbol{F}'_{(h-1)}\boldsymbol{t}_{1(h)}^{(s)} / \|\boldsymbol{F}'_{(h-1)}\boldsymbol{t}_{1(h)}^{(s)}\|$
    Step 1.4: $\boldsymbol{t}_{2(h)}^{(s+1)} = \boldsymbol{F}_{(h-1)}\boldsymbol{w}_{2(h)}^{(s)}$
  until convergence of $\boldsymbol{w}_{1(h)}$
  Step 2: $\boldsymbol{p}_{(h)} = \boldsymbol{E}'_{(h-1)}\boldsymbol{t}_{1(h)} / (\boldsymbol{t}'_{1(h)}\boldsymbol{t}_{1(h)})$
  Step 3: $b_{(\boldsymbol{t}_{2(h)}|\boldsymbol{t}_{1(h)})} = \boldsymbol{t}'_{2(h)}\boldsymbol{t}_{1(h)} / (\boldsymbol{t}'_{1(h)}\boldsymbol{t}_{1(h)})$
  Step 4: $\boldsymbol{E}_{(h)} = \boldsymbol{E}_{(h-1)} - \boldsymbol{t}_{1(h)}\boldsymbol{p}'_{(h)}$
  Step 5: $\boldsymbol{F}_{(h)} = \boldsymbol{F}_{(h-1)} - b_{(\boldsymbol{t}_{2(h)}|\boldsymbol{t}_{1(h)})}\boldsymbol{t}_{1(h)}\boldsymbol{w}'_{2(h)}$
end for

2.3. PLS approach to Structural Equation Models: the PLS Path Modeling

Structural Equation Models (SEMs) [2] describe and estimate conceptual structures where Q latent variables (LVs) $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_q, \ldots, \boldsymbol{\xi}_Q$, linked to each other by a network of linear relationships (usually highlighted by means of a path diagram), are measured by Q sets of manifest variables (MVs) $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_q, \ldots, \boldsymbol{X}_Q$ observed on N observations. Each SEM has two levels of relationships. The first level concerns relationships between LVs (structural model); these relations can be stored in an $Q \times Q$ lower triangular binary matrix $\boldsymbol{L}$, whose generic element $l_{qq'}$ is one if $\boldsymbol{\xi}_q$ depends on $\boldsymbol{\xi}_{q'}$ and zero otherwise. The other level of relations concerns the links between each LV and its own block of MVs (measurement model). According to Lohmöller [17], the PLS approach to SEM is a predictive

Path Modeling procedure oriented towards reconstruction of the data matrix, in contrast to the covariance-based approach to SEM by Jöreskog [15], which models relationships between the variables. The PLS-PM estimation procedure is described in detail in Algorithm 3. First, a PLS loop is iterated to obtain LV scores through a system of multiple and simple linear regressions (Steps 1-2). Next, structural (or path) coefficients are estimated through OLS multiple/simple regressions on the LV estimates $\hat{\boldsymbol{\xi}}_q$, according to the path diagram structure (Step 3). In the iterative part of the algorithm (Step 1) outer weights w_{pq} $(p = \{1, \ldots, P_q\})$, linking each MV $\boldsymbol{x}_{pq}$ to the corresponding LV, are estimated by a procedure in which the final estimates of the latent variables are obtained through the alternation of their *outer* and *inner* estimations. In the outer estimation phase, a proxy $\boldsymbol{t}_q$ for each LV is computed as a linear combination of its own MVs (Step 1.3-1.4). In the inner estimation phase, a proxy $\boldsymbol{z}_q$ for each LV is computes as a linear combination of the connected LVs, according to the path diagram structure (Steps 1.1-1.2). In Wold's original algorithm (the one presented in Algorithm 3), inner estimate $\boldsymbol{z}_q^{(s)}$ at the s-th iteration is a function of $\boldsymbol{t}_{q'}^{(s+1)}$ if $q^{'} < q$ and is a function of $\boldsymbol{t}_{q'}^{(s)}$ if $q^{'} > q$, where q' indexes LVs connected to ξ_q in the path diagram. Wold's original algorithm [38, 6] was simplified by Lohmöller [17]. In Lohmöller's algorithm, $\boldsymbol{z}_q^{(s)}$ is always a function of $\boldsymbol{t}_{q'}^{(s)}$. This procedure, used in most commercial softwares, is easier to implement, but it takes longer to converge.

The PLS-PM algorithm is extremely flexible. Various options, providing slightly different results, can be used to compute weights for inner and outer estimates of the LVs. There are three ways to calculate the inner weights $(e_{qq'})$:

1. the centroid scheme (Wold's original scheme), where $e_{qq'}$ is equal to the sign of the correlation between $\boldsymbol{t}_q$ and $\boldsymbol{t}_{q'}$;
2. the factorial scheme (the Lohmöller scheme), where $e_{qq'}$ is equal to the correlation between $\boldsymbol{t}_q$ and $\boldsymbol{t}_{q'}$;
3. the path weighting scheme, where, for each $\boldsymbol{\xi}_q$: if $\boldsymbol{\xi}_{q'}$ is a latent predictor of $\boldsymbol{\xi}_q$, $e_{qq'}$ is equal to the coefficient of $\boldsymbol{t}_{q'}$ in the regression of $\boldsymbol{t}_q$ on the inner estimates of its latent predictors; if $\boldsymbol{\xi}_{q'}$ is a latent response variable of $\boldsymbol{\xi}_q$, $e_{qq'}$ is equal to the correlation between $\boldsymbol{t}_q$ and $\boldsymbol{t}_{q'}$.

Two main methods exist to estimate weights in the outer estimation process: *Mode A* (for outwards directed or reflective blocks) and *Mode B* (for inwards directed or formative blocks). In *Mode B* a multiple regression model defines the relation between the latent and manifest variables. The outer weights are thus the regression coefficients of a multiple regression model of the inner estimate of each LV on its own MVs. In *Mode A* each inner estimate is modeled as a predictor of corresponding MVs. Hence, the outer weights are the coefficients of the simple regressions of each MV on the corresponding LV inner estimate. Therefore, in *Mode A* PLS-PM, the generic outer weight w_{pq} is a measure of the linear relationship between each $\boldsymbol{x}_{pq}$ and $\boldsymbol{z}_q$, which is a weighted sum of the outer estimates $\boldsymbol{t}_{q'}$ of the LVs linked to $\boldsymbol{\xi}_q$.

The rest of this discussion focuses on *Mode A* PLS-PM, i.e. on the case where

Algorithm 3 Wold's PLS Path Modeling algorithm.

Input: $\boldsymbol{X}_1 \ldots \boldsymbol{X}_Q, \boldsymbol{L}, \boldsymbol{C}$
Output: $\boldsymbol{\beta}_1 \ldots \boldsymbol{\beta}_J$, $\boldsymbol{w}_1 \ldots \boldsymbol{w}_Q$, $\hat{\boldsymbol{\xi}}_1 \ldots \hat{\boldsymbol{\xi}}_Q$
Step 0: Initialize $\boldsymbol{t}_q = \boldsymbol{t}_q^{(0)}$ **for all** $q = 1, \ldots, Q$
Step 1: Iteration
repeat
  for all $q = 1, \ldots, Q$ **do**
    Step 1.1: Computation of the inner weights
    if $q < q^{'}$ **then**
      $e_{qq'}^{(s)} = f\left(\boldsymbol{t}_q^{(s)}, \boldsymbol{t}_{q'}^{(s)}\right)$
    else
      $e_{qq'}^{(s)} = f\left(\boldsymbol{t}_q^{(s)}, \boldsymbol{t}_{q'}^{(s+1)}\right)$
    end if
    Step 1.2: Inner estimation of the LVs
    $\boldsymbol{z}_q^{(s)} = \sum_{q'<q}\left(c_{qq'} e_{qq'}^{(s)} \boldsymbol{t}_{q'}^{(s+1)}\right) + \sum_{q'>q}\left(c_{qq'} e_{qq'}^{(s)} \boldsymbol{t}_{q'}^{(s)}\right)$
    Step 1.3: Computation of the outer weights
    $$\begin{cases} \boldsymbol{w}_q^{(s)} = (\boldsymbol{X}_q^{'}\boldsymbol{X}_q)^{-1}\boldsymbol{X}_q^{'}\boldsymbol{z}_q^{(s)} & \text{if Mode B} \\ \boldsymbol{w}_q^{(s)} = (1/N)\boldsymbol{X}_q^{'}\boldsymbol{z}_q^{(s)} & \text{if Mode A} \\ \boldsymbol{w}_q^{(s)} = \boldsymbol{X}_q^{'}\boldsymbol{z}_q^{(s)}/\|\boldsymbol{X}_q^{'}\boldsymbol{z}_q^{(s)}\| & \text{if new Mode A} \end{cases}$$
    Step 1.4: Outer estimation of the LVs
    $$\begin{cases} \boldsymbol{t}_q^{(s+1)} = \boldsymbol{X}_q\boldsymbol{w}_q^{(s)} & \text{if new Mode A} \\ \boldsymbol{t}_q^{(s+1)} = (N^{1/2}\boldsymbol{X}_q\boldsymbol{w}_q^{(s)})/\|\boldsymbol{X}_q\boldsymbol{w}_q^{(s)}\| & \text{otherwise} \end{cases}$$
  end for
until convergence of $\boldsymbol{w}_q$
Step 2: Computation of the LVs
for all $q = 1, \ldots, Q$ **do**
  $\hat{\boldsymbol{\xi}}_q = (N^{1/2}\boldsymbol{X}_q\boldsymbol{w}_q)/\|\boldsymbol{X}_q\boldsymbol{w}_q\|$
end for
Step 3: Computation of the Path Coefficients
for all $j = 1, \ldots, J$ **do**
  $\boldsymbol{\beta}_j = \left(\hat{\boldsymbol{\Xi}}'_{\to j}\hat{\boldsymbol{\Xi}}_{\to j}\right)^{-1}\hat{\boldsymbol{\Xi}}'_{\to j}\hat{\boldsymbol{\xi}}_j$
  $\{\hat{\boldsymbol{\Xi}}'_{\to j}$ is the matrix of the latent predictors of the endogenous LV $\hat{\boldsymbol{\xi}}_j\}$
end for

Mode A is used for all the blocks. *Mode A* PLS-PM does not appear to optimize any criterion: Krämer [10] showed that Wold's *Mode A* algorithm is not based on stationary equations related to the optimization of a twice differentiable function. However, Tenenhaus and Tenenhaus [31] have recently proposed a new framework, called Regularized Generalized Canonical Component Analysis (RGCCA), where *Mode B* and a slightly adjusted *Mode A*, called *new Mode A*, are unified towards a regularization parameter τ_q ($0 \leq \tau_q \leq 1$), which provides a connection between the two. The *new Mode A* differs from the original *Mode A* in normalization constraints: outer estimates are not constrained to unitary variance, while outer weights are constrained to unitary norm (see Algorithm 3). Tenenhaus and Tenenhaus [31] also demonstrated that *new Mode A* PLS-PM, when centroid or factorial schemes are used, monotonically converges, for each

TABLE 1
Iterative steps of the NIPALS algorithm, PLS-R algorithm (first component) and PLS-PM algorithm with new Mode A option.

