NeuroImage xxx (2010) xxx–xxx

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 xxxx

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.

1. 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

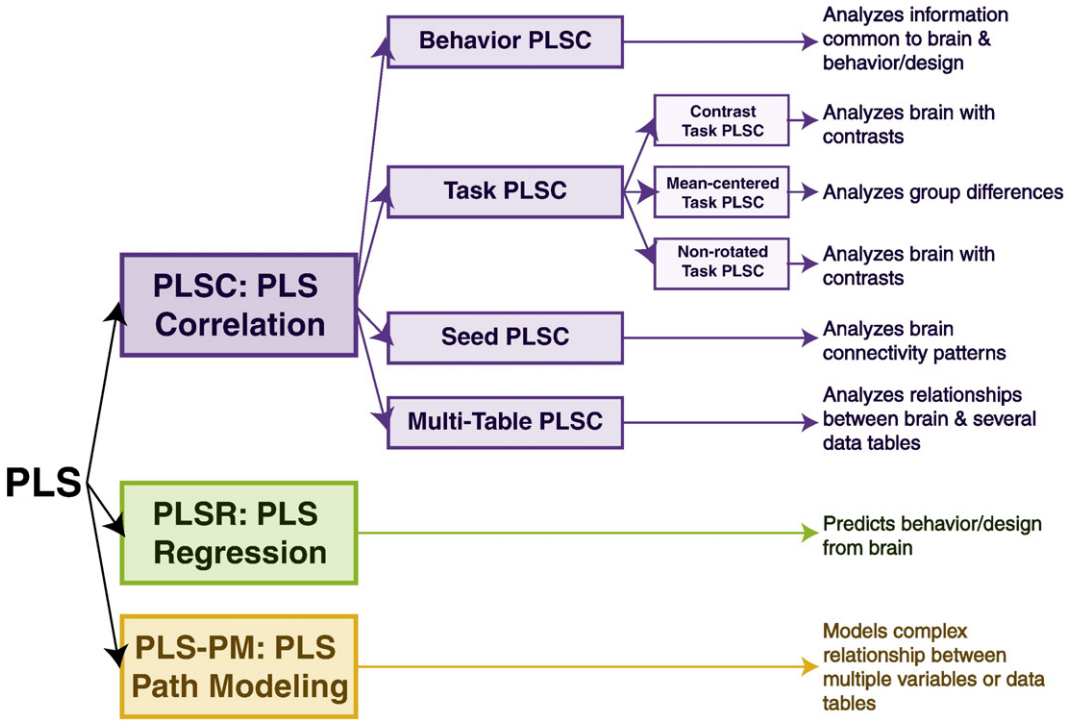

Fig. 1. The PLS family.

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

2. 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., $\mathbf{X}$). The *identity* matrix is denoted $\mathbf{I}$. Column vectors are denoted by lower case bold letters (e.g., $\mathbf{x}$). 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 $\mathbf{X}$ whose generic element is denoted $x_{i,j}$ and where the rows are observations and the columns are variables. Matrix $\mathbf{X}$ 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 $\mathbf{X}$ (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 $\mathbf{X}$ (whose total number of columns J is the sum of all the J_t; see Fig. 3).

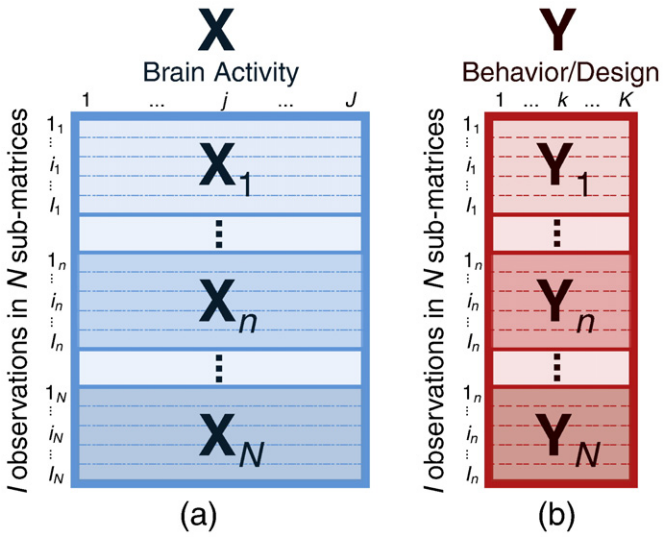

Fig. 2. Data representation for matrices (a) $\mathbf{X}$ and (b) $\mathbf{Y}$. Note that the I observations of $\mathbf{X}$ and $\mathbf{Y}$ are composed of N sub-matrices, $\mathbf{X}_1 \ldots \mathbf{X}_n \ldots \mathbf{X}_N$ and $\mathbf{Y}_1 \ldots \mathbf{Y}_n \ldots \mathbf{Y}_N$ representing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

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

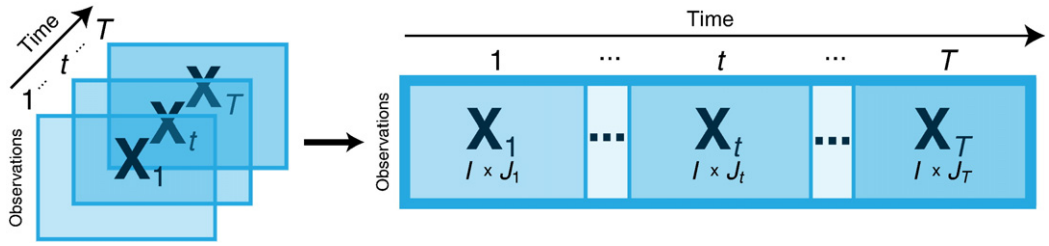

Fig. 3. Data Matrix for EEG, MEG, *f*MRI, and NIRS experiments: The data (scans) **X** for each time point *t* has *I* rows and *J* columns. Each scan from $\mathbf{X}_1$ to $\mathbf{X}_T$ is concatenated by rows so that each row contains all the voxel activity at all time points.

Behavioral or design data are stored in an *I* by *K* matrix, **Y**, whose generic element is denoted $y_{i,k}$. Each **Y** matrix is made up of the same *N a priori* sub-matrices as **X** (Fig. 2b), with *K* behavioral measures or contrasts that code for different aspects of the experimental design (e.g., category membership). In general, matrices **X** and **Y** are statistically preprocessed in order to make the variables comparable. For example, the mean of each column can be subtracted from all its elements or each column can be transformed into *Z*-scores. In some cases a normalization is performed on a sub-matrix basis (e.g., the mean of each column of each sub-matrix is zero and its standard deviation is one).

2.1. The main tool: the singular value decomposition

The main analytical tool for PLS is the singular value decomposition (SVD) of a matrix (see Abdi, 2007; de Leeuw, 2007; Greenacre, 1984; Takane, 2002 for details and tutorials). Recall that the SVD of a given $J \times K$ matrix **Z** decomposes it into three matrices as:

$$\mathbf{Z} = \mathbf{U}\Delta\mathbf{V}^T = \sum_{\ell}^{L} \delta_{\ell}\mathbf{u}_{\ell}\mathbf{v}_{\ell}^T \tag{1}$$

where **U** is the *J* by *L* matrix of the normalized left singular vectors (with *L* being the rank of **Z**), **V** the *K* by *L* matrix of the normalized right singular vectors, Δ the *L* by *L* diagonal matrix of the *L* singular values, also δ_{ℓ}, $\mathbf{u}_{\ell}$, and $\mathbf{v}_{\ell}$ are respectively the ℓth singular value, left, and right singular vectors. Matrices **U** and **V** are orthonormal matrices (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$). The SVD is closely related to and generalizes the well-known *eigen-decomposition* as **U** is also the matrix of the normalized eigenvectors of $\mathbf{Z}\mathbf{Z}^T$, **V** is the matrix of the normalized eigenvectors of $\mathbf{Z}^T\mathbf{Z}$, and the singular values are the square root of the eigenvalues of $\mathbf{Z}\mathbf{Z}^T$ and $\mathbf{Z}^T\mathbf{Z}$ (these two matrices have the same eigenvalues). Key property: the SVD provides the best reconstitution (in a least squares sense) of the original matrix by a matrix with a lower rank.

3. Partial Least Squares Correlation

Partial Least Squares Correlation (PLSC) analyzes the relationship between the matrices **X** and **Y**. These matrices store measurements collected on the same observations (e.g., scans or participants). The *I* by *J* matrix **X** corresponds to brain activity and the *I* by *K* matrix **Y** corresponds to behavioral or design variables. The relationship between the *j*th column of **X** and the *k*th column of **Y** is measured by the dot (i.e., scalar) product between these two columns. When these two columns are centered, the dot product gives the *covariance* between these two columns. When, in addition, these two columns are normalized (e.g., are *Z*-scores, or when the sum of the squared values of each column is equal to one), the dot product expresses the *correlation* between these two columns. Because covariance and correlation are not directional (i.e., correlation and covariance do not depend upon the order of the variables), the roles of **X** and **Y** are symmetric and the analysis focuses on shared information.

3.1. Overview of PLSC

PLSC can be subdivided into four variants: (1) behavior PLSC, (2) task PLSC, (3) seed PLSC, and (4) multi-table or multi-block PLSC (McIntosh and Lobaugh, 2004; McIntosh et al., 1998). The difference between the techniques rests in the **Y** matrix. **Y** is a matrix of behavioral variables in behavior PLSC, a matrix of contrasts or design variables in task PLSC, a matrix of voxel activity from the regions of interest (ROIs) in seed PLSC, and there are multiple **Y** matrices in multi-table PLSC, each consisting of behavioral, design, or ROI variables. Note that for all versions of PLSC, all scans are co-registered onto a common brain space to remove differences in brain volume (e.g., Talairach coordinates; see Talairach and Tournoux, 1988).

3.2. Formal expression of PLSC

Formally, the relationship between the columns of **X** and **Y** are stored in a cross-product matrix, denoted **R**, which is computed as:

$$\mathbf{R} = \mathbf{Y}^T\mathbf{X}. \tag{2}$$

In general, **R** is a matrix of correlations because **X** and **Y** are centered and normalized (e.g., expressed as *Z*-scores).

The SVD [see Eq. (1)] of **R** decomposes it into three matrices:

$$\mathbf{R} = \mathbf{U}\Delta\mathbf{V}^T. \tag{3}$$

In the PLSC vocabulary the *singular vectors* **U** and **V** are also called *saliences* (Bookstein, 1994) and in this paper these two terms are synonymous. The *L* left singular vectors of **R** (i.e., **U**) represent the design or behavioral profiles that best characterize **R**, whereas the *L* right singular vectors of **R** (i.e., **V**) represent the voxels or brain images that best characterize **R**.

To express the saliences relative to brain activity and behavior (or design), the original matrices **X** and **Y** are projected onto their respective saliences. This creates *latent variables*—which are linear combinations of the original variables—that are computed as:

$$\mathbf{L}_{\mathbf{X}} = \mathbf{X}\mathbf{V}, \tag{4}$$

where the matrix of latent variables of **X** (i.e., the *I* by *L* matrix $\mathbf{L}_{\mathbf{X}}$) is called "brain scores," and

$$\mathbf{L}_{\mathbf{Y}} = \mathbf{Y}\mathbf{U}, \tag{5}$$

where the matrix of latent variables of **Y** (i.e., the *I* by *L* matrix $\mathbf{L}_{\mathbf{Y}}$) is called "behavior" or "design scores." A pair of vectors $\ell_{\mathbf{X},\ell}$ (which is the ℓth column of $\mathbf{L}_{\mathbf{X}}$) and $\ell_{\mathbf{Y},\ell}$ (which is the ℓth column of $\mathbf{L}_{\mathbf{Y}}$) reflects a relationship between brain activity and behavior (Nestor et al., 2002). PLSC searches for latent variables that express the largest amount of information common to both **X** and **Y**. Specifically, PLSC computes latent variables with maximal covariance.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

3.2.1. What does PLSC optimize?

The goal of PLSC is to find pairs of latent vectors $\ell_{\mathbf{X},\ell}$ and $\ell_{\mathbf{Y},\ell}$ with maximal covariance and with the additional constraints that (1) the pairs of latent vectors made from two different indices are uncorrelated and (2) the coefficients used to compute the latent variables are normalized (see Tucker, 1958; Tenenhaus, 1998, for proofs). Formally, we want to find

$$\ell_{\mathbf{X},\ell} = \mathbf{X}\mathbf{v}_\ell \quad \text{and} \quad \ell_{\mathbf{Y},\ell} = \mathbf{Y}\mathbf{u}_\ell \tag{6}$$

$$\text{such that } \operatorname{cov}\left(\ell_{\mathbf{X},\ell}, \ell_{\mathbf{Y},\ell}\right) \propto \ell_{\mathbf{X},\ell}^{\mathbf{T}} \ell_{\mathbf{Y},\ell} = \max$$

[where $\operatorname{cov}(\ell_{\mathbf{X},\ell}, \ell_{\mathbf{Y},\ell})$ denotes the covariance between $\ell_{\mathbf{X},\ell}$ and $\ell_{\mathbf{Y},\ell}$] under the constraints that

$$\ell_{\mathbf{X},\ell}^{\mathbf{T}} \ell_{\mathbf{Y},\ell'} = 0 \text{ when } \ell \neq \ell' \tag{7}$$

(note that $\ell_{\mathbf{X},\ell}^{\mathbf{T}} \ell_{\mathbf{X},\ell'}$ and $\ell_{\mathbf{Y},\ell}^{\mathbf{T}} \ell_{\mathbf{Y},\ell'}$ are *not* required to be null) and

$$\mathbf{u}_\ell^T \mathbf{u}_\ell = \mathbf{v}_\ell^T \mathbf{v}_\ell = 1. \tag{8}$$

It follows from the properties of the SVD (see, e.g., Abdi and Williams, 2010d; de Leeuw, 2007; Greenacre, 1984; Takane, 2002) that $\mathbf{u}_\ell$ and $\mathbf{v}_\ell$ are singular vectors of $\mathbf{R}$. In addition, from Eqs. (3–5), the covariance of a pair of latent variables $\ell_{\mathbf{X},\ell}$ and $\ell_{\mathbf{Y},\ell}$ is equal to the corresponding singular value:

$$\ell_{\mathbf{X},\ell}^{\mathbf{T}} \ell_{\mathbf{Y},\ell} = \delta_\ell. \tag{9}$$

So, when $\ell = 1$, we have the largest possible covariance between the pair of latent variables. When $\ell = 2$ we have the largest possible *covariance* for the latent variables under the constraints that the latent variables are uncorrelated with the first pair of latent variables [as stated in Eq. (7), e.g., $\ell_{\mathbf{X},1}$ and $\ell_{\mathbf{Y},2}$ are uncorrelated], and so on for larger values of ℓ.

