NeuroImage 56 (2011) 455–475

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review

Anjali Krishnan [a], Lynne J. Williams [b], Anthony Randal McIntosh [c,d,*], Hervé Abdi [a,*]

[a] School of Behavioral and Brain Sciences, The University of Texas at Dallas, MS: GR4.1, 800 West Campbell Road Richardson, TX 75080-3021, USA
[b] The Kunen-Luenfeld Applied Research Unit, The Rotman Research Institute, Baycrest, 3560 Bathurst Street, Toronto, ON, Canada M6A 2E1
[c] Department of Psychology, Sidney Smith Hall, 4th Floor, University of Toronto, 100 St. George Street, Toronto, ON, Canada M5S 3G3
[d] The Rotman Research Institute, Baycrest, 3560 Bathurst Street, Toronto, ON, Canada M6A 2E1

ARTICLE INFO

Article history:
Received 21 January 2010
Revised 1 July 2010
Accepted 19 July 2010
Available online 23 July 2010

Keywords:
Partial least squares correlation
Partial least squares regression
Partial least squares path modeling
PLS
Symmetric PLS
Asymmetric PLS
Task PLS
Behavior PLS
Seed PLS
Multi-block PLS
Multi-table PLS
Canonical variate analysis
Co-inertia analysis
Multiple factor analysis
STATIS
Barycentric discriminant analysis
Multiple factor analysis
Common factor analysis

ABSTRACT

Partial Least Squares (PLS) methods are particularly suited to the analysis of relationships between measures of brain activity and of behavior or experimental design. In neuroimaging, PLS refers to two related methods: (1) symmetric PLS or Partial Least Squares *Correlation* (PLSC), and (2) asymmetric PLS or Partial Least Squares *Regression* (PLSR). The most popular (by far) version of PLS for neuroimaging is PLSC. It exists in several varieties based on the type of data that are related to brain activity: *behavior* PLSC analyzes the relationship between brain activity and behavioral data, *task* PLSC analyzes how brain activity relates to pre-defined categories or experimental design, *seed* PLSC analyzes the pattern of connectivity between brain regions, and *multi-block* or *multi-table* PLSC integrates one or more of these varieties in a common analysis. PLSR, in contrast to PLSC, is a predictive technique which, typically, predicts behavior (or design) from brain activity. For both PLS methods, statistical inferences are implemented using cross-validation techniques to identify significant patterns of voxel activation. This paper presents both PLS methods and illustrates them with small numerical examples and typical applications in neuroimaging.

Introduction

Originally developed for econometrics and chemometrics (Wold, 1982), Partial Least Squares (PLS) is a multivariate statistical technique first introduced to functional neuroimaging by McIntosh et al. (1996) with the goal of extracting commonalities between brain activity and behavior or experimental design. In neuroimaging there are two basic types of PLS methods, which we call Partial Least Squares Correlation (PLSC; McIntosh et al., 1996), and Partial Least Squares Regression (PLSR; Wold, 1982; de Jong, 1993; Wold et al., 2001). PLSC (Tucker, 1958; Bookstein, 1982; Streissguth et al., 1993; Bookstein, 1994; McIntosh et al., 1996) is a correlational technique that analyzes *associations* between two sets of data (e.g., behavior and brain activity), while PLSR (Wold, 1982; Martens and Naes, 1989; de Jong and Phatak, 1997; Tenenhaus, 1998; Martens and Martens, 2001; Wold et al., 2001; Abdi, 2010) is a regression technique that *predicts* one set of data from another (e.g., predicts behavior from brain activity). A third, closely related, technique called partial least squares path modeling (see, e.g., Esposito-Vinzi et al., 2010 for a recent comprehensive review) can be seen as a least squares equivalent of structural equation modeling (which is a maximum likelihood technique). Despite PLS path modeling's obvious relevance, this method has not yet been applied to neuroimaging, and therefore we will not include it in this review. For both PLSC and PLSR, statistical inferences are implemented using computational cross-validation methods (e.g., jackknife, bootstrap). As a distinct advantage, PLS techniques are tailored to handle the very large data sets which are typical of current neuroimaging research.

In this paper we will present PLSC, PLSR and their main variants used in neuroimaging. We introduce each technique with a small artificial example in order to describe the main computational steps. For each technique we also present and review major applications

* Corresponding authors.
E-mail addresses: herve@utdallas.edu (H. Abdi), rmcintosh@rotman-baycrest.on.ca (A.R. McIntosh).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.07.034