NIPALS loop	PLS-R loop	new Mode A PLS-PM loop
$\boldsymbol{w}_{(1)} = \boldsymbol{X}'(\boldsymbol{X}\boldsymbol{w}_{(1)})$	$\boldsymbol{w}_{q(1)} = \boldsymbol{X}'_q(\boldsymbol{X}_{q'}\boldsymbol{w}_{q'(1)})$	$\boldsymbol{w}_q = \boldsymbol{X}'_q(\sum_{q'} c_{qq'}e_{qq'}\boldsymbol{X}_{q'}\boldsymbol{w}_{q'})$
normalize $\boldsymbol{w}_{(1)}$	normalize $\boldsymbol{w}_{q(1)}$	normalize $\boldsymbol{w}_q$
$\boldsymbol{t}_{(1)} = \boldsymbol{X}\boldsymbol{w}_{(1)}$	$\boldsymbol{t}_{q(1)} = \boldsymbol{X}_q\boldsymbol{w}_{q(1)}$	$\boldsymbol{t}_q = \boldsymbol{X}_q\boldsymbol{w}_q$

q, to the criterion

$$\arg\max_{\boldsymbol{w}_q} \sum_{q \neq q'} c_{qq'}\, g(\mathrm{cov}(\boldsymbol{X}_q\boldsymbol{w}_q, \boldsymbol{X}_{q'}\boldsymbol{w}_{q'})) \tag{6}$$

under the constraint $||\boldsymbol{w}_q|| = 1$. Above, $c_{qq'}$ represents the generic element of the binary matrix $\boldsymbol{C} = \boldsymbol{L}' + \boldsymbol{L}$ defining the path diagram, and function $g(\cdot)$ depends on the scheme used for the inner estimation of the LVs: it represents the absolute value function for the centroid scheme and the square function for the factorial scheme. Criterion (9) can be rewritten as

$$\arg\max_{\forall \boldsymbol{w}_q} \sum_{q} \mathrm{cov}(\boldsymbol{X}_q\boldsymbol{w}_q, \boldsymbol{z}_q) \tag{7}$$

where, for each q, $||\boldsymbol{w}_q|| = 1$ [31]. As the next section will show, this criterion can be used to define a unifying framework which includes NIPALS, PLS-R and *new Mode A* PLS-PM loops, and the corresponding convergence criteria.

2.4. A bridge between the three PLS approaches

The algorithmic core of NIPALS, PLS-R and PLS-PM is the iterative process in which weight and score vectors are alternately estimated. Each iteration of a PLS algorithm can be summarized in three steps: a weight estimation step, in which weights are updated; a score estimation step, in which they are linearly combined to build the score vector(s); a normalization step, where the score vector(s) or the weight vector(s) (depending on the algorithm used) are normalized. In *Mode A* PLS-PM the normalization constraint concerns the score vectors, whereas in NIPALS, PLS-R and *new Mode A* PLS-PM loops it concerns the weight vectors. The strong similarity between these last three loops can be observed in Table 1. In PLS-PM and PLS-R the steps are repeated for each q and q' in $1, \ldots, Q$ (in PLS-R $Q = 2$) with $q \neq q'$. Moreover, in PLS-PM $c_{qq'}e_{qq'}$ is null if $\boldsymbol{\xi}_{q'}$ is not connected to $\boldsymbol{\xi}_q$; otherwise, it represents the corresponding inner weight.

In order to optimize criteria (1), (5) and (7), weights are always worked out in a way that maximizes the squared correlation of the corresponding variables with a latent construct. From now onwards I will refer to this latent construct as *Latent Criterion* (LC), in contrast to what Hayashi [7] called the *outside criterion* in his seminal paper on quantification methods. The LC is an unknown vector of order N, centered by construction. For each PLS method different LCs are considered:

- In NIPALS, the relevant LC is the first principal component $\boldsymbol{t}_{(1)} = \boldsymbol{X}\boldsymbol{w}_{(1)}$.
- In PLS-R the relevant LCs are vector scores $\boldsymbol{t}_{1(1)} = \boldsymbol{X}_1\boldsymbol{w}_{1(1)}$ for the predictor block $\boldsymbol{X}_1$ and the vector score $\boldsymbol{t}_{2(1)} = \boldsymbol{X}_2\boldsymbol{w}_{2(1)}$ for the response block $\boldsymbol{X}_2$.
- In *(new) Mode A* PLS-PM framework, there is one relevant LC for each block of manifest variables $\boldsymbol{X}_q$, i.e. the corresponding inner estimate $\boldsymbol{z}_q = \sum_{q'} c_{qq'} e_{qq'} \boldsymbol{X}_{q'} \boldsymbol{w}_{q'}$.

All of these LCs are then expressed using the generic notation $\boldsymbol{\gamma} = f(\boldsymbol{w})$. Through the concept of LC, criteria expressed by equations (1), (5) and (7) can be summed up in one general criterion expressed as a function of $\boldsymbol{\gamma}$:

$$\arg\max_{\forall \boldsymbol{w}_q} \sum_q \operatorname{cov}(\boldsymbol{X}_q\boldsymbol{w}_q, \boldsymbol{\gamma}_q) \tag{8}$$

under the constraint that $||\boldsymbol{w}_q|| = 1$. Optimization of criterion (8) yields the one-component NIPALS solution when $q \in \{1\}$, the one-component PLS-R solution when $q \in \{1, 2\}$ and the *new Mode A* PLS-PM solution when $q \in \{1, 2, \ldots, Q\}$.

3. Handling non-metric data in PLS: the statement of the problem

In order to satisfy criterion (8), in NIPALS, PLS-R and *new Mode A* PLS-PM, when working on standardized variables, optimal weights are calculated as Pearson's product-moment correlation coefficients between each variable and the LC. This leads to two basic assumptions underlying PLS models:

- Each variable is measured on a interval (or ratio) scale.
- Relationships between variables and latent constructs are linear and, consequently, monotonic.

Therefore, standard PLS methods cannot handle data measured on a scale which does not have metric properties.

There is a simple way to overcome this problem: replacement each non-metric variable by the corresponding indicator matrix. Most of the softwares currently used to perform PLS analyses use this type of coding in order to handle categorical variables, but it is not a valid solution to the problem considered here, for three reasons:

- First of all, a complete disjunctive coding conflicts with the idea of the variable as a whole, because it considers categories as if they were variables in themselves. Any PLS analysis performed on dummy variables yields a weight for each category, rather then for the whole categorical variable. Such weights measure the impact of each individual category on the latent construct, while PLS weights should measure the intensity of the relationship between the original variables and latent constructs. Binary coding makes it impossible to evaluate the importance of the whole variable in the model, or to compare the weight of one variable with the weights of the other variables.

- Secondly, the binary coding inflates the dimensionality of the data matrix, since each categorical variable is transformed into as many binary variables as the number of its categories. As a result, the measurement of a variable's overall impact on the LC is influenced by the number of its categories. Moreover, if the number of categories is very large, binary coding generates sparse matrices.
- Finally, the weight of a dummy variable representing a category mainly associated with central values of the corresponding LC score distribution is systematically underestimated, because such a binary variable is linked to the LC by a non-monotonic relationship.

To overcome the drawbacks of binary coding, a new scaling approach has been proposed in a regression framework: the PLS algorithm for CAtegorical Predictors (PLS-CAP) [24, 25]. This approach was suggested for handling nominal predictors, as well as non-linearity, in a PLS Regression. In a Path Modeling framework, Jakobowicz and Darquenne [14] proposed a modified PLS-PM algorithm, called Partial Maximum Likelihood (PML), that they recommend for cases where some block are made up of both numeric and nominal MVs. Yet no comprehensive approach seems to exist for analysis of variables measured on a variety of measurement scales in a more general PLS framework.

4. Non-Metric PLS approach

The Non-Metric approach for handling measurement heterogeneity in PLS framework (NM-PLS) is based on the concept of Optimal Scaling (OS). OS has been extensively implemented in multivariate analysis by iterative algorithms belonging to the Alternating Least Squares (ALS) family [16, 5]: for this reason, these algorithms are also called ALSOS (Alternating Least Squares approach to Optimal Scaling) [41]. The OS principle sees observations as categorical, and represents each observation category by a scaling parameter. This parameter is subject to constraints deriving from the measurement characteristics of the variables. In the ALS approach parameters are divided into two subsets: the parameters of the model and the parameters of the data (or scaling parameters). A loss function is then optimized by alternately optimizing with respect to each subset, keeping the other fixed.

NM-PLS algorithms exploit a PLS-type iteration to implement an OS procedure. This leads to a new class of PLS algorithms that generalize standard PLS methods to the treatment of non-metric variables. They are called Non-Metric PLS (NM-PLS) methods [23], because they are able to provide data with a new metric structure, that does not depend on the metric properties of the original data. In other words, NM-PLS methods provides non-metric data with a metric, and provide metric data with a new metric, making relationships between variables and latent constructs linear, as required in standard PLS models. These methods could also be named non-linear PLS methods as well, since they discard the linearity assumption intrinsic to the standard PLS methods. However,

the adjective Non-Metric is preferred because it highlights their suitability for data without a metric structure.

In the rest of this article, I will refer to a variable $\boldsymbol{x}^*$ which has been observed for N units on a given measurement scale and needs to be provided with a metric as a *raw* variable. In the OS process a scaling (numeric) value is assigned to each category (or distinct value) ϕ_k $(k = 1, \ldots, K \leq N)$ of $\boldsymbol{x}^*$, such that

- it is coherent with the chosen scaling level;
- it optimizes the model criterion.

Each raw variable is transformed as $\hat{\boldsymbol{x}} \propto \tilde{\boldsymbol{X}}\boldsymbol{\phi}$, where $\boldsymbol{\phi}' = (\phi_1, \ldots, \phi_K)$ is the vector of optimal scaling parameters and the matrix $\tilde{\boldsymbol{X}}$ defines a space in which constraints imposed by the scaling level are respected. The symbol $\propto$ means that the left side of the equation corresponds to the right side normalized to unitary variance.

In order to optimize NM-PLS model criteria, for any raw variable $\boldsymbol{x}^*$ in the model, the correspondent scaling vector must satisfy the following criterion:

$$\arg\max_{\boldsymbol{\phi}} \operatorname{cor}^2(\tilde{\boldsymbol{X}}\boldsymbol{\phi}, \boldsymbol{\gamma}_{\boldsymbol{x}^*}) \tag{9}$$

Criterion (9) is optimized by means of the ordinary least squares regression coefficients of the LC on $\tilde{\boldsymbol{X}}$, i.e. by projecting $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ on the space defined by the columns of $\tilde{\boldsymbol{X}}$. The resulting projection, normalized to unitary variance, is the geometric representation of the scaled variable $\hat{\boldsymbol{x}}$.