3.3. Deciding which latent variables to keep

The SVD of $\mathbf{R}$ corresponds to a *fixed effect* model; therefore, the results can only be interpreted with respect to the original data sets. Yet, in the framework of PLSC, the goal is to extract information common to the two sets of data (e.g., brain activity and behavioral measures) which can generalize to the population (i.e., a *random effect* model; Abdi, 2010).

To generalize the results (i.e., to create a random effect model), we could use an inferential analytical approach such as the one defined by Tucker (1958), but this approach makes too many parametric assumptions to be used routinely. Instead, we use computational approaches, such as permutation tests, to obtain *p*-values, which can then be used to identify the generalizable latent variables (McIntosh and Lobaugh, 2004; McIntosh et al., 2004). In a permutation test, a new data set, called a *permutation sample*, is obtained by randomly reordering the rows (i.e., observations) of $\mathbf{X}$ and leaving $\mathbf{Y}$ unchanged. The PLSC model used to compute the fixed effect model is then recomputed for the permutation sample to obtain a new matrix of singular values. This procedure is repeated for a large number of permutation samples, say 1000 or 10,000. The set of all the singular values provides a sampling distribution of the singular values under the null hypothesis and, therefore can be used as a null hypothesis test.

When a vector of saliences is considered generalizable and is kept for further analysis, we need to identify its elements that are *stable* through resampling. In practice, the stability of an element is evaluated by dividing it by its standard error. Specifically, if $\widehat{\sigma}(u_i)$ and $\widehat{\sigma}(v_i)$ denote the standard errors of u_i and v_i, the stability of the ith element of $\mathbf{u}$ and $\mathbf{v}$ are obtained (respectively) as:

$$\frac{u_i}{\widehat{\sigma}(u_i)} \quad \text{and} \quad \frac{v_i}{\widehat{\sigma}(v_i)}. \tag{10}$$

To estimate the standard errors, we create *bootstrap samples* which are obtained by sampling *with replacement* the observations in $\mathbf{X}$ and $\mathbf{Y}$ (Efron and Tibshirani, 1986). A salience standard error is then estimated as the standard error of the saliences from a large number of these bootstrap samples (say 1,000 or 10,000). The ratios from Eq. (10) are akin to a *Z*-score, therefore when they are larger than 2 the corresponding saliences are considered significantly stable. Stable saliences determine which voxels show reliable responses to the experimental conditions (McIntosh et al., 2004; Efron and Tibshirani, 1986) and indicate the important saliences in the brain activity network (Mentis et al., 2003).

As a technical aside, one problem when using permutation or bootstrap methods is that resampling may cause axis rotation or reflection. Axis rotation refers to a change in the order of the latent variables that are extracted with each permutation. Reflection refers to a change in the sign of the saliences for each bootstrap sample. A Procrustes rotation or a variation of multidimensional scaling, such as DISTATIS (Abdi et al., 2005; Abdi et al., 2009a), can be used to correct for these rotations and reflections (see McIntosh and Lobaugh, 2004, for more details).

Although it may seem that using both permutation tests and the bootstrap is somewhat redundant, these two methods provide, in fact, different information. Permutation tests indicate whether a signal can be differentiated from noise, but do not index signal reliability which is, by contrast, provided by the bootstrap (see McIntosh and Lobaugh, 2004, for a discussion of detection and reliability).

3.4. PLS: mini-example

In order to illustrate the various versions of PLS we have chosen a hypothetical neuroimaging experiment that analyzes data from participants grouped into three experimental conditions comprising two clinical populations, Alzheimer's disease (AD) and Parkinson's disease (PD), and an age-matched normal control group (NC). Each participant is scanned once using PET imaging. Voxel activity values from the PET scans are collected separately for each participant and stored in $\mathbf{X}$. The scans are coregistered to Talairach coordinates so that voxels are in the same location for each participant. In this example, we have three participants per clinical category (hence a total of $3 \times 3 = 9$ participants). The PET results are:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} = \begin{bmatrix} 2 & 5 & 6 & 1 & 9 & 1 & 7 & 6 & 2 & 1 & 7 & 3 \\ 4 & 1 & 5 & 8 & 8 & 7 & 2 & 8 & 6 & 4 & 8 & 2 \\ 5 & 8 & 7 & 3 & 7 & 1 & 7 & 4 & 5 & 1 & 4 & 3 \\ 3 & 3 & 7 & 6 & 1 & 1 & 10 & 2 & 2 & 1 & 7 & 4 \\ 2 & 3 & 8 & 7 & 1 & 6 & 9 & 1 & 8 & 8 & 1 & 6 \\ 1 & 7 & 3 & 1 & 1 & 3 & 1 & 8 & 1 & 3 & 9 & 5 \\ 9 & 0 & 7 & 1 & 8 & 7 & 4 & 2 & 3 & 6 & 2 & 7 \\ 8 & 0 & 6 & 5 & 9 & 7 & 4 & 4 & 2 & 10 & 3 & 8 \\ 7 & 7 & 4 & 5 & 7 & 6 & 7 & 6 & 5 & 4 & 8 & 8 \end{bmatrix}, \tag{11}$$

where the columns give voxel activity and the rows are participants AD_1, AD_2, AD_3 (which correspond to $\mathbf{X}_1$) PD_1, PD_2, PD_3 (which correspond to $\mathbf{X}_2$), NC_1, NC_2 and NC_3 (which correspond to $\mathbf{X}_3$). Matrix $\mathbf{Y}$ (and sometimes the preprocessing of $\mathbf{X}$) will differ based on the version of PLSC (i.e., behavior, task, seed, or multi-table).

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

3.5. Behavior PLSC

Behavior PLSC analyzes the relationship between the behavioral characteristics of groups and their functional brain activity. Matrix **X** contains voxel activity [Eq. (11)] and Matrix **Y**, in this case, contains various demographic (e.g., age) and/or behavioral data (e.g., neuropsychological tests, reaction times).

For our example, (fictitious) participants underwent behavioral testing using a memory test for word recall. The behavioral measures were the number of words correctly recalled and the average reaction time (in ms). The behavioral data are:

$$\mathbf{Y}_{\text{behavior}} = \begin{bmatrix} \mathbf{Y}_{\text{behavior},1} \\ \mathbf{Y}_{\text{behavior},2} \\ \mathbf{Y}_{\text{behavior},3} \end{bmatrix} = \begin{bmatrix} 15 & 600 \\ 19 & 520 \\ 18 & 545 \\ \\ 22 & 426 \\ 21 & 404 \\ 23 & 411 \\ \\ 29 & 326 \\ 30 & 309 \\ 30 & 303 \end{bmatrix}, \tag{12}$$

where the rows are the same participants as in matrix **X** and the columns are the participants' number of words recalled and reaction time scores, respectively. Note that both **X** and **Y** contain information from the same participants and hence have the same number of rows but are likely to have a different number of columns.

Both **X** and **Y** are centered and normalized within each condition n (i.e., each $\mathbf{X}_n$ and $\mathbf{Y}_n$ is centered and normalized independently, and the sum of squares of a column in one condition is equal to 1, note also then when all values are equal to their mean, they are all normalized to zero). This normalization gives the following matrices:

$$\mathbf{X} = \begin{bmatrix} -0.77 & 0.07 & 0.00 & -0.59 & 0.71 & -0.41 & 0.41 & 0.00 & -0.79 & -0.41 & 0.23 & 0.41 \\ 0.15 & -0.74 & -0.71 & 0.78 & 0.00 & 0.82 & -0.82 & 0.71 & 0.57 & 0.82 & 0.57 & -0.82 \\ 0.62 & 0.67 & 0.71 & -0.20 & -0.71 & -0.41 & 0.41 & -0.71 & 0.23 & -0.41 & -0.79 & 0.41 \\ \\ 0.71 & -0.41 & 0.27 & 0.29 & 0.00 & -0.66 & 0.48 & -0.31 & -0.31 & -0.59 & 0.23 & -0.71 \\ 0.00 & -0.41 & 0.53 & 0.51 & 0.00 & 0.75 & 0.33 & -0.50 & 0.81 & 0.78 & -0.79 & 0.71 \\ -0.71 & 0.82 & -0.80 & -0.81 & 0.00 & -0.09 & -0.81 & 0.81 & -0.50 & -0.20 & 0.57 & 0.00 \\ \\ 0.71 & -0.41 & 0.62 & -0.82 & 0.00 & 0.41 & -0.41 & -0.71 & -0.15 & -0.15 & -0.51 & -0.82 \\ 0.00 & -0.41 & 0.15 & 0.41 & 0.71 & 0.41 & -0.41 & 0.00 & -0.62 & 0.77 & -0.29 & 0.41 \\ -0.71 & 0.82 & -0.77 & 0.41 & -0.71 & -0.82 & 0.82 & 0.71 & 0.77 & -0.62 & 0.81 & 0.41 \end{bmatrix} \tag{13}$$

and

$$\mathbf{Y}_{\text{behavior}} = \begin{bmatrix} -0.79 & 0.78 \\ 0.57 & -0.60 \\ 0.23 & -0.17 \\ \\ 0.00 & 0.78 \\ -0.71 & -0.61 \\ 0.71 & -0.17 \\ \\ -0.82 & 0.79 \\ 0.41 & -0.22 \\ 0.41 & -0.57 \end{bmatrix}. \tag{14}$$

The matrix of correlations for each condition n is then computed as (Fig. 4):

$$\mathbf{R}_{\text{behavior},n} = \mathbf{Y}_{\text{behavior},n}^{T}\mathbf{X}_n. \tag{15}$$

All the condition-wise matrices of correlations are stacked one on top of the other to form the combined matrix of correlations $\mathbf{R}_{\text{behavior}}$, which is the input for the SVD. The $\mathbf{R}_{\text{behavior}}$ matrix contains the

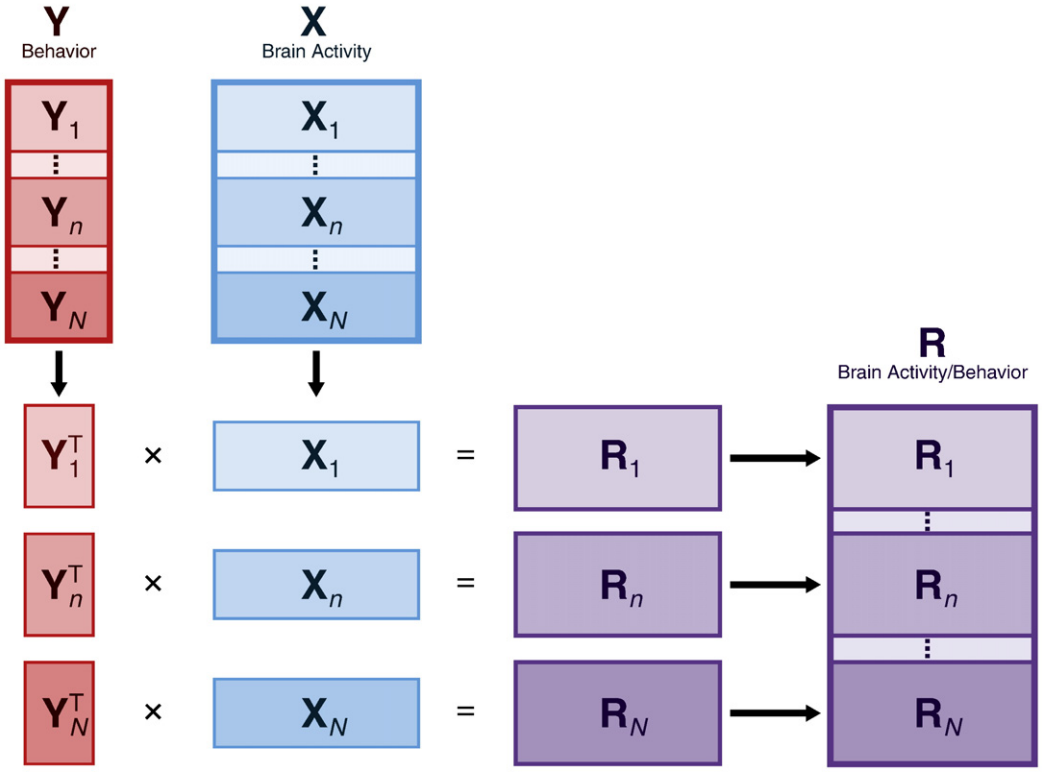

Fig. 4. Matrix **X** and matrix **Y** for Behavior PLSC: The observations are arranged according to N conditions in both matrices and are normalized within condition. The matrix of correlations ($\mathbf{R}_n$) between each condition-wise sub-matrix ($\mathbf{X}_n$ and $\mathbf{Y}_n$) are stacked one below the other to form a combined matrix of correlations **R** which is then decomposed by SVD.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