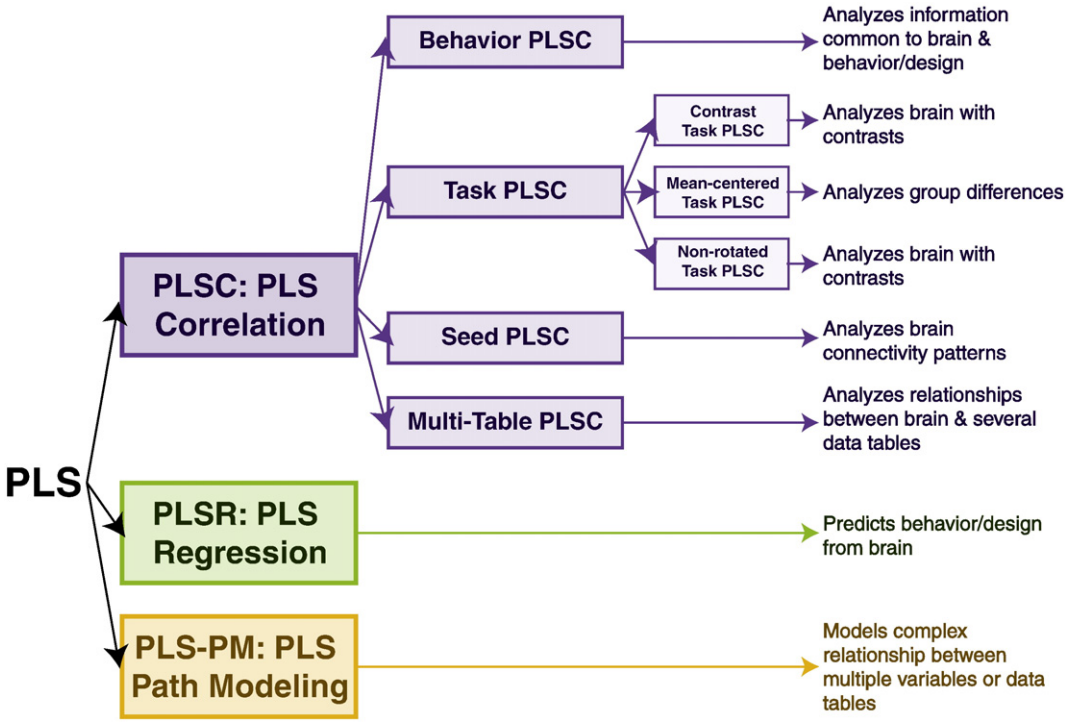

Fig. 1. The PLS family.

from the neuroimaging literature. A diagram outlining the various PLS methods is shown in Fig. 1.

Notations

In this section, we review the main notations used in this paper. For convenience, Appendix A also lists our main notations and acronyms (see also Abdi and Williams, 2010c, for more details on matrices).

Data are stored in matrices which are denoted by upper case bold letters (e.g., **X**). The *identity* matrix is denoted **I**. Column vectors are denoted by lower case bold letters (e.g., **x**). Matrix or vector transposition is denoted by an uppercase superscript T (e.g., $\mathbf{X}^T$). Two bold letters placed next to each other imply matrix or vector multiplication unless otherwise mentioned. The number of rows, columns, or sub-matrices is denoted by an uppercase italic letter (e.g., I) and a given row, column, or sub-matrix is denoted by a lowercase italic letter (e.g., i).

Brain activity is stored in an I by J matrix denoted **X** whose generic element is denoted $x_{i,j}$ and where the rows are observations and the columns are variables. Matrix **X** is made up of N *a priori* sub-matrices, with I_n being the number of observations in sub-matrix n. The sum of the number of observations in all of the sub-matrices is the number of rows of **X** (i.e., $I = \sum I_n$; see Fig. 2a). When dealing with spatio-temporal neuroimaging methods (e.g., EEG, *f*MRI, NIRS), there are T scans where the set of scans for all I observations at time t corresponds to an I by J_t matrix denoted $\mathbf{X}_t$. The $\mathbf{X}_t$ matrices are concatenated by row to form the larger matrix **X** (whose total number of columns J is the sum of all the J_t; see Fig. 3).

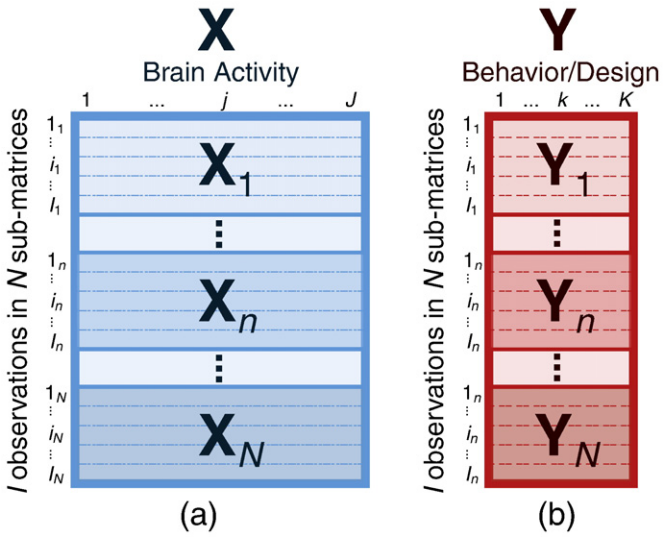

Fig. 2. Data representation for matrices (a) **X** and (b) **Y**. Note that the I observations of **X** and **Y** are composed of N sub-matrices, $\mathbf{X}_1 \ldots \mathbf{X}_n \ldots \mathbf{X}_N$ and $\mathbf{Y}_1 \ldots \mathbf{Y}_n \ldots \mathbf{Y}_N$ representing the groups or trial types.