Three levels of scaling are considered here: nominal, ordinal and polynomial. Each level of scaling has a corresponding *ad hoc* scaling function $\mathcal{Q}$, which is the projection operator of the LC in a suitable space spanned by $\tilde{\boldsymbol{X}}$-columns. While nominal and ordinal scaling involve the quantification of numerals (i.e. numeric labels with no quantitative meaning), polynomial scaling exclusively addresses non-linearity, as it involves the transformation of a metric raw variable. The following explanations refer to $\mathcal{Q}$ as either a quantification or a transformation function, depending on the type of scaling. It is noteworthy that other smoothing procedures involving an orthogonal projection on a predefined basis (a spline basis, for instance) could be used as well.

In nominal scaling, a variable is quantified as the orthogonal projection of the LC $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ linked to $\boldsymbol{x}^*$ on the space spanned by the columns of the indicator matrix $\tilde{\boldsymbol{X}}^{\mathrm{n}}$ ("n" standing for nominal) generated by the K categories of $\boldsymbol{x}^*$:

$$\mathcal{Q}(\tilde{\boldsymbol{X}}^{\mathrm{n}}, \boldsymbol{\gamma}_{\boldsymbol{x}^*}) = \tilde{\boldsymbol{X}}^{\mathrm{n}}(\tilde{\boldsymbol{X}}^{\mathrm{n}'}\tilde{\boldsymbol{X}}^{\mathrm{n}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{n}'}\boldsymbol{\gamma}_{\boldsymbol{x}^*}. \tag{10}$$

The scaling function $\mathcal{Q}(\tilde{\boldsymbol{X}}^{\mathrm{n}}, \boldsymbol{\gamma}_{\boldsymbol{x}^*})$ maximizes (9) under the grouping constraint that, for each pair of observations i and i',

$$(x_i^* \sim x_{i'}^*) \Rightarrow (\hat{x}_i = \hat{x}_{i'}), \tag{11}$$

where the symbol $\sim$ indicates membership of the same category. The resulting scaling values for the categories of $\boldsymbol{x}^*$ are the K least squares regression coefficients of $\tilde{\boldsymbol{X}}^{\mathrm{n}}$ on $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$, which correspond to the averages of $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ conditioned to

categories of $\boldsymbol{x}^*$. The scaled variable contains the LC values predicted by the regression of $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ on $\tilde{\boldsymbol{X}}^{\mathrm{n}}$.

Ordinal scaling may involve ordinal or metric variables, since it retains the order property of $\boldsymbol{x}^*$. In ordinal scaling the following scaling function [40] is used:

$$\mathcal{Q}(\tilde{\boldsymbol{X}}^{\mathrm{o}}, \boldsymbol{\gamma}_{\boldsymbol{x}^*}) = \tilde{\boldsymbol{X}}^{\mathrm{o}}(\tilde{\boldsymbol{X}}^{\mathrm{o}'}\tilde{\boldsymbol{X}}^{\mathrm{o}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{o}'}\boldsymbol{\gamma}_{\boldsymbol{x}^*}. \tag{12}$$

where $\boldsymbol{X}^{\mathrm{o}}$ ("o" standing for ordinal) is constructed by Kruskal's secondary least squares monotonic transformation [11] of $\boldsymbol{x}^*$. Kruskal's up-and-down block algorithm, also known as the pool-adjacent-violators algorithm, may be implemented in order to obtain $\tilde{\boldsymbol{X}}^{\mathrm{o}}$ [12]. The vector of the regression coefficient $(\tilde{\boldsymbol{X}}^{\mathrm{o}'}\tilde{\boldsymbol{X}}^{\mathrm{o}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{o}'}\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ contains the optimal scaling values which preserve the order of the categories of $\boldsymbol{x}^*$, as required by the conditions

$$(x_i^* \sim x_{i'}^*) \Rightarrow (\hat{x}_i = \hat{x}_{i'}) \text{ and } (x_i^* \prec x_{i'}^*) \Rightarrow (\hat{x}_i \leq \hat{x}_{i'}). \tag{13}$$

where the symbol $\prec$ indicates empirical order.

Nominal and ordinal quantification functions provide easily and clearly interpretable scalings, due to the fact that in a PLS framework the weight of a variable is a function of its correlation with the corresponding LC. When function $\mathcal{Q}(\tilde{\boldsymbol{X}}^{\mathrm{n}}, \boldsymbol{\gamma}_{\boldsymbol{x}^*})$ is used, the relationship between $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ and $\boldsymbol{x}^*$ in terms of linear correlation can be expressed as the Pearson's correlation ratio $\eta_{\boldsymbol{\gamma}_{\boldsymbol{x}^*}|\boldsymbol{x}^*}$:

$$\mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}}) = \eta_{\boldsymbol{\gamma}_{\boldsymbol{x}^*}|\boldsymbol{x}^*}. \tag{14}$$

Since $0 \leq \eta \leq 1$, this correlation will never be negative. This implies that the relation between a variable generated by a nominal scaling function and the LC can be interpreted in terms of intensity, but not in terms of sign. This makes sense, as it is a conceptual error to expect a sign in the covariation between a numerical variable (the LC) and a nominal variable, even after providing the nominal variable with a metric: by definition, this type of variable neither increases nor decreases.

When using Kruskal's secondary transformation, instead, the weight of the scaled variable indicates the extent to which relationship between $\boldsymbol{x}^*$ and the LC approaches monotonicity, as it is closely related to Kruskal's badness-of-fit STRESS index [11] by the following equations:

$$\mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}}) = \begin{cases} \sqrt{1 - STRESS^2_{(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \boldsymbol{x}^*)}} & \text{if } \mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}}) \geq 0 \\ -\sqrt{1 - STRESS^2_{(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \boldsymbol{x}^*)}} & \text{if } \mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}}) < 0 \end{cases}$$

The STRESS index is bounded between zero and one. It expresses the departure of the relationship between $\boldsymbol{x}^*$ and $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ from the assumption of monotonicity. The sign of $\mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}})$ depends on the type of monotonic transformation is applied to $\boldsymbol{x}^*$, which can be increasing or decreasing. Therefore, if $\mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}}) = 1$ a perfect increasing monotonic relationship exists between $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ and $\boldsymbol{x}^*$, while if $\mathrm{cor}(\boldsymbol{\gamma}_{\boldsymbol{x}^*}, \hat{\boldsymbol{x}}) = -1$ this relationships is perfectly decreasing and monotonic.

Advance knowledge of the degree D of a polynomial relationship between a raw numerical variable and the LC makes it possible to use the polynomial scaling. According to Young [41], optimal parameters for the polynomial transformation are found by projecting $\boldsymbol{\gamma}_{\boldsymbol{x}^*}$ on the space spanned by the columns of matrix $\tilde{\boldsymbol{X}}^{\mathrm{p}}$ ("p" standing for polynomial). Matrix $\tilde{\boldsymbol{X}}^{\mathrm{p}}$ is built with one row for each observation and with $D+1$ columns, each column being an integer power of the vector $\boldsymbol{x}^*$:

$$\mathcal{Q}(\tilde{\boldsymbol{X}}^{\mathrm{p}}, \boldsymbol{\gamma}_{\boldsymbol{x}^*}) = \tilde{\boldsymbol{X}}^{\mathrm{p}}(\tilde{\boldsymbol{X}}^{\mathrm{p}'}\tilde{\boldsymbol{X}}^{\mathrm{p}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{p}'}\boldsymbol{\gamma}_{\boldsymbol{x}^*}. \tag{15}$$

Assuming that the raw variable and the LC are linked by a linear relationship, all that is required is to set $D = 1$. If this is the case for all of the variables, NM-PLS methods provide the same results as standard PLS methods applied to standardized data.

It is noteworthy that scaling functions (10), (12) and (15) cannot be directly applied to raw variables, as the LC is unknown by definition. NM-PLS algorithms use PLS iteration in order to overcome this issue: model and scaling parameters are alternately estimated in a modified PLS-type loop with an added quantification step. In standard PLS steps the model parameters are estimated for given scaling parameters, while in the quantification step scaling parameters are estimated for given model parameters and raw variables are suitably transformed into interval variables normalized to unitary variance.

The Non-Metric approach to NIPALS, PLS-R and (*new*) *Mode A* PLS-PM algorithms are described in detail in the following sections.

5. A PLS algorithm for non-metric PCA: Non-Metric NIPALS

Non-Metric NIPALS (NM-NIPALS) performs a non-metric PCA on a matrix $\boldsymbol{X}^*$ representing a set of P raw variables observed on N units. In particular, it yields a matrix $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1 \ldots \hat{\boldsymbol{x}}_P]$ of standardized scaled variables $\hat{\boldsymbol{x}}_p \propto \tilde{\boldsymbol{X}}_p\boldsymbol{\phi}_p$ satisfying criterion (2), to which scaling parameters are added. From a mathematical point of view, the NM-NIPALS criterion can be expressed as

$$\arg\max_{\forall \boldsymbol{\phi}_p, \boldsymbol{w}_{(1)}} \sum_p \mathrm{cor}^2(\tilde{\boldsymbol{X}}_p\boldsymbol{\phi}_p, \hat{\boldsymbol{X}}\boldsymbol{w}_{(1)}) \tag{16}$$

under the constraints $\|\boldsymbol{w}_{(1)}\| = 1$ and $\mathrm{var}(\hat{\boldsymbol{x}}_p) = 1$. A further restriction is given by matrices $\tilde{\boldsymbol{X}}_p$, defined in advance for each variable according to the chosen scaling level.

For fixed $\boldsymbol{t}_{(1)} = \hat{\boldsymbol{X}}\boldsymbol{w}_{(1)}$, criterion (16) is separable with respect to the optimally scaled data for each variable and can be considered as a sum of criteria each of which is a function of the vector $\boldsymbol{\phi}_p$ only. Therefore, the scaling parameters can be identified by separately maximizing $\forall p \in \{1, \ldots, P\}$

$$\arg\max_{\boldsymbol{\phi}_p} \mathrm{cor}^2(\tilde{\boldsymbol{X}}_p\boldsymbol{\phi}_p, \boldsymbol{t}_{(1)}). \tag{17}$$

This criterion is optimized by the scaling functions $\mathcal{Q}(\tilde{\boldsymbol{X}}_p, \boldsymbol{t}_{(1)})$ discussed in section 4.