correlation of each of the J voxels in $\mathbf{X}$ with each of the K behavioral measures in $\mathbf{Y}$ within each of the N conditions. Therefore $\mathbf{R}$ will have $N \times K$ rows and J columns. For our example, the matrix of correlations is:

$$\mathbf{R}_{\text{behavior}} = \begin{bmatrix} \mathbf{R}_{\text{behavior},1} \\ \mathbf{R}_{\text{behavior},2} \\ \mathbf{R}_{\text{behavior},3} \end{bmatrix} = \begin{bmatrix} 0.84 & -0.32 & -0.24 & 0.87 & -0.72 & 0.69 & -0.69 & 0.24 & 1.00 & 0.69 & -0.04 & -0.69 \\ -0.80 & 0.38 & 0.31 & -0.90 & 0.67 & -0.74 & 0.74 & -0.31 & -1.00 & -0.74 & -0.03 & 0.74 \\ -0.50 & 0.87 & -0.94 & -0.93 & 0.00 & -0.60 & -0.81 & 0.92 & -0.92 & -0.69 & 0.96 & -0.50 \\ 0.67 & -0.21 & 0.02 & 0.05 & 0.00 & -0.95 & 0.30 & -0.07 & -0.65 & -0.90 & 0.56 & -0.98 \\ -0.87 & 0.50 & -0.76 & 1.00 & 0.00 & -0.50 & 0.50 & 0.87 & 0.19 & 0.19 & 0.63 & 1.00 \\ 0.96 & -0.70 & 0.90 & -0.97 & 0.25 & 0.70 & -0.70 & -0.96 & -0.43 & 0.06 & -0.80 & -0.97 \end{bmatrix} \tag{16}$$

where each of the $N \times K$ rows gives the kth behavioral measure for the nth condition and where the J columns represent the voxel activity.

The SVD [*cf.*, Eq. (1)] of $\mathbf{R}_{\text{behavior}}$ is computed as:

$$\mathbf{R}_{\text{behavior}} = \mathbf{U}\Delta\mathbf{V}^T, \tag{17}$$

which gives:

$$\mathbf{U} = \begin{bmatrix} 0.41 & -0.42 & 0.37 & -0.02 & -0.14 & -0.71 \\ -0.41 & 0.44 & -0.40 & 0.00 & -0.14 & -0.68 \\ -0.43 & 0.25 & 0.74 & 0.46 & 0.01 & -0.02 \\ -0.07 & 0.31 & 0.40 & -0.86 & 0.08 & -0.01 \\ -0.44 & -0.47 & -0.06 & -0.09 & 0.74 & -0.15 \\ 0.53 & 0.51 & 0.00 & 0.20 & 0.64 & -0.13 \end{bmatrix}, \tag{18}$$

where $\mathbf{U}$ is the $N \times K$ row and L column matrix of the saliences for the behavioral measures, where L is the rank of $\mathbf{R}_{\text{behavior}}$,

$$\Delta = \begin{bmatrix} 3.80 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.46 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.64 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.08 \end{bmatrix}, \tag{19}$$

where Δ is the diagonal matrix of singular values, and

$$\mathbf{V} = \begin{bmatrix} 0.46 & 0.09 & 0.23 & -0.34 & 0.07 & -0.42 \\ -0.32 & -0.04 & 0.10 & 0.24 & -0.28 & -0.54 \\ 0.26 & 0.25 & -0.35 & -0.12 & -0.03 & -0.14 \\ 0.04 & -0.59 & -0.02 & -0.47 & 0.37 & -0.18 \\ -0.12 & 0.22 & -0.22 & 0.04 & 0.49 & 0.30 \\ 0.39 & -0.14 & -0.09 & 0.43 & 0.03 & 0.26 \\ -0.22 & -0.03 & -0.43 & -0.49 & -0.24 & 0.24 \\ -0.28 & -0.28 & 0.33 & 0.13 & 0.13 & 0.05 \\ 0.25 & -0.49 & -0.07 & 0.01 & -0.57 & 0.26 \\ 0.24 & -0.34 & -0.13 & 0.27 & 0.35 & -0.04 \\ -0.30 & -0.09 & 0.36 & -0.15 & 0.05 & 0.32 \\ -0.33 & -0.24 & -0.56 & 0.21 & 0.08 & -0.31 \end{bmatrix}, \tag{20}$$

where $\mathbf{V}$ is the $J \times L$ matrix of the saliences for brain activity.

Behavior saliences (i.e., $\mathbf{U}$) indicate task-dependent differences in the brain-behavior correlation (i.e., the interaction of the experimental conditions with the behavioral measures). Brain saliences (i.e., $\mathbf{V}$) indicate voxel-dependent differences in the brain behavior correlation.

The correlation between the brain scores and each of the behavioral measures gives a pattern of scores similar to the behavior saliences ($\mathbf{U}$), depicting experimental differences in behavior. Confidence intervals on these correlations can be used to assess the reliability of the brain-behavior relationship (McIntosh and Lobaugh, 2004).

A separate bar plot of each column of $\mathbf{U}$ against each behavioral measure shows how the experimental conditions (i.e., groups) interact with the behavioral measures. Fig. 5 shows that the first behavior salience differentiates between the word recall performance of the AD group from the PD and NC group and differentiates between the reaction times of the AD and NC groups. The second behavior salience differentiates the word recall performance of the PD group from the other two groups, but does not differentiate between the reaction times for all three groups.

From the saliences, we compute the latent variables [see Eqs. (4) and (5)]. Recall that the $\mathbf{X}$ latent variables, $\mathbf{L}_\mathbf{X}$, are called *brain scores* and the $\mathbf{Y}$ latent variables, $\mathbf{L}_\mathbf{Y}$, are called *behavior scores*. Both $\mathbf{L}_\mathbf{X}$ and $\mathbf{L}_\mathbf{Y}$ have I rows and L columns.

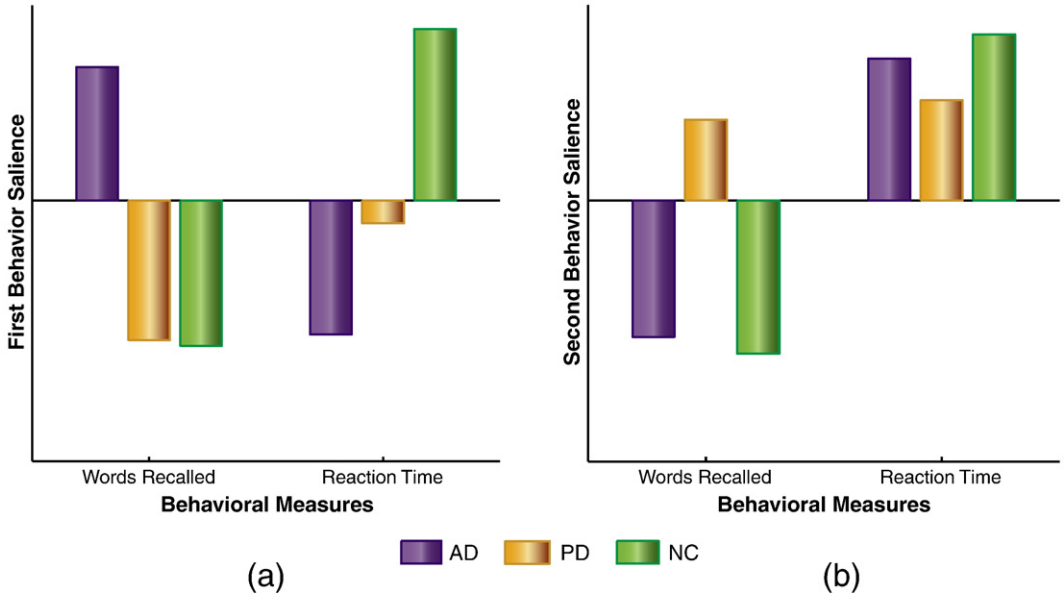

Fig. 5. Mini-example—Behavior PLSC: Behavior saliences 1 and 2. (a) The first behavior salience reveals the difference between AD and the other groups for words recalled and the difference between AD and NC for reaction time; (b) The second behavior salience reveals the difference between PD and the other groups for words recalled but no difference in reaction times for all groups.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

The latent variables for brain activity (brain scores) are computed as (*cf.*, Eq. (4)):

$$\mathbf{L_X} = \mathbf{XV} = \begin{bmatrix} -1.23 & 0.90 & -0.49 & 0.13 & 0.30 & 0.35 \\ 0.90 & -1.31 & 1.21 & 0.31 & 0.78 & 0.90 \\ 0.33 & 0.41 & -0.72 & -0.44 & -1.08 & -1.25 \\ 0.21 & 0.64 & 0.35 & -1.41 & 0.02 & 0.00 \\ 1.05 & -0.89 & -1.46 & 0.17 & -0.01 & 0.00 \\ -1.25 & 0.25 & 1.10 & 1.24 & -0.01 & 0.00 \\ 1.38 & 1.24 & 0.12 & 0.12 & -0.20 & 0.01 \\ 0.34 & -0.11 & -0.51 & 0.42 & 1.36 & -0.06 \\ -1.73 & -1.13 & 0.39 & -0.55 & -1.16 & 0.05 \end{bmatrix}. \quad (21)$$

The computation of the behavior scores is slightly more complicated due to the block structure of the data. Recall that $\mathbf{R}_{\text{behavior}}$ has $N \times K$ rows where each row depicts the kth behavioral measure for the nth condition. In our example, the first two rows of $\mathbf{R}_{\text{behavior}}$ represent the behavioral measures (i.e., words recalled and reaction time) for AD, the next two rows of $\mathbf{R}_{\text{behavior}}$ represent the behavioral measures for PD and the last two rows of $\mathbf{R}_{\text{behavior}}$ represent the behavioral measures for NC. The product the rows of $\mathbf{Y}$ and $\mathbf{U}$ corresponding to each experimental condition separately gives the latent variables for behavior (behavior scores) per condition. These condition-wise behavior scores are stacked one of top of the other to form the combined behavior scores. So, with the latent variable for the n-th condition computed as:

$$\mathbf{L}_{\mathbf{Y},n} = \mathbf{Y}_n \mathbf{U}_n, \quad (22)$$

we obtain

$$\mathbf{L_Y} = \begin{bmatrix} \mathbf{L}_{\mathbf{Y},1} \\ \mathbf{L}_{\mathbf{Y},2} \\ \mathbf{L}_{\mathbf{Y},3} \end{bmatrix} = \begin{bmatrix} -0.64 & 0.67 & -0.61 & 0.01 & 0.00 & 0.03 \\ 0.48 & -0.50 & 0.45 & -0.01 & 0.01 & 0.01 \\ 0.16 & -0.17 & 0.15 & 0.00 & -0.01 & -0.04 \\ -0.05 & 0.24 & 0.31 & -0.67 & 0.06 & 0.00 \\ 0.34 & -0.36 & -0.76 & 0.19 & -0.06 & 0.02 \\ -0.29 & 0.12 & 0.45 & 0.47 & 0.00 & -0.02 \\ 0.78 & 0.79 & 0.05 & 0.23 & -0.10 & 0.03 \\ -0.30 & -0.30 & -0.02 & -0.08 & 0.16 & -0.04 \\ -0.48 & -0.48 & -0.02 & -0.15 & -0.06 & 0.01 \end{bmatrix}, \quad (23)$$

The latent variables for brain activity (brain scores or $\mathbf{L_X}$) and behavior (behavior scores or $\mathbf{L_Y}$) are not typically illustrated. However PCA style plots show the effect of two latent variables at once. In our example, the PCA style plot of $\mathbf{L_X}$ (Fig. 6a) shows that the first brain score separates AD_1, PD_3 and NC_3 from the other participants, while the second brain score separates AD_2, PD_2, NC_2 and NC_3 from the other participants. The PCA style plot of $\mathbf{L_Y}$ (Fig. 6b) shows that the first behavior score separates NC_1 from NC_2 and NC_3, while the second behavior score separates AD_1, PD_2 and PD_3 from AD_2, AD_3 and PD_2.

Because of the small sample size, the bootstrap and permutation tests were not performed in this analysis and, hence, these results represent a fixed effect model (i.e., the interpretation is limited to the given data sets and cannot be generalized to the population).

3.5.1. Applications

Behavior PLSC has been used in several neuroimaging domains to examine the relationship between brain activity and behavior. For example, Price et al. (2004) used behavior PLSC to examine the relationship between voxel activity from PET scans of individuals with Alzheimer's disease and normal controls and descriptive and cognitive measures of age, white matter volume, and mental status on the Mini-Mental State Examination (MMSE; Folstein et al., 1975). The saliences for $\mathbf{Y}_{\text{behavior}}$ showed that tracer retention differences between patients with Alzheimer's disease and normal controls (i.e., brain activity differences between groups coded by contrasts) contributed most to the correlation. The MMSE scores had the second highest contribution.