For fixed scaling parameters, on the other hand, criterion (16) is optimized by iterating the standard NIPALS loop until convergence. However, in criterion (17) the scaling parameters are functions of $\boldsymbol{t}_{(1)}$, while $\boldsymbol{t}_{(1)}$ in its turn is a function of the quantified variables $\hat{\boldsymbol{x}}_p$. The NM-NIPALS algorithm overcomes this issue by integrating the standard NIPALS loop with a quantification step in which scaling functions are applied. Scaling parameters can thus be estimated simultaneously with the PCA parameters by alternating maximization of criterion (16) with respect to scaling and model parameters.

The NM-NIPALS loop starts with a quantification step in which the functions $\mathcal{Q}(\tilde{\boldsymbol{X}}_p, \boldsymbol{t}_{(1)})$ are applied to each raw variable in order to obtain a first estimate of each $\hat{\boldsymbol{x}}_p$ (Step 1.1.1 of Algorithm 4). Then, for fixed $\hat{\boldsymbol{x}}_p$, optimal model parameters are determined each one a function of the previous one, as in the standard NIPALS loop (Steps 1.1.2-1.1.4). This procedure is iterated until convergence. Since in each iteration the quantification step and the the standard NIPALS steps yield scaling and model parameter proxies which increase criterion (16), the procedure monotonically converges to a maximum for criterion (16). After obtaining convergence, the matrix $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1 \dots \hat{\boldsymbol{x}}_p]$ is deflated (Step 1.2), and the second principal component is added by implementing the standard NIPALS algorithm to the residual matrix $\boldsymbol{E}_{(1)}$. Higher order components are extracted in a similar way.

Algorithm 4 NM-NIPALS algorithm

Input: $\boldsymbol{X}^*$
Output: $\boldsymbol{W} = [\boldsymbol{w}_{(1)} \dots \boldsymbol{w}_{(H)}], \boldsymbol{T} = [\boldsymbol{t}_{(1)} \dots \boldsymbol{t}_{(H)}], \hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1 \dots \hat{\boldsymbol{x}}_p]$
Step 1.0: **Initialize** $\boldsymbol{t}_{(1)} = \boldsymbol{t}_{(1)}^{(0)}$
Step 1.1:
repeat
  for all $p = 1, \dots, P$ **do**
    Step 1.1.1: $\hat{\boldsymbol{x}}_p^{(s)} \propto \mathcal{Q}(\tilde{\boldsymbol{X}}_p, \boldsymbol{t}_{(1)}^{(s)})$
  end for
  Step 1.1.2: $\boldsymbol{w}_{(1)}^{(s)} = \hat{\boldsymbol{X}}^{(s)'} \boldsymbol{t}_{(1)}^{(s)} / (\boldsymbol{t}_{(1)}^{(s)'} \boldsymbol{t}_{(1)}^{(s)})$
  Step 1.1.3: $\boldsymbol{w}_{(1)}^{(s)} = \boldsymbol{w}_{(1)}^{(s)} / \|\boldsymbol{w}_{(1)}^{(s)}\|$
  Step 1.1.4: $\boldsymbol{t}_{(1)}^{(s+1)} = \hat{\boldsymbol{X}}^{(s)} \boldsymbol{w}_{(1)}^{(s)}$
until convergence of $\boldsymbol{w}_{(1)}$
Step 1.2: $\boldsymbol{E}_1 = \hat{\boldsymbol{X}} - \boldsymbol{t}_{(1)} \boldsymbol{w}_{(1)}'$
for all $h = 2, \dots, H$ **do**
  Step 2.0: Initialize $\boldsymbol{t}_{(h)}^{(0)}$
  Step 2.1:
  repeat
    Step 2.1.1: $\boldsymbol{w}_{(h)}^{(s)} = \boldsymbol{E}_{(h-1)}' \boldsymbol{t}_{(h)}^{(s)} / (\boldsymbol{t}_{(h)}^{(s)'} \boldsymbol{t}_{(h)}^{(s)})$
    Step 2.1.2: $\boldsymbol{w}_{(h)}^{(s)} = \boldsymbol{w}_{(h)}^{(s)} / \|\boldsymbol{w}_{(h)}^{(s)}\|$
    Step 2.1.3: $\boldsymbol{t}_{(h)}^{(s+1)} = \boldsymbol{E}_{h-1} \boldsymbol{w}_{(h)}^{(s)}$
  until convergence of $\boldsymbol{w}_{(h)}$
  Step 2.2: $\boldsymbol{E}_{(h)} = \boldsymbol{E}_{(h-1)} - \boldsymbol{t}_{(h)} \boldsymbol{w}_{(h)}'$
end for

5.1. Links with other non-metric approaches to PCA

If all the variables are analyzed at a nominal scaling level, the NM-NIPALS solution for $\boldsymbol{t}_{(1)}$ corresponds, up to a multiplying factor, to the dominant (non-trivial) eigenvector of $(1/N)\tilde{\tilde{\boldsymbol{X}}}\boldsymbol{D}^{-1}\tilde{\tilde{\boldsymbol{X}}}'$, where $\boldsymbol{D} = \text{diag}(\tilde{\tilde{\boldsymbol{X}}}'\tilde{\tilde{\boldsymbol{X}}})$ is the diagonal matrix of the frequencies of the modalities and $\tilde{\tilde{\boldsymbol{X}}} = [\tilde{\boldsymbol{X}}^{\text{n}}_1|\ldots|\tilde{\boldsymbol{X}}^{\text{n}}_P]$. In fact, if all of the variables are nominal, step 1.1.1 of Algorithm 4, for each p, becomes

$$\hat{\boldsymbol{x}}_p \propto \tilde{\boldsymbol{X}}^{\text{n}}_p(\tilde{\boldsymbol{X}}^{\text{n}'}_p\tilde{\boldsymbol{X}}^{\text{n}}_p)^{-1}\tilde{\boldsymbol{X}}^{\text{n}'}_p\boldsymbol{t}_{(1)}.$$

For these scaling values

$$\text{cor}^2(\boldsymbol{t}_{(1)}, \hat{\boldsymbol{x}}_p) \propto \boldsymbol{t}'_{(1)}\tilde{\boldsymbol{X}}^{\text{n}}_p(\tilde{\boldsymbol{X}}^{\text{n}'}_p\tilde{\boldsymbol{X}}^{\text{n}}_p)^{-1}\tilde{\boldsymbol{X}}^{\text{n}'}_p\boldsymbol{t}_{(1)}$$

and criterion (16) can be written

$$\arg\max_{\forall\boldsymbol{\phi}_p}\sum_p \boldsymbol{t}'_{(1)}\tilde{\boldsymbol{X}}^{\text{n}}_p(\tilde{\boldsymbol{X}}^{\text{n}'}_p\tilde{\boldsymbol{X}}^{\text{n}}_p)^{-1}\tilde{\boldsymbol{X}}^{\text{n}'}_p\boldsymbol{t}_{(1)}$$

which is equal to

$$\boldsymbol{t}'_{(1)}\tilde{\tilde{\boldsymbol{X}}}\boldsymbol{D}^{-1}\tilde{\tilde{\boldsymbol{X}}}'\boldsymbol{t}_{(1)}.$$

Therefore, for less than a constant factor, the NM-NIPALS first component equals the one dimensional solution for the unit scaling in well-known methods such as Multiple Correspondence Analysis [1] and Homogeneity Analysis [5].

When the analysis is generalized at a variety of scaling levels, for less than a constant factor due to different normalization constraints, the NM-NIPALS solution coincides with the one-dimensional solution of others algorithms implementing non-metric PCA, as it optimizes the PCA criterion subject to the same scaling restrictions. In particular:

- In nominal scaling, we observe the restrictions used by the discrete-nominal option in PRINCALS [5, 18] and PRINCIPALS [42] and in PRINQUALS [27] algorithms.
- In ordinal scaling, we observe the restrictions used by the discrete-ordinal option in PRINCALS and PRINCIPALS, and by secondary approach to monotonicity in Kruskal and Shepard's non-metric PCA [13].

Finally, when all raw variables are observed on a metric scale and a polynomial transformation of degree $D = 1$ is chosen for all of them, we observe the restrictions involved in standard PCA (variables are centered and scaled to unitary variance).

6. Non-Metric PLS Regression

Let two blocks of raw variables $\boldsymbol{X}^*_1$ and $\boldsymbol{X}^*_2$ be measured on a variety of measurement scales. The generic variable of $\boldsymbol{X}^*_2$ is $\boldsymbol{x}^*_{p2}$ (with $p = 1\ldots P_2$), while the

generic variable of $\boldsymbol{X}_1^*$ is $\boldsymbol{x}_{p1}^*$ (with $p = 1 \dots P_1$). NM-PLSR searches for optimally scaled data matrices $\hat{\boldsymbol{X}}_1$, with the generic column $\hat{\boldsymbol{x}}_{p1} \propto \tilde{\boldsymbol{X}}_{p1}\boldsymbol{\phi}_{p1}$, and $\hat{\boldsymbol{X}}_2$, with the generic column $\hat{\boldsymbol{x}}_{p2} \propto \tilde{\boldsymbol{X}}_{p2}\boldsymbol{\phi}_{p2}$. For the p-th variable, matrix $\tilde{\boldsymbol{X}}_{pq}$ ($q \in \{1, 2\}$) defines the constraints due to the scaling level and $\boldsymbol{\phi}_{pq}$ represents the vectors of the scaling values.

NM-PLSR scalings are optimal in the sense that they optimize the single component PLS-R criterion

$$\underset{\boldsymbol{w}_{p1(1)}, \boldsymbol{w}_{p2(1)}, \forall \boldsymbol{\phi}_{p1}, \forall \boldsymbol{\phi}_{p2}}{\arg\max} \operatorname{cov}^2(\hat{\boldsymbol{X}}_2\boldsymbol{w}_{1(1)}, \hat{\boldsymbol{X}}_1\boldsymbol{w}_{2(1)}) \tag{18}$$

under the constraints $\|\boldsymbol{w}_{1(1)}\| = \|\boldsymbol{w}_{2(1)}\| = \operatorname{var}(\hat{\boldsymbol{x}}_{p1}) = \operatorname{var}(\hat{\boldsymbol{x}}_{p2}) = 1$. It is noteworthy that, since this criterion involves only the first component, it can also be applied to the other PLS approaches to two-block cross-covariance analysis cited in section 2.2. Criterion (18) depends on two sets of parameters. The first set consists of the model parameters, constrained to unitary norm. The other set consists of scaling parameters, which must respect restrictions due to the scaling level chosen for each variable and to the standardization constraint applicable to the scaled variables.