Behavior PLSC has also been used (1) to identify grey matter systems correlated with response inhibition in obsessive compulsive disorder (Menzies et al., 2007), (2) to examine relationships between brain structures and neuropsychological tests in schizophrenia (Nestor et al., 2002), (3) to analyze the relationship between cerebral blood flow (as measured by SPECT) and visuo-spatial tasks in Alzheimer's patients (Tippett and Black, 2008), and, (4) to relate brain shape, neuropsychological behavior, and cortical functioning in traumatic brain injury patients (Bookstein et al., 2002; Fujiwara et al., 2008).

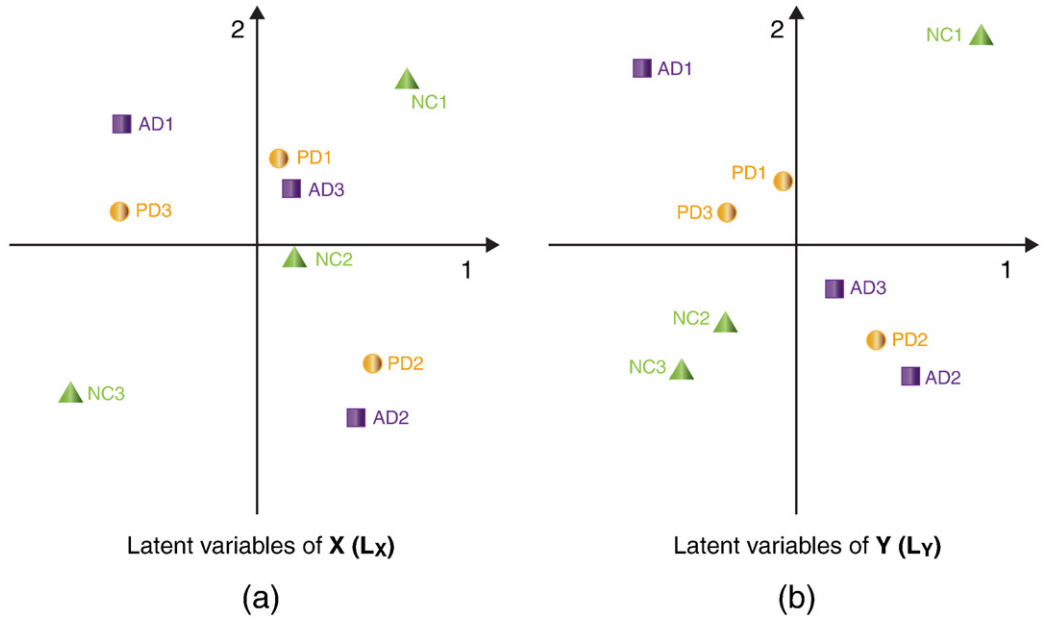

Fig. 6. Mini-example—Behavior PLSC: Brain and Behavior scores. (a) Plot of $\mathbf{L_X}$; latent variables 1 and 2 of $\mathbf{X}$ (Brain scores) and (b) Plot of $\mathbf{L_Y}$; latent variables 1 and 2 of $\mathbf{Y}$ (Behavior scores). The dimensions represent the latent variables, the horizontal (resp. vertical) coordinate of each point is given by its value for the first (resp. second) latent variable.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

3.6. Task PLSC

In contrast with behavior PLSC, task PLSC analyzes the relationship between brain activity and experimental design. There are two closely related types of task PLSC: (1) contrast, and (2) mean-centered. When differences in scaling are considered, the contrast and mean-centered approaches give almost the same results (McIntosh and Lobaugh, 2004). The mean-centered approach is currently the version of task PLSC used, and we present contrast task PLSC essentially for historical reasons and to be exhaustive.

3.6.1. Contrast task PLSC

Contrast task PLSC analyzes the relationship between brain measurements and a set of contrasts which represent specific experimental questions. As with behavior PLSC, when using contrast task PLSC, matrix **X** [Eq. (11)] contains voxel activity. However, instead of demographic or behavioral measures, matrix **Y** contains a set of *orthonormal* contrasts (i.e., these contrasts are pairwise orthogonal and the sum of squares of the coefficients of a given contrast is equal to one), which reflects the experimental question at hand. The method for contrast task PLSC is illustrated in Fig. 7.

The matrix **X** is obtained from the data described in Eq. (11). These data are then centered and normalized such that the mean of each column is zero and the sum of the squared elements of each column is equal to one. This gives:

$$\mathbf{X} = \begin{bmatrix} 0.31 & 0.14 & 0.02 & -0.41 & 0.33 & -0.42 & 0.15 & 0.20 & -0.27 & -0.35 & 0.19 & -0.33 \\ -0.07 & -0.32 & -0.19 & 0.51 & 0.23 & 0.34 & -0.42 & 0.47 & 0.34 & -0.02 & 0.30 & -0.49 \\ 0.05 & 0.48 & 0.24 & -0.14 & 0.13 & -0.42 & 0.15 & -0.08 & 0.19 & -0.35 & -0.17 & -0.33 \\ -0.19 & -0.09 & 0.24 & 0.25 & -0.46 & -0.42 & 0.50 & -0.35 & -0.27 & -0.35 & 0.19 & -0.17 \\ -0.31 & -0.09 & 0.46 & 0.38 & -0.46 & 0.21 & 0.38 & -0.48 & 0.64 & 0.41 & -0.53 & 0.14 \\ -0.44 & 0.37 & -0.63 & -0.41 & -0.46 & -0.17 & -0.54 & 0.47 & -0.42 & -0.13 & 0.42 & -0.02 \\ 0.55 & -0.43 & 0.24 & -0.41 & 0.23 & 0.34 & -0.19 & -0.35 & -0.12 & 0.19 & -0.41 & 0.30 \\ 0.42 & -0.43 & 0.02 & 0.12 & 0.33 & 0.34 & -0.19 & -0.08 & -0.27 & 0.63 & -0.29 & 0.45 \\ 0.30 & 0.37 & -0.41 & 0.12 & 0.13 & 0.21 & 0.15 & 0.20 & 0.19 & -0.02 & 0.30 & 0.45 \end{bmatrix}. \tag{24}$$

Because there are three experimental groups in our example, the maximum number of orthonormal contrasts is two (see e.g., Abdi et al., 2009b), and this matrix of contrasts, denoted $\mathbf{Y}_{\text{contrast}}$, is:

$$\mathbf{Y}_{\text{contrast}} = \begin{bmatrix} -0.24 & -0.41 \\ -0.24 & -0.41 \\ -0.24 & -0.41 \\ -0.24 & 0.41 \\ -0.24 & 0.41 \\ -0.24 & 0.41 \\ 0.47 & 0.00 \\ 0.47 & 0.00 \\ 0.47 & 0.00 \end{bmatrix}, \tag{25}$$

where the rows represent participants AD_1, AD_2, AD_3, PD_1, PD_2, PD_3, NC_1, NC_2 and NC_3, and the columns represent contrasts ψ_1 and ψ_2. Contrast ψ_1 predicts that the AD and PD participants will differ from the NC participants and contrast ψ_2 predicts that the AD participants will differ from the PD participants.

The matrix of correlations between **X** and $\mathbf{Y}_{\text{contrast}}$ (i.e., $\mathbf{R}_{\text{contrast}}$) is computed as:

$$\mathbf{R}_{\text{contrast}} = \mathbf{Y}_{\text{contrast}}^T \mathbf{X}. \tag{26}$$

For our example, the cross-product matrix is:

$$\mathbf{R}_{\text{contrast}} = \begin{bmatrix} 0.90 & -0.35 & -0.10 & -0.12 & 0.49 & 0.63 & -0.16 & -0.16 & -0.14 & 0.57 & -0.28 & 0.85 \\ -0.25 & -0.05 & 0.00 & 0.11 & -0.85 & 0.05 & 0.19 & -0.39 & -0.12 & 0.27 & -0.10 & 0.45 \end{bmatrix}, \tag{27}$$

which is decomposed by the SVD as [*cf.*, Eq. (1)]:

$$\mathbf{R}_{\text{contrast}} = \mathbf{U}\Delta\mathbf{V}^T. \tag{28}$$

with the left singular vectors

$$\mathbf{U} = \begin{bmatrix} -1.0000 & 0.0014 \\ -0.0014 & -1.0000 \end{bmatrix}, \tag{29}$$

the singular values

$$\Delta = \begin{bmatrix} 1.67 & 0 \\ 0 & 1.13 \end{bmatrix}, \tag{30}$$

and the right singular vectors

$$\mathbf{V} = \begin{bmatrix} -0.54 & 0.22 \\ 0.21 & 0.04 \\ 0.06 & 0.00 \\ 0.07 & -0.09 \\ -0.29 & 0.75 \\ -0.38 & -0.05 \\ 0.10 & -0.17 \\ 0.10 & 0.34 \\ 0.09 & 0.11 \\ -0.34 & -0.24 \\ 0.17 & 0.09 \\ -0.51 & -0.39 \end{bmatrix}. \tag{31}$$

The brain scores (latent variables for brain activity) are computed as [*cf.*, Eq. (4)]:

$$\mathbf{L}_{\mathbf{X}} = \mathbf{XV} = \begin{bmatrix} 0.56 & 0.48 \\ 0.14 & 0.57 \\ 0.48 & 0.33 \\ 0.64 & -0.46 \\ 0.01 & -0.82 \\ 0.52 & -0.11 \\ -0.94 & 0.00 \\ -1.08 & -0.08 \\ -0.34 & 0.08 \end{bmatrix}, \tag{32}$$

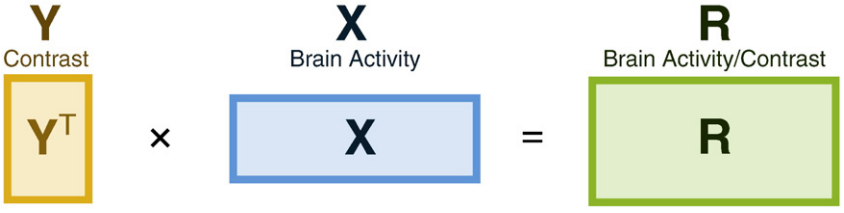

Fig. 7. Matrix **X** and matrix $\mathbf{Y}_{\text{contrast}}$ for Contrast approach to task PLSC: The matrix of correlations (**R**) between **X** and $\mathbf{Y}_{\text{contrast}}$ is decomposed by SVD.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

and the design scores (latent variables for the design) are computed as [*cf.*, Eq. (5)]:

$$\mathbf{L_Y} = \mathbf{YU} = \begin{bmatrix} 0.24 & 0.41 \\ 0.24 & 0.41 \\ 0.24 & 0.41 \\ 0.24 & -0.41 \\ 0.24 & -0.41 \\ 0.24 & -0.41 \\ -0.47 & 0.00 \\ -0.47 & 0.00 \\ -0.47 & 0.00 \end{bmatrix}. \tag{33}$$

For spatio-temporal data, the matrix $\mathbf{X}$ is made up of T matrices, each denoted $\mathbf{X}_t$. The product of each of these $\mathbf{X}_t$ matrices with the brain saliences gives a matrix of brain scores for a given time point, called the *temporal scores*, which depict differences in the time-course of brain activation.

Incidentally, contrast task PLSC is related to Principal Component Analysis (PCA) because, like PCA, contrast task PLSC boils down to computing the SVD of a correlation (or covariance) matrix. Therefore, the plots for contrast task PLSC are similar to PCA plots, except that for the contrast approach there are separate plots for the brain scores ($\mathbf{L_X}$) and for the design scores ($\mathbf{L_Y}$). Also, as in PCA, the number of latent variables is limited to the rank of $\mathbf{R}_{\text{contrast}}$. For our example, the rank of $\mathbf{R}_{\text{contrast}}$ is two and, hence, there are only two brain or design latent variables.

Even though, in general, PLSC results are not illustrated with PCA style graphs, such a representation can be informative because it can help integrate the effects of two latent variables at once. As an illustration, PCA style plots of the latent variables, $\mathbf{L_X}$ and $\mathbf{L_Y}$, for our example are shown in Fig. 8a and b, respectively and indicate that the three groups are separated in the space of the first two latent variables. Note that $\mathbf{L_X}$ is mean-centered and the first brain score is plotted against the second brain score. The first brain score separates NC from AD and PD. The second brain score separates AD from the PD group (Fig. 8a). Note that because $\mathbf{L_Y}$ is a product of a set of contrasts and brain activity, each observation in an experimental condition is weighted equally. Therefore a plot of each latent variable for the design (design score) shows a condition effect (i.e., the scores within a condition fall on the same point in the space). For our example, the first design score separates NC from AD and PD, and the second design score separates AD from PD (Fig. 8b).

3.6.2. Mean-centered task PLSC

Mean-centered task PLSC is used when the I observations are structured into N groups or conditions. When using mean-centered task PLSC, the observations in $\mathbf{X}$ are arranged according to these N experimental conditions and $\mathbf{Y}$ is a matrix of dummy coding that codes for the experimental groups or conditions. For our example, the matrix $\mathbf{X}$ is the data matrix described in Eq. (11) and $\mathbf{Y}_{\text{dummy coding}}$ is:

$$\mathbf{Y}_{\text{dummy coding}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \tag{34}$$

where the first column represents participants in the AD group, the second column represents participants in the PD group, and the third column represents participants in the NC group.