For fixed scaling parameters, the NM-PLSR optimization problem becomes

$$\underset{\|\boldsymbol{w}_{1(1)}\|=1, \|\boldsymbol{w}_{2(1)}\|=1}{\arg\max} \operatorname{cov}^2(\hat{\boldsymbol{X}}_2\boldsymbol{w}_{1(1)}, \hat{\boldsymbol{X}}_1\boldsymbol{w}_{2(1)}) \tag{19}$$

Optimal $\boldsymbol{w}_{2(1)}$ and $\boldsymbol{w}_{1(1)}$ are respectively dominant right and left singular vectors of matrix $(1/N)\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{X}}_1$ (see section 2.2). Therefore, in order to optimize criterion (19), conditions

$$\boldsymbol{w}_{2(1)} = \hat{\boldsymbol{X}}_2'\hat{\boldsymbol{X}}_1\boldsymbol{w}_{1(1)} / \|\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{X}}_1\boldsymbol{w}_{1(1)}\| \tag{20}$$

and

$$\boldsymbol{w}_{1(1)} = \hat{\boldsymbol{X}}_1'\hat{\boldsymbol{X}}_2\boldsymbol{w}_{2(1)} / \|\hat{\boldsymbol{X}}_1'\hat{\boldsymbol{X}}_2\boldsymbol{w}_{2(1)}\| \tag{21}$$

must be respected.

When the other parameters remain fixed, the response scaling parameters must optimize the following criterion:

$$\begin{aligned} \underset{\forall \boldsymbol{\phi}_{p1}}{\arg\max} \quad & \operatorname{cov}^2(\hat{\boldsymbol{X}}_2\boldsymbol{w}_{1(1)}, \hat{\boldsymbol{X}}_1\boldsymbol{w}_{2(1)}) \\ &= \frac{1}{N^2}\boldsymbol{w}_{2(1)}'\hat{\boldsymbol{X}}_2'\hat{\boldsymbol{X}}_1\hat{\boldsymbol{X}}_1'\hat{\boldsymbol{X}}_2\boldsymbol{w}_{2(1)} \\ &= \frac{1}{N^2}\boldsymbol{t}_{2(1)}'\hat{\boldsymbol{X}}_1\hat{\boldsymbol{X}}_1'\boldsymbol{t}_{2(1)} \\ &= \sum_p^{P_1} \operatorname{cov}^2(\hat{\boldsymbol{x}}_{p1}, \boldsymbol{t}_{2(1)}) \\ &= \sum_p^{P_1} \operatorname{cor}^2(\hat{\boldsymbol{x}}_{p1}, \boldsymbol{t}_{2(1)}) \operatorname{var}(\hat{\boldsymbol{x}}_{p1}) \operatorname{var}(\boldsymbol{t}_{2(1)}) \end{aligned} \tag{22}$$

Since $\text{var}(\hat{\boldsymbol{x}}_{p1}) = 1$ and $\text{var}(\boldsymbol{t}_{2(1)})$ is fixed with respect to the sum, criterion (22) can be rewritten as

$$\arg\max_{\forall \boldsymbol{\phi}_{p1}} \sum_{p}^{P_1} \text{cor}^2(\tilde{\boldsymbol{X}}_{p1}\boldsymbol{\phi}_{p1}, \boldsymbol{t}_{2(1)}) \tag{23}$$

This function is separable with respect to each $\boldsymbol{\phi}_{p1}$, and it can be considered as a sum of P_1 components, each of which is a function of the scaling parameters of one variable only:

$$\forall p \in \{1 \dots P_1\} \quad \arg\max_{\boldsymbol{\phi}_{p1}} \text{cor}^2(\tilde{\boldsymbol{X}}_{p1}\boldsymbol{\phi}_{p1}, \boldsymbol{t}_{2(1)}). \tag{24}$$

Since optimizing each criterion (24) is equivalent to optimize criterion (22), scaling parameters can be independently calculated as OLS coefficients of the regressions of $\boldsymbol{t}_{2(1)}$ on each matrix $\tilde{\boldsymbol{X}}_{p1}$. Hence, each optimally scaled predictor $\hat{\boldsymbol{x}}_{p1}$ can be calculated as

$$\hat{\boldsymbol{x}}_{p1} \propto \mathcal{Q}(\tilde{\boldsymbol{X}}_{p1}, \boldsymbol{t}_{2(1)}). \tag{25}$$

Specular reasoning can be used to find the optimal quantifications for response variables, leading, for each $p \in \{1 \dots P_2\}$, to the criterion

$$\arg\max_{\boldsymbol{\phi}_{p2}} \text{cor}^2(\tilde{\boldsymbol{X}}_{p2}\boldsymbol{\phi}_{p2}, \boldsymbol{t}_{1(1)}), \tag{26}$$

which is satisfied by

$$\hat{\boldsymbol{x}}_{p2} \propto \mathcal{Q}(\tilde{\boldsymbol{X}}_{p2}, \boldsymbol{t}_{1(1)}). \tag{27}$$

This specular treatment is due to the fact that the single component PLS-R model is symmetric. The PLS-R model becomes asymmetric only when successive latent dimensions are computed by deflating both predictors and responses with respect to the components in the predictor space.

In the NM-PLSR algorithm conditions (20), (21), (25) and (27) are integrated into a modified PLS-R loop which starts by quantifying each predictor through the quantification functions $\mathcal{Q}(\tilde{\boldsymbol{X}}_{p1}, \boldsymbol{t}_{2(1)})$ (Step 1.1.1 of Algorithm 5), so as to get a first approximation of the matrix of the quantified predictors $\hat{\boldsymbol{X}}_1$. Once this is done, $\boldsymbol{w}_{1(1)}$ is calculated as a function of $\hat{\boldsymbol{X}}_1$ and $\boldsymbol{t}_{2(1)}$ (Step 1.1.2). Then, $\boldsymbol{t}_{1(1)}$ is computed as a function of $\hat{\boldsymbol{X}}_1$ and $\boldsymbol{w}_{1(1)}$ (Step 1.1.3), and $\boldsymbol{X}_2$-variables are quantified by means of the quantification functions $\mathcal{Q}(\tilde{\boldsymbol{X}}_{p2}, \boldsymbol{t}_{1(1)})$ (Step 1.1.4). The weights $\boldsymbol{w}_{2(1)}$ are then computed as a function of $\hat{\boldsymbol{X}}_2$ and $\boldsymbol{t}_{1(1)}$ (Step 1.1.5). Finally, the vector $\boldsymbol{t}_{2(1)}$ is computed as a function of $\hat{\boldsymbol{X}}_2$ and $\boldsymbol{w}_{2(1)}$ (Step 1.1.6). This loop is repeated until convergence is reached. In each step of the loop, criterion (18) is maximized with respect to one of the parameters, keeping the others constant.

Since in each loop of the iterative process the criterion increases, convergence is reached when a maximum is encountered. The MN-PLSR algorithm continues

by calculating residuals matrices $\hat{\boldsymbol{E}}_1$ and $\hat{\boldsymbol{F}}_1$ as functions of $\hat{\boldsymbol{X}}_1$ and $\hat{\boldsymbol{X}}_2$ respectively (Steps 1.2-1.5). Residuals matrices are then entered into the standard PLS-R loop in order to extract a second set of components, and so on.

This procedure is based on the PLS-R algorithm described in [8] and shown in Algorithm 5; however, the version presented in [28] can be modified in a similar way in order to deal with non-metric variables.

Algorithm 5 NM-PLSR algorithm

Input: $\boldsymbol{X}_1^*$, $\boldsymbol{X}_2^*$
Output: $\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{T}_1, \boldsymbol{T}_2, \boldsymbol{P}, \hat{\boldsymbol{X}}_1, \hat{\boldsymbol{X}}_2$
Step 1.0: **Initialize** $\boldsymbol{t}_{2(1)} = \boldsymbol{t}_{2(1)}^{(0)}$
Step 1.1:
repeat
  for all $p = 1, \ldots, P_1$ **do**
    Step 1.1.1: $\hat{\boldsymbol{x}}_{p1}^{(s)} \propto \mathcal{Q}(\tilde{\boldsymbol{X}}_{p1}, \boldsymbol{t}_{2(1)}^{(s)})$
  end for
  Step 1.1.2: $\boldsymbol{w}_{1(1)}^{(s)} = \hat{\boldsymbol{X}}_1^{(s)'} \boldsymbol{t}_{2(1)}^{(s)} / \| \hat{\boldsymbol{X}}_1^{(s)'} \boldsymbol{t}_{2(1)}^{(s)} \|$
  Step 1.1.3: $\boldsymbol{t}_{1(1)}^{(s)} = \hat{\boldsymbol{X}}_1^{(s)} \boldsymbol{w}_{1(1)}^{(s)}$
  for all $p = 1, \ldots, P_2$ **do**
    Step 1.1.4: $\hat{\boldsymbol{x}}_{p2}^{(s)} \propto \mathcal{Q}(\tilde{\boldsymbol{X}}_{p2}, \boldsymbol{t}_{1(1)}^{(s)})$
  end for
  Step 1.1.5: $\boldsymbol{w}_{2(h)}^{(s)} = \hat{\boldsymbol{X}}_2^{(s)'} \boldsymbol{t}_{1(1)}^{(s)} / \| \hat{\boldsymbol{X}}_2^{(s)'} \boldsymbol{t}_{1(1)}^{(s)} \|$
  Step 1.1.6: $\boldsymbol{t}_{2(1)}^{(s+1)} = \hat{\boldsymbol{X}}_2^{(s)} \boldsymbol{w}_{2(1)}^{(s)}$
until convergence of $\boldsymbol{w}_{1(1)}$
Step 1.2: $\boldsymbol{p}_{(1)} = \hat{\boldsymbol{X}}_1' \boldsymbol{t}_{1(1)} / (\boldsymbol{t}_{1(1)}' \boldsymbol{t}_{1(1)})$
Step 1.3: $b_{(\boldsymbol{t}_{2(1)} | \boldsymbol{t}_{1(1)})} = \boldsymbol{t}_{2(1)}' \boldsymbol{t}_{1(1)} / (\boldsymbol{t}_{1(1)}' \boldsymbol{t}_{1(1)})$
Step 1.4: $\boldsymbol{E}_{(1)} = \hat{\boldsymbol{X}}_1 - \boldsymbol{t}_{1(1)} \boldsymbol{p}_{(1)}'$
Step 1.5: $\boldsymbol{F}_{(1)} = \hat{\boldsymbol{X}}_2 - b_{(\boldsymbol{t}_{2(1)} | \boldsymbol{t}_{1(1)})} \boldsymbol{t}_{1(1)} \boldsymbol{w}_{2(1)}'$
for all $h = 2, \ldots, H$ **do**
  Step 2.0: Initialize $\boldsymbol{t}_{2(h)} = \boldsymbol{t}_{2(h)}^{(0)}$
  Step 2.1:
  repeat
    Step 2.1.1: $\boldsymbol{w}_{1(h)}^{(s)} = \boldsymbol{E}_{(h-1)}' \boldsymbol{t}_{2(h)}^{(s)} / \| \boldsymbol{E}_{(h-1}' \boldsymbol{t}_{2(h)}^{(s)} \|$
    Step 2.1.2: $\boldsymbol{t}_{1(h)}^{(s)} = \boldsymbol{E}_{(h-1)} \boldsymbol{w}_{1(h)}^{(s)}$
    Step 2.1.3: $\boldsymbol{w}_{2(h)}^{(s)} = \boldsymbol{F}_{(h-1)}' \boldsymbol{t}_{1(h)}^{(s)} / \| \boldsymbol{F}_{(h-1)}' \boldsymbol{t}_{1(h)}^{(s)} \|$
    Step 2.1.4: $\boldsymbol{t}_{2(h)}^{(s+1)} = \boldsymbol{F}_{(h-1)} \boldsymbol{w}_{2(h)}^{(s)}$
  until convergence of $\boldsymbol{w}_{1(h)}$
  Step 2.2: $\boldsymbol{p}_{(h)} = \boldsymbol{E}_{(h-1)}' \boldsymbol{t}_{1(h)} / (\boldsymbol{t}_{1(h)}' \boldsymbol{t}_{1(h)})$
  Step 2.3: $b_{(\boldsymbol{t}_{2(h)} | \boldsymbol{t}_{1(h)})} = \boldsymbol{t}_{2(h)}' \boldsymbol{t}_{1(h)} / (\boldsymbol{t}_{1(h)}' \boldsymbol{t}_{1(h)})$
  Step 2.4: $\boldsymbol{E}_{(h)} = \boldsymbol{E}_{(h-1)} - \boldsymbol{t}_{1(h)} \boldsymbol{p}_{(h)}'$
  Step 2.5: $\boldsymbol{F}_{(h)} = \boldsymbol{F}_{(h-1)} - b_{(\boldsymbol{t}_{2(h)} | \boldsymbol{t}_{1(h)})} \boldsymbol{t}_{1(h)} \boldsymbol{w}_{2(h)}'$
end for