The average for each condition is then computed and stored in an $N \times J$ matrix, denoted $\mathbf{M}$, which is computed as:

$$\mathbf{M} = \text{diag}\left\{\mathbf{1}^T\mathbf{Y}_{\text{dummy coding}}\right\}^{-1}\mathbf{Y}_{\text{dummy coding}}^T\mathbf{X} \tag{35}$$

where $\mathbf{1}$ is a column vector of ones with the same number of rows as matrix $\mathbf{Y}_{\text{dummy coding}}$ and $\text{diag}\left\{\mathbf{1}^T\mathbf{Y}_{\text{dummy coding}}\right\}$ is the diagonal matrix of the sum of the columns of $\mathbf{Y}_{\text{dummy coding}}$. For our example, $\mathbf{M}$ is:

$$\mathbf{M} = \begin{bmatrix} 3.67 & 4.67 & 6.00 & 4.00 & 8.00 & 3.00 & 5.33 & 6.00 & 4.33 & 2.00 & 6.33 & 2.67 \\ 2.00 & 4.33 & 6.00 & 4.67 & 1.00 & 3.33 & 6.67 & 3.67 & 3.67 & 4.00 & 5.67 & 5.00 \\ 8.00 & 2.33 & 5.67 & 3.67 & 8.00 & 6.67 & 5.00 & 4.00 & 3.33 & 6.67 & 4.33 & 7.67 \end{bmatrix}. \tag{36}$$

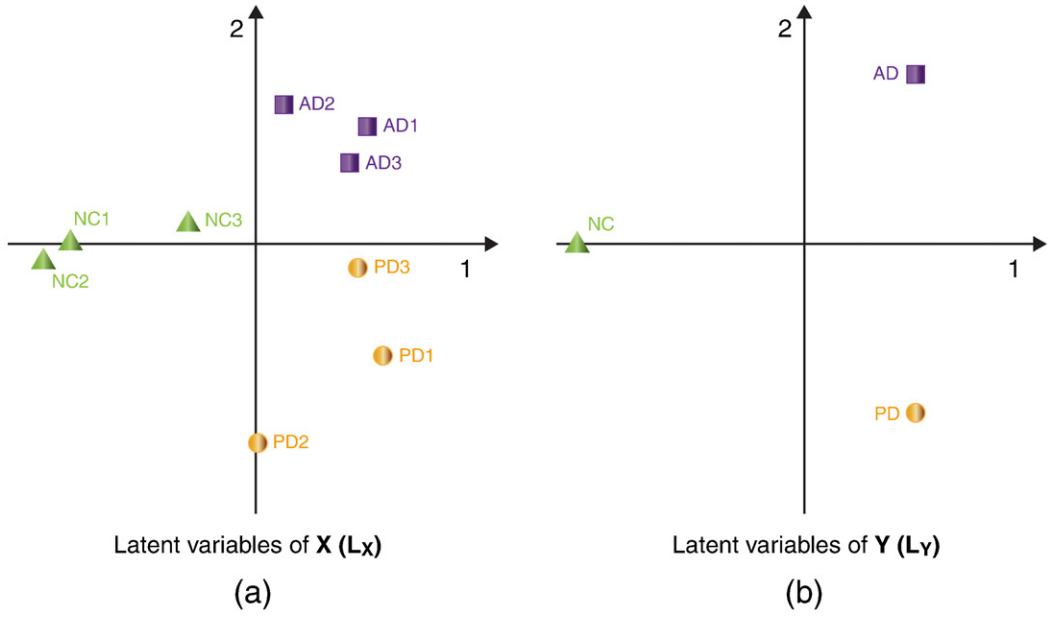

Fig. 8. Mini-example—Task PLSC: (a) Plot of $\mathbf{L_X}$; latent variables 1 and 2 of $\mathbf{X}$ (Brain scores) and (b) Plot of $\mathbf{L_Y}$; latent variables 1 and 2 of $\mathbf{Y}$ (Design scores). The dimensions represent the latent variables, the horizontal (resp. vertical) coordinate of each point is given by its value for the first (resp. second) latent variable.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

Each column of matrix **M** is then mean-centered by subtracting the mean of the column from each value of that column. This column-wise mean-centered matrix is denoted $\mathbf{R}_{\text{mean-centered}}$ (note that $\mathbf{R}_{\text{mean-centered}}$ is *not* a matrix of correlations). Formally, $\mathbf{R}_{\text{mean-centered}}$ is computed as:

$$\mathbf{R}_{\text{mean-centered}} = \mathbf{M} - \mathbf{1}\left[\frac{1}{N}\mathbf{1}^T\mathbf{M}\right] \tag{37}$$

where **1** is an N by 1 vector of ones. For our example, $\mathbf{R}_{\text{mean-centered}}$ is:

$$\mathbf{R}_{\text{mean-centered}} = \begin{bmatrix} -0.89 & 0.89 & 0.11 & -0.11 & 2.33 & -1.33 & -0.33 & 1.44 & 0.56 & -2.22 & 0.89 & -2.44 \\ -2.56 & 0.56 & 0.11 & 0.56 & -4.67 & -1.00 & 1.00 & -0.89 & -0.11 & -0.22 & 0.22 & -0.11 \\ 3.44 & -1.44 & -0.22 & -0.44 & 2.33 & 2.33 & -0.67 & -0.56 & -0.44 & 2.44 & -1.11 & 2.56 \end{bmatrix}. \tag{38}$$

The SVD [*cf.*, Eq. (1)] on $\mathbf{R}_{\text{mean-centered}}$ gives:

$$\mathbf{R}_{\text{mean-centered}} = \mathbf{U}\Delta\mathbf{V}^T \tag{39}$$

with left singular vectors

$$\mathbf{U} = \begin{bmatrix} 0.20 & 0.79 \\ 0.59 & -0.57 \\ -0.79 & -0.22 \end{bmatrix}, \tag{40}$$

singular values

$$\Delta = \begin{bmatrix} 7.86 & 0 \\ 0 & 5.73 \end{bmatrix}, \tag{41}$$

and right singular vectors

$$\mathbf{V} = \begin{bmatrix} -0.56 & 0.00 \\ 0.21 & 0.12 \\ 0.03 & 0.01 \\ 0.08 & -0.05 \\ -0.52 & 0.69 \\ -0.34 & -0.18 \\ 0.13 & -0.12 \\ 0.03 & 0.31 \\ 0.05 & 0.11 \\ -0.32 & -0.38 \\ 0.15 & 0.14 \\ -0.33 & -0.43 \end{bmatrix}. \tag{42}$$

Because mean-centered task PLSC differentiates between experimental groups or conditions, the saliences for the groups/conditions **U** have a pattern similar to the design latent variables ($\mathbf{L}_{\mathbf{Y}}$) in the contrast approach of task PLSC. The first salience differentiates the PD group from the NC group and the second salience differentiates the AD group from the NC and PD groups (Fig. 9). Note that in mean-centered task PLSC the group saliences optimally separate the groups. This makes mean-centered task PLSC akin to discriminant analysis. Specifically mean-centered task PLSC is equivalent to "barycentric discriminant analysis," because these two techniques compute the SVD of the matrix $\mathbf{R}_{\text{mean-centered}}$ (see, e.g., Abdi and Williams, 2010a; Williams et al., in press, for details).

3.6.3. Applications

3.6.3.1. Contrast task PLSC. Contrast task PLSC has been used to extract both the spatial and temporal features characteristic of *f*MRI data (Addis et al., 2004). For example, Keightley et al. (2003b) used contrast task PLSC to study the relationship between attention-demanding cognitive tasks and brain activity associated with emotional processing.

In addition, many PET studies have used contrast task PLSC to study the relationship between memory and different experimental conditions (e.g., Iidaka et al., 2000; Nyberg et al., 1996; Nyberg et al., 2002; Rajah and McIntosh, 2005; Köhler et al., 1998). For example, (Nyberg et al., 2002) used contrast task PLSC to study the similarities and differences in brain activity for working memory, semantic memory, and episodic memory. Their design matrix contained dummy coding for different tasks. Contrast task PLSC revealed spatial patterns of brain activity that showed optimal association with the experimental conditions.

3.6.3.2. Mean-centered task PLSC. Mean-centered task PLSC is now used instead of contrast task PLSC. McIntosh and Lobaugh (2004) demonstrated that the two approaches produce identical results save for differences in the singular values. The mean-centered approach has been used in several ERP studies (Bergström et al., 2007; West and Krompinger, 2005; West and Wymbs, 2004) and recent *f*MRI work (McIntosh et al., 2004; Martinez-Montes et al., 2004).

3.6.3.3. "Non-rotated" PLSC. A recent version of PLSC serves as a variation of the contrast task PLSC and mean-centered task PLSC. This version uses *a priori* (preferably orthogonal) contrasts, but, instead of using matrix **X**, it uses the matrix $\mathbf{R}_{\text{mean-centered}}$ [see Eq. (37)], and does *not* perform the SVD of $\mathbf{R}_{\text{mean-centered}}$ (hence the term "non-rotated" because the SVD can be interpreted as a rotation of an orthogonal basis). Instead the matrix obtained by the product of the contrast matrix with the $\mathbf{R}_{\text{mean-centered}}$ matrix is used directly for inferential analyses. This version acts as a complement to mean-centered task PLSC because it can test specific contrasts and therefore can be used to decompose interaction terms into interpretable dimensions (see e.g., Protzner and McIntosh, 2007; McIntosh et al., 2004).

3.7. Seed PLSC

Seed PLSC is used to analyze functional connectivity (McIntosh et al., 1997; McIntosh and Gonzalez-Lima, 1991). Although seed PLSC can analyze connectivity over the whole brain, this method has been commonly used to determine the connectivity between one (or a few) region(s) of interest (ROI) and the rest of the brain.

The first step is to determine a *seed*, which is a representative voxel or set of voxels for a particular ROI. Ideally, a seed is chosen for theoretical reasons, but voxel seeds are also selected using behavior or task PLSC, by univariate multiple regression, or based on experimental hypotheses. One seed is selected for every ROI for each experimental condition. The activity of these seed voxels are stored in $\mathbf{Y}_{\text{seed}}$ (McIntosh et al., 1997). Because $\mathbf{Y}_{\text{seed}}$ already accounts for seed activity, the seed voxels are removed from **X**.

In our example (for simplicity), we have chosen the first column and last column of **X** to be the seeds [*cf.*, Eq. (11)]. Therefore, $\mathbf{Y}_{\text{seed}}$ is represented as:

$$\mathbf{Y}_{\text{seed}} = \begin{bmatrix} 2 & 3 \\ 4 & 2 \\ 5 & 3 \\ 3 & 4 \\ 2 & 6 \\ 1 & 5 \\ 9 & 7 \\ 8 & 8 \\ 7 & 8 \end{bmatrix}, \tag{43}$$

where the rows represent the participants from matrix **X** and the columns represent the seed voxels. Because we have used 2 vectors of

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

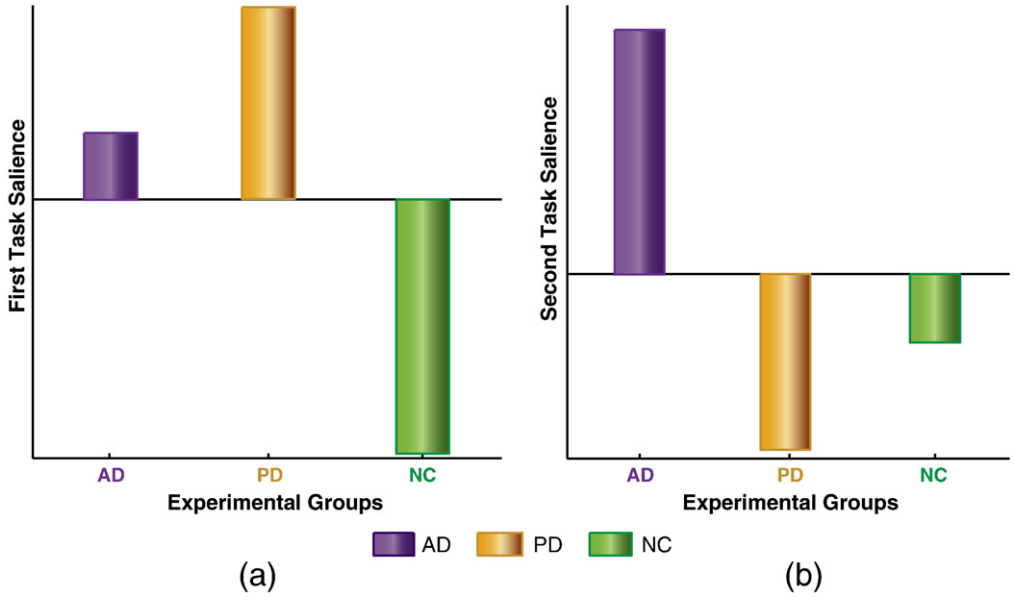

Fig. 9. Mini-example—Mean-centered task PLSC: Plot of saliences 1 and 2 of $\mathbf{R}_{\text{mean-centered}}$.