7. Non-Metric PLS Path Modeling

The basic principles of the NM-PLSR algorithm can be extended to the PLS approach to predictive Path Modeling, as the single component PLS-R algorithm is the same as the two-block PLS-PM algorithm, except for normalization constraints. In the PLS-PM vocabulary, $\boldsymbol{t}_{2(1)}$ can be interpreted as the outer estimate of the LV associated with block $\boldsymbol{X}_2$ ($\boldsymbol{t}_{2(1)} = \boldsymbol{X}_2\boldsymbol{w}_{2(1)}$), as well as the inner estimate of the LV underlying $\boldsymbol{X}_1$ by means of which the outer weights ($\boldsymbol{w}_{1(1)} = \boldsymbol{X}'\boldsymbol{t}_{2(1)}/\|\boldsymbol{X}'\boldsymbol{t}_{2(1)}\|$) are calculated. Symmetrically, $\boldsymbol{t}_{1(1)}$ can be considered simultaneously as the outer estimate of the LV underlying $\boldsymbol{X}_1$ and the inner estimate of the LV underlying $\boldsymbol{X}_2$. These double functions are justified by the inner linear relationship $\boldsymbol{t}_{2(1)} = \alpha\boldsymbol{t}_{1(1)}$. This is a hidden step in the PLS-R algorithm, which, in PLS-PM terms, can be interpreted as: the outer estimate of the LV in one block is used the as inner estimate of the LV in the other block.

In view of these considerations, it can been stated that in the NM-PLSR model, interpreted as a two-block path model, any quantified variable is computed as a function of the inner estimate of the corresponding LV. Therefore, the non-metric extension of the PLS-R algorithm can also be applied to PLS-PM algorithm by adding to the PLS-PM loop a quantification step in which any scaled MV is computed as a function of the inner estimate of the corresponding LV.

The NM-PLSPM algorithm is described in Algorithm 6. It starts by initializing LV outer estimates by means of centered vectors $\boldsymbol{t}_q^{(0)}$. A first approximation of the inner weights (Step 1.1) and of the LV inner estimates $\boldsymbol{z}_q^{(0)}$ (Step 1.2) are then computed as functions of $\boldsymbol{t}_q^{(0)}$. The LV inner estimates are in turn used for computing a first quantification $\hat{\boldsymbol{x}}_{pq}^{(0)}$ of the MVs by means of the functions $\mathcal{Q}(\tilde{\boldsymbol{X}}_{pq}, \boldsymbol{z}_q^{(0)})$ (Step 1.3). Parameter $\tilde{\boldsymbol{X}}_{pq}$ of the scaling functions depends on the scaling level defined in advance for each MV. Afterwards, a first estimate of the outer weights is computed as a function of $\boldsymbol{z}_q^{(0)}$ and $\hat{\boldsymbol{X}}_q^{(0)}$ (Step 1.4). Finally, a new proxy of LV outer estimates $\boldsymbol{t}_q^{(1)}$ is computed by *Mode A* estimation mode (Step 1.5). This loop is repeated upon convergence. It yields final outer weights, which are used to calculate the final LV scores (Step 2). Path coefficients are then calculated as in the standard PLS-PM algorithm.

7.1. The optimizing criterion of Non-Metric PLS Path Modeling

Unfortunately, *Mode A* NM-PLSPM suffers from the same drawbacks as *Mode A* PLS-PM. That is, since the criterion to which the *Mode A* PLS-PM algorithm converges is unknown, it is not possible to state that scalings provided by the *Mode A* NM-PLSPM algorithm are mathematically optimal with respect to the model. However, this problem disappears if the *new Mode A* scheme is used, i.e. by implementing a RGCCA with the regularization parameter $\tau_q = 1$ for all the blocks [31].

As seen in section 2.3, *new Mode A* PLS-PM has recently been shown to optimize criterion (7) when a centroid or factorial scheme is used. In the non-

Algorithm 6 NM-PLS Path Modeling algorithm: estimation of the LVs.

Input: $\boldsymbol{X}_1^* \dots \boldsymbol{X}_Q^*$, $\boldsymbol{C}$
Output: $\hat{\boldsymbol{X}}_1 \dots \hat{\boldsymbol{X}}_Q$, $\boldsymbol{w}_1 \dots \boldsymbol{w}_Q$, $\hat{\boldsymbol{\xi}}_1 \dots \hat{\boldsymbol{\xi}}_Q$
Step 0: Initialize $\boldsymbol{t}_q = \boldsymbol{t}_q^{(0)}$ **for all** $q = 1, \dots, Q$
Step 1: Iteration
repeat
  for all $q = 1, \dots, Q$ **do**
    Step 1.1: Computation of the inner weights
    if $q < q'$ **then**
      $e_{qq'}^{(s)} = f\left(\boldsymbol{t}_q^{(s)}, \boldsymbol{t}_{q'}^{(s)}\right)$
    else
      $e_{qq'}^{(s)} = f\left(\boldsymbol{t}_q^{(s)}, \boldsymbol{t}_{q'}^{(s+1)}\right)$
    end if
    Step 1.2: Inner estimation of the LVs
    $\boldsymbol{z}_q^{(s)} = \sum_{q'<q}\left(c_{qq'}e_{qq'}^{(s)}\boldsymbol{t}_{q'}^{(s+1)}\right) + \sum_{q'>q}\left(c_{qq'}e_{qq'}^{(s)}\boldsymbol{t}_{q'}^{(s)}\right)$
    Step 1.3: Quantification step
    for all $p = 1, \dots, P_q$ **do**
      $\hat{\boldsymbol{x}}_{pq}^{(s)} \propto \mathcal{Q}(\tilde{\boldsymbol{X}}_{pq}, \boldsymbol{z}_q^{(s)})$
    end for
    Step 1.4: Computation of the outer weights
    $$\begin{cases} \boldsymbol{w}_q^{(s)} = (1/N)\hat{\boldsymbol{X}}_q^{(s)'}\boldsymbol{z}_q^{(s)} & \text{if Mode A} \\ \boldsymbol{w}_q^{(s)} = \hat{\boldsymbol{X}}_q^{(s)'}\boldsymbol{z}_q^{(s)}/\|\hat{\boldsymbol{X}}_q^{(s)'}\boldsymbol{z}_q^{(s)}\| & \text{if new Mode A} \end{cases}$$
    Step 1.5: Outer estimation of the LVs
    $$\begin{cases} \boldsymbol{t}_q^{(s+1)} = \hat{\boldsymbol{X}}_q^{(s)}\boldsymbol{w}_q^{(s)} & \text{if new Mode A} \\ \boldsymbol{t}_q^{(s+1)} = (N^{1/2}\hat{\boldsymbol{X}}_q^{(s)}\boldsymbol{w}_q^{(s)})/\|\hat{\boldsymbol{X}}_q^{(s)}\boldsymbol{w}_q^{(s)}\| & \text{otherwise} \end{cases}$$
  end for
until convergence of $\boldsymbol{w}_q$
Step 2: Computation of the LVs
for all $q = 1, \dots, Q$ **do**
  $\hat{\boldsymbol{\xi}}_q = (N^{1/2}\hat{\boldsymbol{X}}_q\boldsymbol{w}_q)/\|\hat{\boldsymbol{X}}_q\boldsymbol{w}_q\|$
end for

metric version of the *new Mode A* PLS-PM, scaling parameters $\boldsymbol{\phi}_{pq}$ are added in criterion (7) in order to obtain quantified variables $\hat{\boldsymbol{x}}_{pq} \propto \tilde{\boldsymbol{X}}_{pq}\boldsymbol{\phi}_{pq}$. Particular consideration is given to the following criterion:

$$\begin{aligned} \underset{\forall \boldsymbol{w}_q, \forall \boldsymbol{\phi}_{pq}}{\arg\max} \quad & \sum_q \operatorname{cov}(\hat{\boldsymbol{X}}_q\boldsymbol{w}_q, \boldsymbol{z}_q) \\ & = \sum_q \frac{1}{N}\boldsymbol{w}_q'\hat{\boldsymbol{X}}_q'\boldsymbol{z}_q \\ & = \frac{1}{N}\sum_q \frac{\boldsymbol{z}_q'\hat{\boldsymbol{X}}_q\hat{\boldsymbol{X}}_q'\boldsymbol{z}_q}{||\hat{\boldsymbol{X}}_q'\boldsymbol{z}_q||} \\ & = \frac{1}{N}\sum_q \frac{\sum_p (N \times \operatorname{cov}(\hat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q))^2}{\sqrt{\sum_p (N \times \operatorname{cov}(\hat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q))^2}} \end{aligned}$$

$$= \sum_q \frac{\sum_p \text{cov}^2(\hat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q)}{\sqrt{\sum_p \text{cov}^2(\hat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q)}}$$
$$= \sum_q \sqrt{\sum_p \text{cov}^2(\hat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q)}$$
$$= \sum_q \sqrt{\text{var}(\boldsymbol{z}_q)} \sqrt{\sum_p \text{cor}^2(\hat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q)} \tag{28}$$

under the constraints $||\boldsymbol{w}_q|| = 1$ and $\text{var}(\hat{\boldsymbol{x}}_{pq}) = 1$. For fixed $\boldsymbol{z}_q$, criterion (28) consists of a sum of $p \times q$ criteria, each of which is a function of a scaled variable. Hence, it can be optimized by separately maximizing

$$\forall p, q \qquad \arg\max_{\boldsymbol{\phi}_{pq}} \text{cor}^2(\tilde{\boldsymbol{X}}_{pq}\boldsymbol{\phi}_{pq}, \boldsymbol{z}_q) \tag{29}$$

by means of the scaling functions $\mathcal{Q}(\tilde{\boldsymbol{X}}_{pq}, \boldsymbol{z}_q)$. This implies that when *new Mode A* is used, the quantification step 1.3 in Algorithm 6 optimizes model criterion with respect to scaling parameters for fixed model parameters. Therefore, a procedure that alternates the PLS-PM iterative estimation (for fixed scaling parameters) and the quantification step (for fixed model parameters) is expected to converge to a maximum for criterion (7), as in each of the two steps criterion (7) cannot decrease. However, in order to avoid unnecessary computations, in the NM-PLSPM algorithm the quantification step is directly inserted into the PLS loop, exactly as in Algorithm 6.