$\mathbf{X}$ as seeds, $\mathbf{X}_{\text{seed}}$ now has I rows and $J-2$ columns [compare with $\mathbf{X}$ in Eq. (11)]:

$$\mathbf{X}_{\text{seed}} = \begin{bmatrix} 5 & 6 & 1 & 9 & 1 & 7 & 6 & 2 & 1 & 7 \\ 1 & 5 & 8 & 8 & 7 & 2 & 8 & 6 & 4 & 8 \\ 8 & 7 & 3 & 7 & 1 & 7 & 4 & 5 & 1 & 4 \\ 3 & 7 & 6 & 1 & 1 & 10 & 2 & 2 & 1 & 7 \\ 3 & 8 & 7 & 1 & 6 & 9 & 1 & 8 & 8 & 1 \\ 7 & 3 & 1 & 1 & 3 & 1 & 8 & 1 & 3 & 9 \\ 0 & 7 & 1 & 8 & 7 & 4 & 2 & 3 & 6 & 2 \\ 0 & 6 & 5 & 9 & 7 & 4 & 4 & 2 & 10 & 3 \\ 7 & 4 & 5 & 7 & 6 & 7 & 6 & 5 & 4 & 8 \end{bmatrix}. \tag{44}$$

As with behavior PLSC, both $\mathbf{X}_{\text{seed}}$ and $\mathbf{Y}_{\text{seed}}$ are centered and normalized within each condition n. This gives the following values for $\mathbf{X}$:

$$\mathbf{X}_{\text{seed}} = \begin{bmatrix} 0.07 & 0.00 & -0.59 & 0.71 & -0.41 & 0.41 & 0.00 & -0.79 & -0.41 & 0.23 \\ -0.74 & -0.71 & 0.78 & 0.00 & 0.82 & -0.82 & 0.71 & 0.57 & 0.82 & 0.57 \\ 0.67 & 0.71 & -0.20 & -0.71 & -0.41 & 0.41 & -0.71 & 0.23 & -0.41 & -0.79 \\ -0.41 & 0.27 & 0.29 & 0.00 & -0.66 & 0.48 & -0.31 & -0.31 & -0.59 & 0.23 \\ -0.41 & 0.53 & 0.51 & 0.00 & 0.75 & 0.33 & -0.50 & 0.81 & 0.78 & -0.79 \\ 0.82 & -0.80 & -0.81 & 0.00 & -0.09 & -0.81 & 0.81 & -0.50 & -0.20 & 0.57 \\ -0.41 & 0.62 & -0.82 & 0.00 & 0.41 & -0.41 & -0.71 & -0.15 & -0.15 & -0.51 \\ -0.41 & 0.15 & 0.41 & 0.71 & 0.41 & -0.41 & 0.00 & -0.62 & 0.77 & -0.29 \\ 0.82 & -0.77 & 0.41 & -0.71 & -0.82 & 0.82 & 0.71 & 0.77 & -0.62 & 0.81 \end{bmatrix} \tag{45}$$

and $\mathbf{Y}$:

$$\mathbf{Y}_{\text{seed}} = \begin{bmatrix} -0.77 & 0.41 \\ 0.15 & -0.82 \\ 0.62 & 0.41 \\ 0.71 & -0.71 \\ 0.00 & 0.71 \\ -0.71 & 0.00 \\ 0.71 & -0.82 \\ 0.00 & 0.41 \\ -0.71 & 0.41 \end{bmatrix}. \tag{46}$$

The matrix of correlations for each condition n is then computed as [*cf.*, Eq. (15)]:

$$\mathbf{R}_{\text{seed,n}} = \mathbf{Y}_{\text{seed,n}}^{\text{T}} \mathbf{X}_{\text{seed,n}}. \tag{47}$$

All the condition-wise matrices of correlations are stacked one on top of the other to form the combined matrix of correlations $\mathbf{R}_{\text{seed}}$, which is the input for the SVD. The matrix $\mathbf{R}_{\text{seed}}$ contains the correlation of each of the $J-2$ voxels in $\mathbf{X}_{\text{seed}}$ with each of the K seed voxels in $\mathbf{Y}_{\text{seed}}$ within each of the N conditions. Therefore $\mathbf{R}_{\text{seed}}$ will have $N \times K$ rows and $J-2$ columns (Fig. 10). For our example, the $\mathbf{R}_{\text{seed}}$ matrix of correlations is:

$$\mathbf{R}_{\text{seed}} = \begin{bmatrix} \mathbf{R}_{\text{seed,1}} \\ \mathbf{R}_{\text{seed,2}} \\ \mathbf{R}_{\text{seed,3}} \end{bmatrix}$$

$$= \begin{bmatrix} 0.25 & 0.33 & 0.45 & -0.98 & 0.19 & -0.19 & -0.33 & 0.84 & 0.19 & -0.58 \\ 0.90 & 0.87 & -0.96 & 0.00 & -1.00 & 1.00 & -0.87 & -0.69 & -1.00 & -0.69 \\ -0.87 & 0.76 & 0.78 & 0.00 & -0.40 & 0.91 & -0.79 & 0.13 & -0.28 & -0.24 \\ 0.00 & 0.19 & 0.16 & 0.00 & 0.99 & -0.10 & -0.13 & 0.79 & 0.97 & -0.72 \\ -0.87 & 0.98 & -0.87 & 0.50 & 0.87 & -0.87 & -1.00 & -0.65 & 0.33 & -0.93 \\ 0.50 & -0.76 & 1.00 & 0.00 & -0.50 & 0.50 & 0.87 & 0.19 & 0.19 & 0.63 \end{bmatrix}, \tag{48}$$

where the rows are the seed voxels per condition n and the columns represent the voxel activity.

The remaining steps to compute seed PLSC are identical to those of behavior PLSC. However, there is an important difference in interpretation. The relationship between the seed voxels (i.e., $\mathbf{Y}_{\text{seed}}$) and the voxels in $\mathbf{X}_{\text{seed}}$ represents their functional connectivity.

Specifically, for our example the SVD [*cf.*, Eq. (1)] of $\mathbf{R}_{\text{seed}}$ is computed as:

$$\mathbf{R}_{\text{seed}} = \mathbf{U}\Delta\mathbf{V}^{T}, \tag{49}$$

which gives:

$$\mathbf{U} = \begin{bmatrix} -0.03 & -0.17 & 0.60 & 0.47 & 0.55 & -0.29 \\ -0.42 & 0.76 & 0.08 & 0.40 & -0.26 & -0.08 \\ -0.17 & 0.18 & 0.67 & -0.69 & -0.05 & 0.08 \\ -0.10 & -0.48 & 0.35 & 0.32 & -0.70 & 0.21 \\ -0.70 & -0.34 & -0.21 & -0.19 & -0.04 & -0.56 \\ 0.54 & 0.12 & 0.09 & -0.07 & -0.37 & -0.74 \end{bmatrix}, \tag{50}$$

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

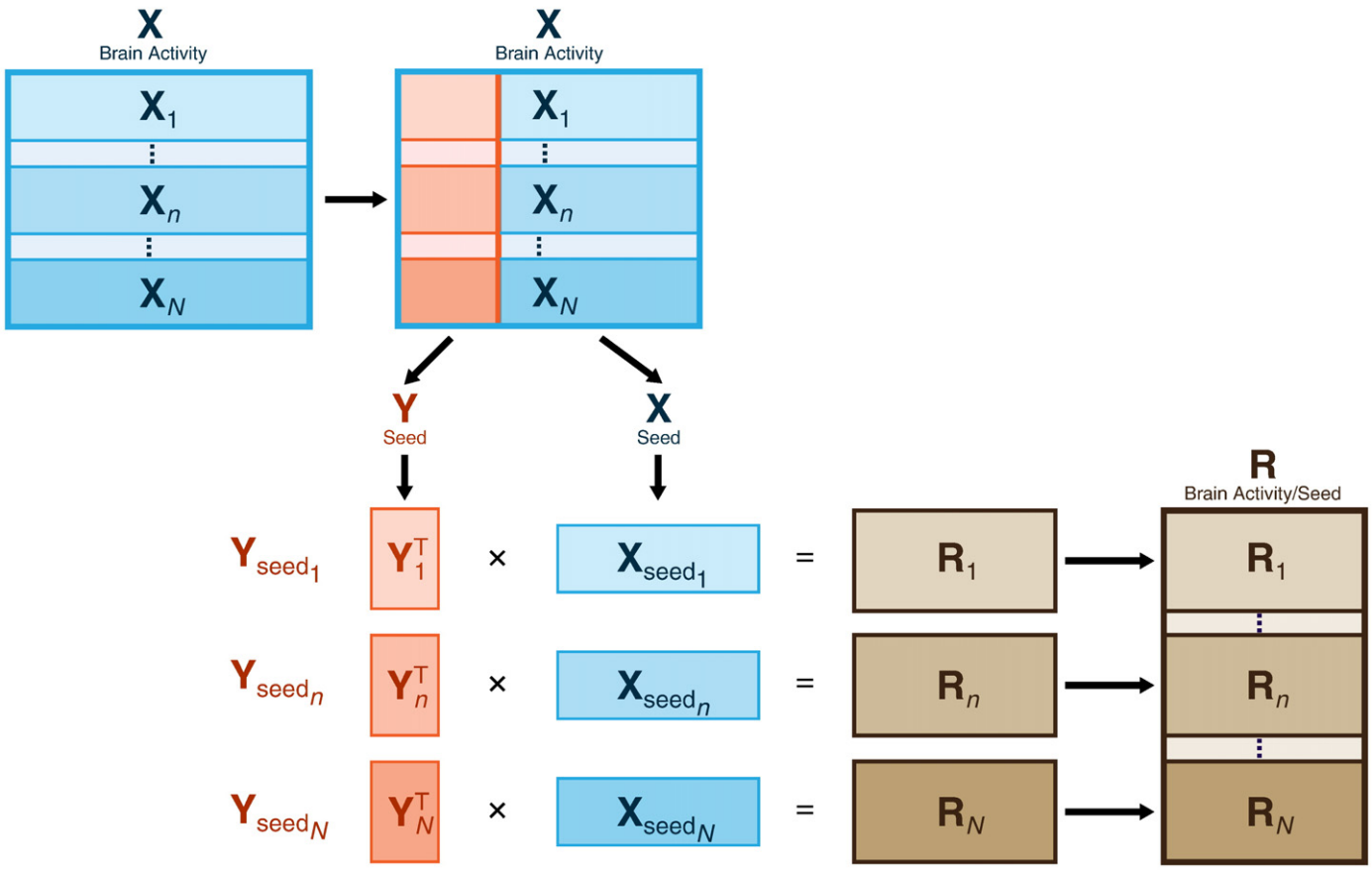

Fig. 10. Matrix **X** and matrix **Y** for Seed PLSC: The *I* observations are arranged according to *N* conditions in both matrices and are normalized within condition. The matrix of correlations ($\mathbf{R}_n$) between each condition-wise sub-matrix ($\mathbf{X}_n$ and $\mathbf{Y}_n$) are stacked one below the other to form a combined matrix of correlations **R** which is then decomposed by SVD.

where **U** are the saliences for the seed voxels,

$$\mathbf{\Delta} = \begin{bmatrix} 3.29 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.88 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.03 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.60 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 \end{bmatrix}, \tag{51}$$

where **Δ** is the diagonal matrix of singular values, and

$$\mathbf{V} = \begin{bmatrix} 0.20 & 0.29 & -0.06 & 0.75 & -0.23 & -0.26 \\ -0.49 & 0.08 & 0.28 & -0.05 & 0.02 & -0.14 \\ 0.42 & -0.12 & 0.52 & -0.35 & 0.02 & -0.53 \\ -0.10 & 0.00 & -0.34 & -0.35 & -0.61 & 0.01 \\ -0.15 & -0.59 & -0.06 & 0.09 & -0.18 & 0.24 \\ 0.10 & 0.47 & 0.38 & -0.14 & -0.55 & 0.35 \\ 0.51 & -0.09 & -0.27 & 0.08 & -0.11 & -0.02 \\ 0.22 & -0.27 & 0.48 & 0.24 & 0.04 & 0.56 \\ 0.07 & -0.49 & 0.06 & 0.07 & -0.42 & -0.27 \\ 0.43 & 0.09 & -0.28 & -0.30 & 0.21 & 0.27 \end{bmatrix}, \tag{52}$$

where **V** are the saliences for the brain activity. The relationship between the seed voxels (i.e., $\mathbf{Y}_{\text{seed}}$) and the voxels in $\mathbf{X}_{\text{seed}}$ represents their functional connectivity. This pattern of connectivity can be illustrated by plotting the saliences for the brain (**V**) into a glass brain to show how strongly the seed voxels correlate with the rest of the brain.

The $N \times K$ rows and L columns (with L being the rank of matrix $\mathbf{R}_{\text{seed}}$) matrix of the seed saliences, (i.e., **U**), indicate differences in the seed voxels across experimental conditions. A separate bar plot of each column of **U** against each seed voxel shows how the experimental conditions interact with the seed voxels. Fig. 11a shows the first seed salience where the first seed voxel is important for the NC group and the second seed voxel differentiates the AD group from the NC group. Fig. 11b shows the second seed salience where the first seed voxel differentiates the PD group from the other groups and the second seed voxel differentiates the AD group from the PD group.

3.7.1. Applications

Several PET studies have used seed PLSC to explore functional connectivity (e.g., see Della-Maggiore et al., 2000; Grady et al., 2003; Keightley et al., 2003a; Nyberg et al., 2000). For example, Keightley et al. (2003a) used seed PLSC to determine networks of brain regions active during neutral and sad internally generated mood states in individuals with positive and negative affect styles. They first used contrast task PLSC to identify the subgenual cingulate as an ROI, a voxel of which was selected as the seed. The **Y** matrix consisted of activity in this voxel across all time points. Via seed PLSC, Keightley et al. (2003a) found a divergent subgenual cingulate-mediated network that differentiated participants based on affect style.

3.8. Multi-Table PLSC

Multi-table PLSC, also known as multi-block PLSC, extends behavior, contrast, and seed PLSC to analyze three or more matrices simultaneously. This means that for every matrix **X** of brain activity, there are two or more **Y** matrices (Fig. 12). The **X** matrix is designated as *causal* and the **Y** matrices are designated as *related* data sets. Multi-table PLSC analyses the pattern(s) of correlation between causal and related data sets (Streissguth et al., 1993). That is, multi-table PLSC identifies the distributed patterns of brain activity and measures

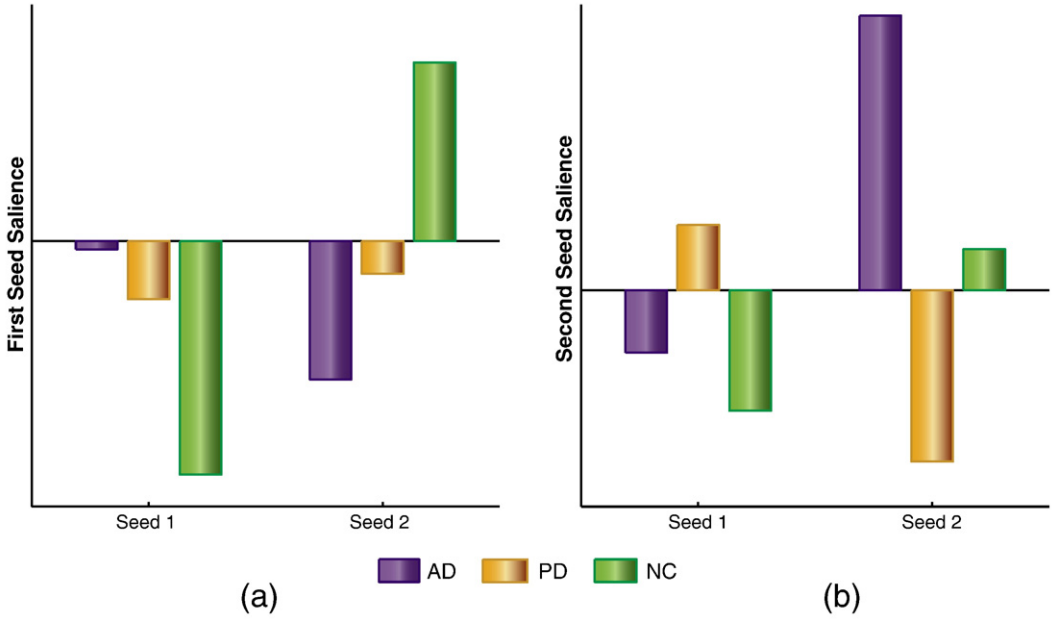

Fig. 11. Mini-example—Seed PLSC: Plot of saliences 1 and 2 of $\mathbf{R}_{seed}$.

related changes in the behavioral measures, contrasts, and/or seed voxel activity across experimental conditions (Caplan et al., 2007).

In our example, $\mathbf{X}$ contains brain activity [as given in Eq. (24)], $\mathbf{Y}_1$ contains the contrasts from Eq. (25) and $\mathbf{Y}_2$ contains the seed voxel activity from Eq. (43). The matrix of correlations $\mathbf{R}_1$ is computed between brain activity and the contrasts as $\mathbf{Y}_1^T\mathbf{X}$ as described in Eq. (27). This gives

$$\mathbf{R}_1 = \begin{bmatrix} 0.90 & -0.35 & -0.10 & -0.12 & 0.49 & 0.63 & -0.16 & -0.16 & -0.14 & 0.57 & -0.28 & 0.85 \\ -0.25 & -0.05 & 0.00 & 0.11 & -0.85 & 0.05 & 0.19 & -0.39 & -0.12 & 0.27 & -0.10 & 0.45 \end{bmatrix}. \tag{53}$$

A second matrix of correlations, $\mathbf{R}_2$, is computed between brain activity and seed voxel activity as $\mathbf{Y}_2^T\mathbf{X}$ as described in Eq. (48). The resulting $\mathbf{R}_2$ matrix has $N \times K$ rows and J columns. Recall that in seed PLSC the seed voxels are removed from $\mathbf{X}$. However in multi-table PLSC, the seed voxels are *not* removed so that $\mathbf{R}_1$ and $\mathbf{R}_2$ have the same number of columns (i.e., $J_1 = J_2$). For the example, $\mathbf{R}_2$ is:

$$\mathbf{R}_2 = \begin{bmatrix} 1.00 & 0.25 & 0.33 & 0.45 & -0.98 & 0.19 & -0.19 & -0.33 & 0.84 & 0.19 & -0.58 & -0.19 \\ -0.19 & 0.90 & 0.87 & -0.96 & 0.00 & -1.00 & 1.00 & -0.87 & -0.69 & -1.00 & -0.69 & 1.00 \\ 1.00 & -0.87 & 0.76 & 0.78 & 0.00 & -0.40 & 0.91 & -0.79 & 0.13 & -0.28 & -0.24 & -0.50 \\ -0.50 & 0.00 & 0.19 & 0.16 & 0.00 & 0.99 & -0.10 & -0.13 & 0.79 & 0.97 & -0.72 & 1.00 \\ 1.00 & -0.87 & 0.98 & -0.87 & 0.50 & 0.87 & -0.87 & -1.00 & -0.65 & 0.33 & -0.93 & -0.87 \\ -0.87 & 0.50 & -0.76 & 1.00 & 0.00 & -0.50 & 0.50 & 0.87 & 0.19 & 0.19 & 0.63 & 1.00 \end{bmatrix}, \tag{54}$$

$\mathbf{R}_1$ and $\mathbf{R}_2$ are concatenated by column to form matrix $\mathbf{R}_{multi}$:

$$\mathbf{R}_{multi} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix} \tag{55}$$

$$= \begin{bmatrix} 0.90 & -0.35 & -0.10 & -0.12 & 0.49 & 0.63 & -0.16 & -0.16 & -0.14 & 0.57 & -0.28 & 0.85 \\ -0.25 & -0.05 & 0.00 & 0.11 & -0.85 & 0.05 & 0.19 & -0.39 & -0.12 & 0.27 & -0.10 & 0.45 \\ 1.00 & 0.25 & 0.33 & 0.45 & -0.98 & 0.19 & -0.19 & -0.33 & 0.84 & 0.19 & -0.58 & -0.19 \\ -0.19 & 0.90 & 0.87 & -0.96 & 0.00 & -1.00 & 1.00 & -0.87 & -0.69 & -1.00 & -0.69 & 1.00 \\ 1.00 & -0.87 & 0.76 & 0.78 & 0.00 & -0.40 & 0.91 & -0.79 & 0.13 & -0.28 & -0.24 & -0.50 \\ -0.50 & 0.00 & 0.19 & 0.16 & 0.00 & 0.99 & -0.10 & -0.13 & 0.79 & 0.97 & -0.72 & 1.00 \\ 1.00 & -0.87 & 0.98 & -0.87 & 0.50 & 0.87 & -0.87 & -1.00 & -0.65 & 0.33 & -0.93 & -0.87 \\ -0.87 & 0.50 & -0.76 & 1.00 & 0.00 & -0.50 & 0.50 & 0.87 & 0.19 & 0.19 & 0.63 & 1.00 \end{bmatrix}. \tag{56}$$

The SVD of $\mathbf{R}_{multi}$ gives:

$$\mathbf{R}_{multi} = \mathbf{U}\Delta\mathbf{V}^T, \tag{57}$$

where $\mathbf{U}$ is the set of saliences for both the contrasts and the seed voxel activity and $\mathbf{V}$ is the set of saliences for the voxels. Consequently $\mathbf{U}$ has both a task or design portion and a seed portion. The task or design portion of $\mathbf{U}$ indicates how the pattern of brain activity varies across conditions. The seed portion of $\mathbf{U}$ indicates how the pattern of brain activity varies with the seed voxel activity across experimental conditions (Caplan et al., 2007).

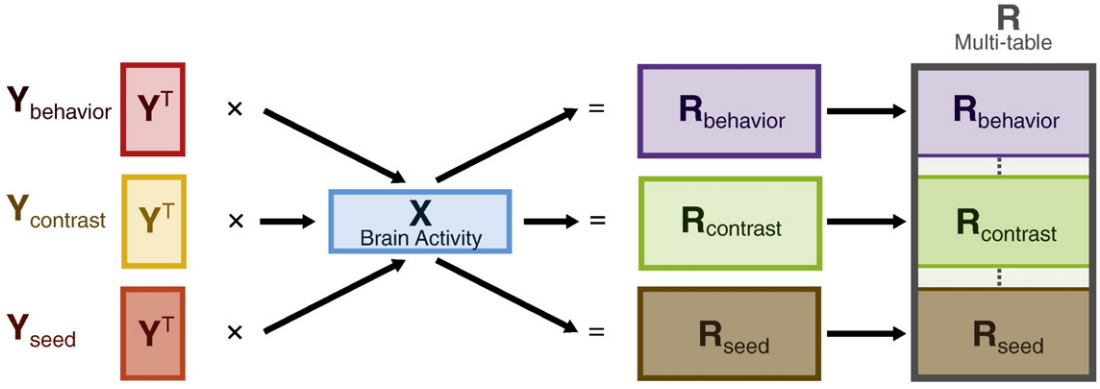

Fig. 12. Matrix $\mathbf{X}$, $\mathbf{Y}_1$ to $\mathbf{Y}_M$ for multi-table PLSC: The condition wise matrix of correlations between $\mathbf{X}$ and each of the $\mathbf{Y}_m$ matrices are stacked one below the other to form a combined matrix of correlations $\mathbf{R}$ which is then decomposed by SVD.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

In our example, the first two rows of **U** are denoted $\mathbf{U}_1$ and they correspond to the contrast matrix. The last six rows of **U** are denoted $\mathbf{U}_2$ and they correspond to the 2 seed voxels for each of the 3 experimental groups:

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} -0.17 & 0.18 & 0.37 & 0.01 & 0.57 & 0.63 & 0.20 & -0.20 \\ 0.04 & -0.04 & 0.22 & -0.22 & -0.18 & -0.25 & 0.90 & -0.07 \\ -0.19 & 0.13 & 0.08 & -0.63 & -0.53 & 0.45 & -0.12 & 0.22 \\ -0.01 & -0.90 & 0.41 & 0.01 & -0.03 & 0.09 & -0.11 & 0.07 \\ -0.29 & -0.19 & -0.31 & -0.65 & 0.52 & -0.30 & -0.06 & -0.07 \\ 0.01 & 0.32 & 0.73 & -0.17 & 0.02 & -0.46 & -0.34 & -0.11 \\ -0.73 & 0.07 & 0.08 & 0.29 & 0.04 & -0.15 & 0.09 & 0.59 \\ 0.57 & 0.07 & 0.07 & -0.15 & 0.32 & 0.03 & 0.05 & 0.73 \end{bmatrix}. \quad (58)$$

The latent variable for the contrasts are computed by:

$$\mathbf{L}_{\mathbf{Y}_1} = \mathbf{Y}_1\mathbf{U}_1. \quad (59)$$

For our example, this gives

$$\mathbf{L}_{\mathbf{Y}_1} = \begin{bmatrix} 0.02 & -0.03 & -0.18 & 0.09 & -0.06 & -0.05 & -0.41 & 0.07 \\ 0.02 & -0.03 & -0.18 & 0.09 & -0.06 & -0.05 & -0.41 & 0.07 \\ 0.02 & -0.03 & -0.18 & 0.09 & -0.06 & -0.05 & -0.41 & 0.07 \\ 0.06 & -0.06 & 0.00 & -0.09 & -0.21 & -0.25 & 0.32 & 0.02 \\ 0.06 & -0.06 & 0.00 & -0.09 & -0.21 & -0.25 & 0.32 & 0.02 \\ 0.06 & -0.06 & 0.00 & -0.09 & -0.21 & -0.25 & 0.32 & 0.02 \\ -0.08 & 0.08 & 0.17 & 0.00 & 0.27 & 0.30 & 0.09 & -0.09 \\ -0.08 & 0.08 & 0.17 & 0.00 & 0.27 & 0.30 & 0.09 & -0.09 \\ -0.08 & 0.08 & 0.17 & 0.00 & 0.27 & 0.30 & 0.09 & -0.09 \end{bmatrix}. \quad (60)$$

Note that, as a consequence of contrast coding (which treats all observations of a group similarly), the rows of a given group have the same value. Therefore the plot of the latent variables for the contrasts will show only the three group values (Fig. 13a) and shows that — as indicated by the first contrast — the first latent variable separates the controls group from the two clinical groups but that, among the clinical groups, the PD group shows the largest difference with the NC group.