8. Missing data

Standard OS algorithms apply different solutions for handling missing data. Gifi [5] distinguishes among three options. In the first option (*missing data passive*) the row of $\tilde{\boldsymbol{X}}$ is a zero row if the corresponding observation is missing for the variable $\boldsymbol{x}^*$; this option corresponds to assigning that category a zero scaling value. The second option is called *missing data multiple category*; it implies that a new category is added, to which all missing terms belong. The third option is the *missing data single category*; it means that for each missing term a new category is added.

All these solutions require advance imputation of missing data, and they can also be implemented in an NM-PLS framework. However, in the PLS framework a fourth option is possible: data pre-handling can be avoided, since PLS algorithms need no advance imputation of missing data. In PLS procedures, each score, as well as each weight, is calculated by means of an iterated sequence of dot products. In the case of missing data, dot products are obtained by pairwise deletion, that is by summing products calculated on available pairs. This characteristic of standard PLS algorithms is retained in NM-PLS quantification procedures, where missing values are left missing and available data are quantified.

9. An application to macroeconomic data

The data-set for this example consists of 9 macroeconomic variables observed for 47 countries. Three variables measure the inequality of land distribution:

- *gini*: Gini's index of concentration;
- *farm*: complement of the percentage of farmers that own half of the lands, starting with the smallest ones. Thus if *farm* is 90%, then 10% of the farmers own half of the lands;
- *rent*: percentage of farm households that rent all their land.

Two variables measure industrial development:

- *gnpr*: gross national product *pro capite* (in U.S. dollars) in 1955;
- *labo*: the percentage of labor force employed in agriculture.

Four variables measure political instability:

- *inst*: an index, bounded from 0 (stability) to 17 (instability), calculated as a function of the number of the chiefs of the executive and of the number of years of independence of the country during the period 1946-1961;
- *ecks*: the Eckstein's index, which measures the number of violent internal war incidents during the same period;
- *death*: number of people killed as a result of violent manifestations during the period 1950-1962;
- *demo*: a categorical variable that classifies countries in three groups: stable democracy, unstable democracy and dictatorship.

These data were collected and analyzed by Russett [22] in order to assess the hypothesis that the political instability of a country depends on the inequality of land distribution and the industrial development. Years later, the same data-set was analyzed by Gifi [5], Tenenhaus [28] and Tenenhaus and Tenenhaus [31]. Gifi used an optimal scaling technique to address the high degree of non-linearity of data, scaled in such a way as to maximize the canonical correlation between the variables reflecting the inequality of land distribution and the industrial development and the variables reflecting political instability. Indeed, Gifi himself noticed that partitioning data in three sets of variables (agricultural inequality, industrial development and political instability) would have been a more rational approach. Starting from this idea, Tenenhaus [28] analyzed the data-set in a multi-block framework; he addressed non-linearity by approximating Gifi's scalings by means of *a priori* monotone functional transformations and handled the variable *demo* by splitting it into three binary variables. However, Tenenhaus [28] pinpointed that approximating Gifi's transformations is not the best choice, while transformations optimized for PLS-PM would be preferable.

Here two approaches are compared. In the first one the standard PLS-PM algorithm is run on the original data[1], and the nominal variable *demo* is split into

[1] In the original data-set three values are missing for variable *rent*. According to Tenenhaus and Tenenhaus [31], they have been replaced for Australia, Nicaragua and Peru by the values 25.3, 9.9 and 12.6.

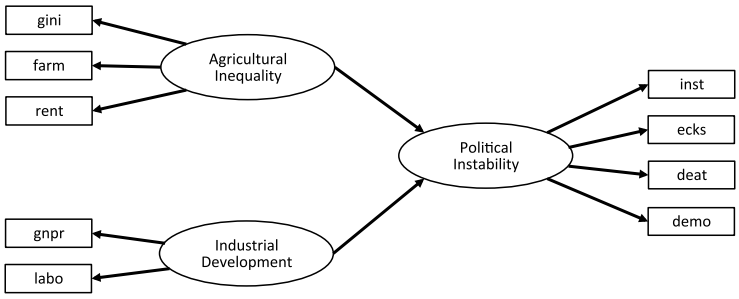


FIG 1. *The path model. Drawing conventions: outward directed arrows joining LVs (in ellipses) to MVs (in boxes) indicate that (New) Mode A estimation mode is used in the algorithm; the arrows linking LVs indicate the dependence flow which specifies the inner model.*

TABLE 2

Parameters of the inner model (path coefficients β) and of the outer model (loadings λ) in standard and Non-Metric PLS path models, with corresponding 95% confidence intervals (C.I.) built by means of 2000 bootstrap samples [3]

(a) Standard PLS

LV	β	C.I.
Agr. In.	0.23	[0.11; 0.43]
Ind. Dev.	-0.67	[−0.72; −0.58]
MV	λ	**C.I.**
gini	0.98	[0.89; 0.99]
farm	0.99	[0.91; 0.99]
rent	0.18	[−0.36; 0.63]
gnpr	0.95	[0.93; 0.97]
labo	-0.95	[−0.97; −0.93]
inst	0.17	[−0.10; 0.45]
ecks	0.67	[0.41; 0.84]
death	0.53	[0.40; 0.69]
stab	-0.87	[−0.94; −0.72]
unstab	0.10	[−0.25; 0.38]
dict	0.74	[0.53; 0.88]

(b) Non-Metric PLS

LV	β	C.I.
Agr. In.	0.30	[0.19; 0.50]
Ind. Dev.	-0.71	[−0.80; −0.54]
MV	λ	**C.I.**
gini	0.95	[0.89; 0.97]
farm	0.96	[0.86; 0.97]
rent	0.63	[−0.60; 0.82]
gnpr	0.96	[0.94; 0.98]
labo	-0.96	[−0.98; −0.95]
inst	0.63	[0.47; 0.80]
ecks	0.90	[0.82; 0.94]
death	0.90	[0.85; 0.94]
demo	0.82	[0.72; 0.90]

three binary variables. In the second one the NM-PLSPM algorithm is run, in order to obtain quantifications coherent with the path model. The quantitative MVs are transformed at the ordinal scaling level to discard the hard assumption of linearity in favor of a milder assumption of monotonicity, and the categorical MV *demo* is analyzed at the nominal scaling level. In both the cases, according to Tenenhaus' previous works, Agricultural Inequality and Industrial Development have been modeled as predictors of Political Instability, as shown in Figure 1; the option *centroid* and both *Mode A* and *new Mode A* estimation procedures are used. Since the two estimation procedures provide outputs identical up to the second decimal digit, the outputs presented in Table 2 can be referred to both, and only the differences between the standard and the non-metric PLS models

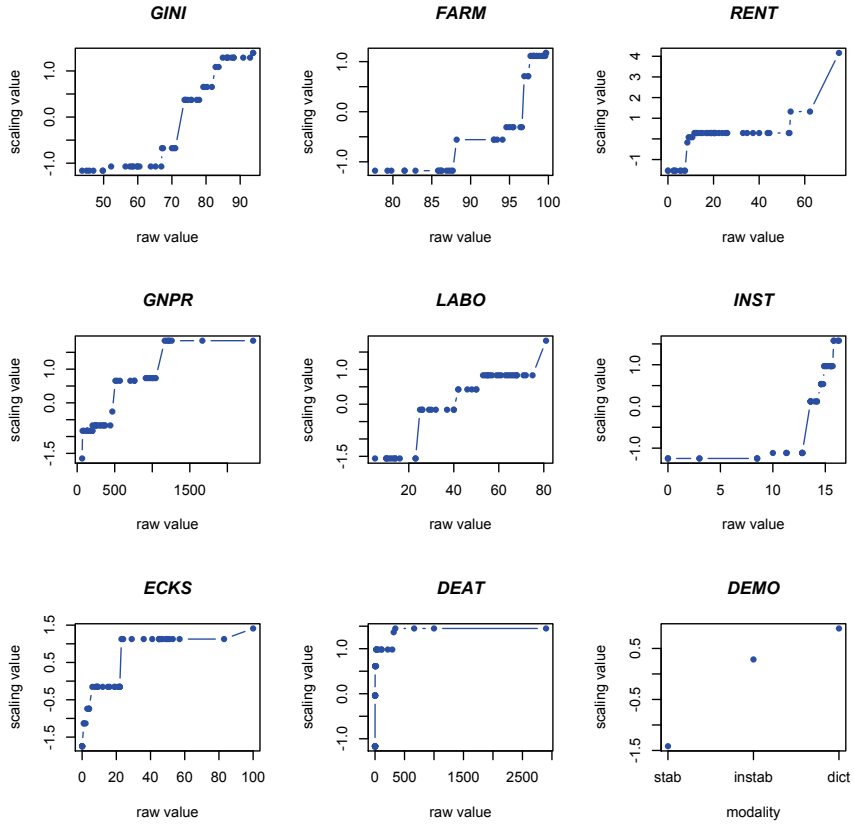


FIG 2. *Variable original values plotted versus corresponding optimal scaling values.*

will be commented. The interpretations of the two inner models are similar: industrial development is the main driver for political instability, on which it has a negative impact. As one can expect, agricultural inequality positively impacts on political instability. However, the criterion optimized by the PLS iteration increases from 2.29 to 3.42. The improvement is due to the monotone transformations of the quantitative MVs (shown in Figure 2), and it is reflected in both the inner and the outer models. The LVs Agricultural Inequality and Industrial Development are more strongly related to Political Instability in NM-PLS model than in the standard PLS model, providing a greater explained variance (the R^2 index increases from 0.60 to 0.76). Moreover, the LVs are on average more correlated to the corresponding MVs; the manifest variables *rent*, *inst*, *ecks* and *deat* show sensibly increased loadings, due to their marked non-linearity.

The comparison of the two models (Table 2) suggests two further remarks:

- In this example the introduction of the scaling parameters did not affect the stability of the estimates and allowed to discover a statistically significant loading for variable *inst*;
- In non-metric approach the loading of the MV *demo* clearly points out a strong relation with the LV Political Instability. On the contrary, in the standard PLS model the three dummy variables labeled as *stab*, *unstab* and *dict* show quite different loadings, which do not make easy the interpretation of the role of the variable *demo* as a whole. Indeed, there is a strong relation between Political Instability and all of the binary variables representing the categories of MV *demo*. However, while relations between binary variables *dict* and *stab* and Political Instability are pretty monotone (and so they can be easily detected in a linear framework), the binary variable *unstab* is non-monotonically related to political instability and consequently its importance is underestimated. This drawback of the binary coding has been already discussed in section 3.

10. Discussion

Ten years after Herman Wold proposed the Nonlinear Iterative LEast Squares (NILES, later renamed NIPALS) procedure [35], an iterative algorithm based on an Alternating Least Squares (ALS) procedure [16] was proposed by Jan de Leeuw, Yoshio Takane and Forrest W. Young for implementing optimal scaling in additive structure analysis. The authors themselves noticed, in respect to ALS, that "*this type of procedure is philosophically much like the NILES/NIPALS procedure developed by Wold and his associates with the distinction that Wold is usually concerned with optimizing only model parameters*" [16].

Over the last 35 years these two procedures have been developed independently for completely different aims. NIPALS-based algorithms have been developed for implementing a number of well-known two- and multi-block analysis methods (see [30] for a review), as well as two new approaches: a component-based regularized regression (the PLS-R) and a predictive Path Modeling in a soft modeling framework (the PLS-PM). Meanwhile, ALS has become the most widely used procedure for optimal scaling in joint non-parametric multivariate analysis of non-metric and metric data. A whole system of non-linear multivariate analysis, based on ALS principles, has been developed by the data theory group of Leiden University [5]. This article shows that PLS algorithms, correctly adjusted, can also work as optimal scaling algorithms. This new and previously totally unexplored feature of PLS make it possible to devise one, two and multi-block PLS methods with optimal scaling features. NM-PLS methods preserve all the features of standard PLS algorithms, since

- they work on landscape matrices;
- they handle missing data without the need for advance data imputations;
- they do not require hard assumptions about population distribution.

Moreover, with NM-PLS methods it is possible to:

- introduce into the models variables observed on different measurement scales;
- investigate non linearity;
- discard the hard assumption of linearity in favor of a milder assumption of monotonicity.

NM-PLS scalings have been proved to be suitable, coherent and optimal. They are suitable because they respect the relevant constraints depending on which properties of the original measurement scale are to be preserved. They are coherent because the statistical relationship of a latent construct with the associated raw variable is reflected into its statistical relationship with the corresponding quantified variable. Finally, they are optimal because they optimize the criterion of the model in which they are involved. However, NM-PLS quantifications are optimal with respect to the first component, while the scaling parameters optimality in multiple component models is not assured. This is not a relevant drawback in PLS-PM framework, where multidimensional latent variables are rarely investigated, but it represents a restriction in Non-Metric PLS approaches to regression and Principal Component Analysis (PCA), which require multi-component models in most of the applications. The impossibility to search for quantifications optimal with respect to multi-component models is due to the fact that the PLS algorithms extract components sequentially, by alternating an estimation step and a deflation step. As a consequence, the extraction of each successive component is conditioned to the knowledge of the previous ones. This procedure is efficient in standard PLS algorithms because PLS solutions are nested, but it prevents from ensuring the optimality of the parameters in multi-component non-metric models, as NM-PLS solutions are not nested.

NM-NIPALS, NM-PLSR and NM-PLSPM algorithms are implemented in R routines [20] which are available from the author upon request.

Acknowledgements

I would like to express my thanks to Dr. Laura Trinchera. The extension of NM-PLS principles to PLS-PM algorithm would not have been possible without her precious assistance.

References

[1] Benzécri, J.P. (1973). *L'analyse des Données*, Dunod, Paris.
[2] Bollen, K.A. (1989). *Structural equations with latent variables*, Wiley, New York. MR0996025
[3] Efron, B. (1982). *The Jackknife, the bootstrap, and other resampling plans*, SIAM, Philadelphia. MR0659849
[4] Esposito Vinzi, V., Chin W.W., Henseler, J. and Wang, H. (Eds.) (2010). *Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications*, Springer Verlag, Berlin Heidelberg. MR2742562

[5] Gifi, A. (1990). *Nonlinear Multivariate Analysis*, Wiley, Chichester, UK. MR1076188
[6] Hanafi, M. (2007). PLS Path Modeling: computation of latent variables with estimation mode B, *Computational Statistics* **22**, 275-292. MR2318460
[7] Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view, *Annals of the Institute of Statistical Mathematics* **3**, 69-98.
[8] Höskuldsson, A. (1988). PLS Regression Methods, *Journal of Chemometrics* **2**, 211-228.
[9] de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression, *Chemometrics and Intelligent Laboratory Systems* **18**, 251-263.
[10] Krämer, N. (2007). *Analysis of high-dimensional data with partial least squares and boosting*, Phd Thesis,Technischen Universität Berlin, Germany.
[11] Kruskal, J. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika*, **29** (1), 1-27. MR0169712
[12] Kruskal, J. (1964). Nonmetric multidimensional scaling: A numerical method, *Psychometrika* **29** (2), 115-129. MR0169713
[13] Kruskal, J. and Shepard, R. (1974). A nonmetric variety of linear factor analysis, *Psychometrika* **39** (2), 123-157. MR0353575
[14] Jakobowicz, E. and Darquenne, C. (2007). A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy, *Computational Statistics & Data Analysis* **51** (8), 3666-3678. MR2364482
[15] Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika* **57**, 239-251. MR0269024
[16] de Leeuw, J., Young, F.W. and Takane Y. (1976). Additive structure in qualitative data: an alternating least squares method with optimal scaling features, *Psychometrika* **41**, 471-503.
[17] Lohmöller, J.B. (1989). *Latent variable path modeling with partial least squares*, Physica-Verlag, Heildelberg. MR1014472
[18] Michailidis, G., de Leeuw, J. (1998). The Gifi system for descriptive multivariate analysis, *Statistical Science* **13** (4), 307-336. MR1705265
[19] Lyttkens, E. (1966). On the fix-point property of Wold's iterative estimation method for principal components. In: Krishnaiah P. R. (Ed.), *Multivariate Analysis*, Academic Press New York, 335-350. MR0211544
[20] R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. Available at: `http://www.r-project.org`.
[21] Rosipal, R. and Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. In: C. Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (Eds.), *Subspace, Latent Structure and Feature Selection Techniques*, Springer, 34-51.
[22] Russett, B.M. (1964), Inequality and Instability: the relation of land tenure to politics, *World Politics* **16**, 442-454.

[23] Russolillo, G. (2009), *PLS methods for Non-Metric data*, PhD thesis, Università degli Studi di Napoli "Federico II", Italy.

[24] Russolillo, G. and Lauro N.C. (2011). A Proposal for Handling Categorical Predictors in PLS Regression Framework. In: Fichet, B.; Piccolo, D.; Verde, R.; Vichi, M. (Eds.), *Classification and Multivariate Analysis for Complex Data Structures*, Springer-Verlag France, 341-348.

[25] Russolillo, G., Trinchera L. and Esposito Vinzi, V. (2009). A Non Linear Regularized Component based Approach to Structural Equation Modeling, *Book of short papers - Convegno Intermedio SIS-2009 Statistical Methods for the analysis of large data-sets*, CLEUP.

[26] Sampson, P.D., Streissguth, A.P., Barr, H.M. and Bookstein, F.L. (1989). Neurobehavioral effects of prenatal alcohol: Part II. Partial Least Squares analysis, *Neurotoxicology and tetralogy* **11** (5), 477-491.

[27] Tenenhaus, M. (1977). Analyse en composantes principales d'un ensemble de variables nominales ou numéric, *Revue de statistique appliquée* **25**, 39-56. MR0483190

[28] Tenenhaus, M. (1998). *La Régression PLS: théorie et pratique*, Technip, Paris. MR1645125

[29] Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., Lauro, N.C. (2005). PLS Path Modeling, *Computational Statistics & Data Analysis* **48**, 159-205. MR2134489

[30] Tenenhaus and M, Hanafi, M. (2010), A bridge between PLS path modeling and multi-block data analysis. In: Esposito Vinzi, V., Chin W.W., Henseler, J., Wang, H. (Eds.), *Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications*, Springer Verlag, Berlin Heidelberg. MR2742562

[31] Tenenhaus, A. and Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika* **76** (2), 257-284. MR2788885

[32] Tishler, A., Dvir, D., Shenhar, A. and Lipovetsky, S. (1996). Identifying critical success factors in defense development projects: a multivariate analysis, *Technological forecasting and social change*, **51**, 151-171.

[33] Tucker, L.R. (1958). An Inter-Battery Method of Factor Analysis, *Psychometrika* **23**, 111-136. MR0099737

[34] Wegelin, J.A. (2000). *A survey on Partial Least Squares (PLS) methods, with emphasis on the two-block case*, Technical Report N.371, Departement of Statistics, University of Washington, USA: Seattle.

[35] Wold, H. (1966). Estimation of principal component and related models by iterative least squares. In: P. R. Krishnaiah (Ed.), *Multivariate Analysis*, Academic Press, New York, 391-420. MR0220397

[36] Wold, H. (1975). Path Models with Latent Variables: The non-linear iterative partial least squares (NIPALS) approach. In: Blalock, H.M., Aganbegian, A., Borodkin F.M., Boudon, R., Capecchi, V. (Eds.), *Quantitative Sociology: Intentional Perspective on Mathematical and Statistical Modeling*, Academic Press, New York, 307-357. MR0444120

[37] Wold, H. (1985). Partial Least Squares. In: Kotz, S., Johnson, N.L. (Eds.), *Encyclopedia of Statistical Sciences (vol. 6)*, Wiley, New York, pp. 581-591.

[38] Wold, S. (1984), Three PLS algorithms according to SW. In: *Report from the symposium MULTDAST (multivariate analysis in science and methodology)*, Wold, S. (Ed.), Umeå, Sweden, 26-30.

[39] Wold, S., Martens, H. and Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In: A. Ruhe, Kagstrom, B. (Eds.), *Proceedings of the Conference on Matrix Pencils. Lectures Notes in Mathematics*, Springer, Heidelberg. MR0697741

[40] Young, F.W. (1975). Methods for describing ordinal data with cardinal models, *Journal of Mathematical Psychology* **12**, 416-436.

[41] Young, F.W. (1981). Quantitative analysis of qualitative data. *Psychometrika* **44** (4), 357-388. MR0668307

[42] Young, F.W., Takane, Y. and de Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features, *Psychometrika* **43** (2), 279-281.