Latent variables could be similarly computed for the seed portion of matrix **U** (i.e., $\mathbf{U}_2$). However, instead when dealing with connectivity data, saliences are, in general, plotted directly as shown in Fig. 13b. In this figure, the plot of the saliences for the first seed [with values of −0.19, −0.29 and −0.73, see Eq. (58)] shows that seed 1 is important for all experimental conditions, although to a varying degree. The plot of the saliences for the second seed (with values of −0.01, 0.01 and 0.57) shows that, in contrast, seed 2 is only relevant for the NC group and could be an indicator of differences between NC and the clinical populations.

Matrix **V** contains the saliences for brain activity across experimental conditions and seed voxels. For our example, **V** is:

$$\mathbf{V} = \begin{bmatrix} -0.48 & 0.04 & -0.16 & -0.32 & 0.19 & 0.76 & 0.02 & -0.05 \\ 0.30 & -0.23 & 0.20 & 0.04 & -0.43 & 0.36 & -0.29 & 0.28 \\ -0.37 & -0.27 & 0.12 & -0.15 & -0.04 & -0.29 & -0.25 & 0.28 \\ 0.24 & 0.27 & -0.21 & -0.55 & 0.26 & -0.13 & -0.01 & 0.53 \\ -0.08 & 0.01 & -0.02 & 0.42 & 0.62 & 0.01 & -0.58 & 0.10 \\ -0.24 & 0.48 & 0.31 & 0.12 & -0.02 & -0.10 & 0.00 & -0.25 \\ 0.18 & -0.40 & -0.01 & -0.36 & 0.35 & -0.21 & -0.03 & -0.37 \\ 0.40 & 0.27 & -0.17 & 0.18 & 0.00 & 0.22 & -0.24 & 0.10 \\ 0.11 & 0.30 & 0.08 & -0.42 & -0.23 & -0.04 & -0.58 & -0.44 \\ -0.04 & 0.47 & 0.30 & -0.04 & 0.10 & -0.11 & 0.20 & 0.22 \\ 0.33 & 0.10 & -0.39 & 0.13 & 0.14 & 0.07 & 0.24 & -0.28 \\ 0.32 & -0.12 & 0.71 & -0.09 & 0.34 & 0.26 & 0.17 & -0.04 \end{bmatrix}. \quad (61)$$

Each salience in **V** shows a distributed pattern of brain activity across conditions (this pattern can be visualized on a "glass brain"). This gives an indication of the interplay of the experimental condition and the functional connectivity between the ROIs represented by the seed voxels.

3.8.1. Applications

In neuroimaging, multi-table PLSC is often used when one of the data sets consists of seed voxel activity. For example, Vallesi et al. (2009) used multi-table PLSC in an *f*MRI experiment to find a functionally connected network of brain regions related to withholding a response. With *f*MRI activity as the *causal* data set (i.e., **X**), the authors used experimental design and seed voxel activity as the two *related* data sets. A multi-table PLSC on the combined matrix of correlations showed a distributed network related to the experimental groups and their functional connectivity with the regions represented by the seed voxels (Vallesi et al., 2009).

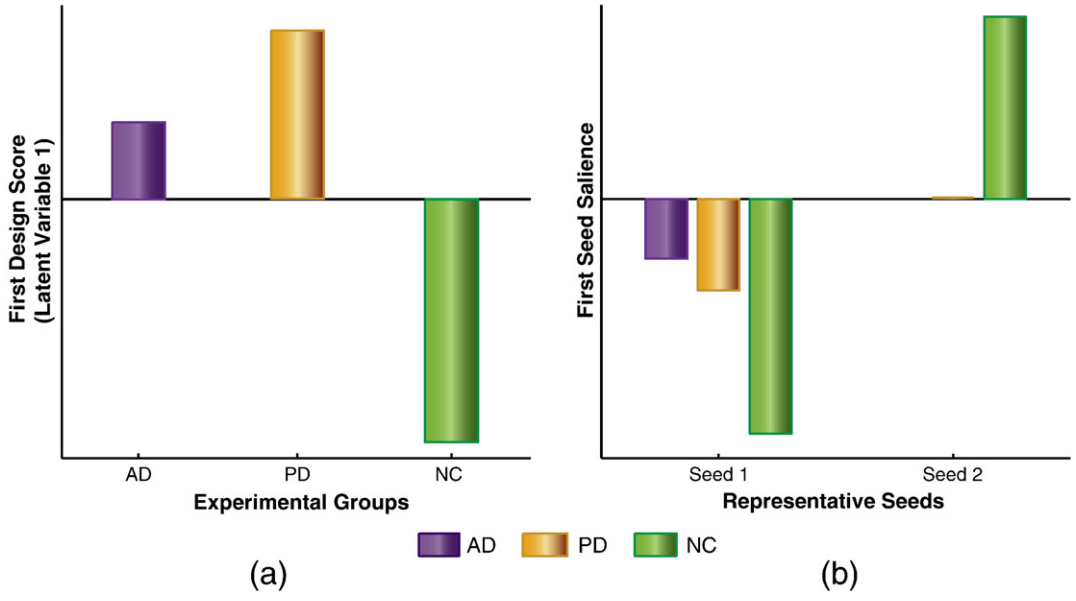

Fig. 13. Mini-example—Multi-table PLSC: (a) First design score of the task portion of multi-table PLSC for AD, PD, and NC groups; and (b) First salience of the seed portion of multi-table PLSC for the AD, PD, and NC groups.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

In another study, Caplan et al. (2007) used multi-table PLSC to identify the functional networks involved in resolving proactive interference. The causal matrix was brain activity measured by *f*MRI. The first related matrix was the experimental design variables and the second related matrix consisted of a combination of behavioral measures (reaction times) and seed voxel activity. The multi-table PLSC generated voxel saliences (i.e., **V**-s) that indicated the degree to which each voxel was related to the experimental design, behavior variables, and seed regions. Multi-table PLSC also generated (1) design latent variables showing changes in brain activity across conditions, and (2) seed saliences characterizing functional connectivity of the brain with the seed voxels (Caplan et al., 2007).

4. Partial Least Squares Regression (PLSR)

So far, we have seen that PLSC correlates brain and behavior (or design) variables by determining brain activity that co-occurs with different experimental conditions. Because PLSC is correlational, it cannot explicitly predict a set of variables from another set. By contrast, if we want to *predict* behavior (or some features of the experimental design) from brain activity, we use Partial Least Squares *Regression* (PLSR; Wold, 1982; Wold et al., 2001, see also Sidtis et al., 2003 for a discussion). The predictors (i.e., the *independent* variables) are stored in **X** and the predicted variables (i.e., the *dependent* variables) are stored in **Y**.

PLSR finds latent variables stored in a matrix **T** that model **X** and simultaneously predict **Y** (i.e., **T** plays a role similar to $\mathbf{L_X}$). Formally this is expressed as a double decomposition of **X** and the predicted $\widehat{\mathbf{Y}}$:

$$\mathbf{X} = \mathbf{TP}^T \quad \text{and} \quad \widehat{\mathbf{Y}} = \mathbf{TBC}^T, \tag{62}$$

where **P** and **C** are loadings (or weights) and **B** is a diagonal matrix. These latent variables are ordered according to the amount of variance of $\widehat{\mathbf{Y}}$ that they explain. Rewriting Eq. (62) shows that $\widehat{\mathbf{Y}}$ can also be expressed as a regression model as:

$$\widehat{\mathbf{Y}} = \mathbf{TBC}^T = \mathbf{XB}_{\mathrm{PLS}} \tag{63}$$

with

$$\mathbf{B}_{\mathrm{PLS}} = \mathbf{P}^{T+}\,\mathbf{BC}^T, \tag{64}$$

(where $\mathbf{P}^{T+}$ is the Moore-Penrose pseudo-inverse of $\mathbf{P}^T$, see, e.g., Abdi and Williams, 2010c, for definitions). The matrix $\mathbf{B}_{\mathrm{PLS}}$ has *J* rows and *K* columns and is equivalent to the regression weights of multiple regression (Note that matrix **B** is diagonal, but that matrix $\mathbf{B}_{\mathrm{PLS}}$ is in general *not* diagonal).

In neuroimaging, each latent variable of **X** describes a unique aspect of the overall variance of brain activity that best predicts behavior or experimental design. Similarly, each latent variable of **Y** describes a unique aspect of the variance of behavior or experimental design that is best predicted by **X** (Fig. 14). In our example, **X** contains brain activity from Eq. (11) and **Y** contains behavioral measures from Eq. (12).

4.1. Iterative computation of the latent variables in PLSR

In PLSR, the latent variables are computed by iterative applications of the SVD. Each run of the SVD produces orthogonal latent variables for **X** and **Y** and corresponding regression weights (see, e.g., Abdi, 2010, for more details and alternative algorithms).

4.1.1. Step one

X and **Y** are mean-centered and normalized (or transformed into *Z*-scores) and stored in matrices $\mathbf{X}_0$ and $\mathbf{Y}_0$. The matrix of correlations (or covariance) between $\mathbf{X}_0$ and $\mathbf{Y}_0$ is computed as:

$$\mathbf{R}_1 = \mathbf{X}_0^T\mathbf{Y}_0. \tag{65}$$

The SVD is then performed on $\mathbf{R}_1$ and produces two sets of orthogonal singular vectors $\mathbf{W}_1$ and $\mathbf{C}_1$, and the corresponding singular values Δ_1 [compare with Eq. (1)]:

$$\mathbf{R}_1 = \mathbf{W}_1\Delta_1\mathbf{C}_1^T. \tag{66}$$

The first pair of singular vectors (i.e., the first columns of $\mathbf{W}_1$ and $\mathbf{C}_1$) are denoted $\mathbf{w}_1$ and $\mathbf{c}_1$ and the first singular value (i.e., the first diagonal entry of Δ_1) is denoted δ_1. The singular value represents the maximum covariance between the singular vectors. The first latent variable of **X** is given by [compare with Eq. (4) defining $\mathbf{L_X}$]:

$$\mathbf{t}_1 = \mathbf{X}_0\mathbf{w}_1 \tag{67}$$

where, $\mathbf{t}_1$ is normalized such that $\mathbf{t}_1^T\mathbf{t}_1 = 1$. For our example, matrices **X** and **Y** [from Eqs. (11) and (12)] were transformed into *Z*-scores and the first set of weights $\mathbf{w}_1$ [obtained from the SVD of $\mathbf{R}_1$, see Eq. (66)] gives:

$$\mathbf{w}_1 = [-0.43 \quad 0.20 \quad 0.10 \quad -0.03 \quad -0.00 \quad -0.41 \quad 0.09 \quad 0.16 \quad 0.07 \quad -0.41 \quad 0.16 \quad -0.59]^T, \tag{68}$$

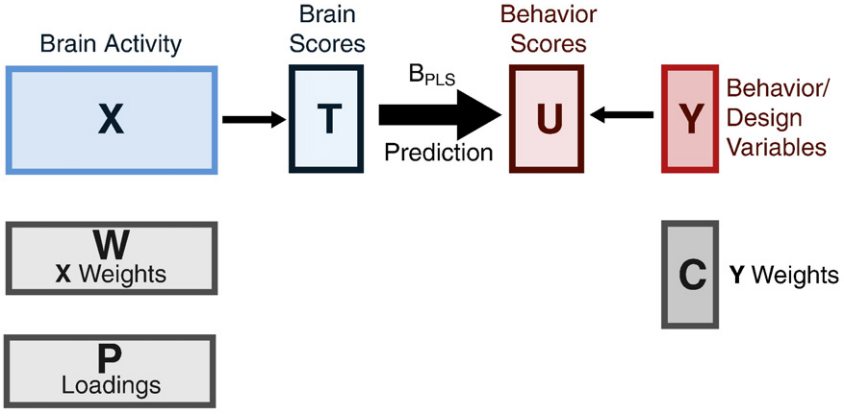

Fig. 14. PLSR: The latent variables of **X** and **Y** are computed iteratively and stored in **T** and **U** respectively. The regression weights from each iteration are stored in $\mathbf{B}_{\mathrm{pls}}$ and are used to predict **Y** from **X** (modified from Wold et al., 2001).

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

which then gives $\mathbf{t}_1$ equal to:

$$\mathbf{t}_1 = [0.41 \quad 0.11 \quad 0.33 \quad 0.28 \quad -0.15 \quad 0.22 \quad -0.45 \quad -0.57 \quad -0.19]^T. \tag{69}$$

The loadings of $\mathbf{X}_0$ on $\mathbf{t}_1$ (i.e., the projection of $\mathbf{X}_0$ onto the space of $\mathbf{t}_1$) are given by:

$$\mathbf{p}_1 = \mathbf{X}_0^T \mathbf{t}_1. \tag{70}$$

The least square estimate of $\mathbf{X}$ from the first latent variable is given by:

$$\widehat{\mathbf{X}}_1 = \mathbf{t}_1^T \mathbf{p}_1. \tag{71}$$

The first latent variable of $\mathbf{Y}$ is obtained as

$$\mathbf{u}_1 = \mathbf{Y}_0 \mathbf{c}_1. \tag{72}$$

For our example $\mathbf{u}_1$ is:

$$\mathbf{u}_1 = [2.16 \quad 1.12 \quad 1.41 \quad 0.12 \quad 0.11 \quad -0.10 \quad -1.43 \quad -1.67 \quad -1.71]^T. \tag{73}$$

So far, PLSR works like PLSC with $\mathbf{t}_1$ and $\mathbf{u}_1$ playing the roles of $\ell_{\mathbf{X},1}$ and $\ell_{\mathbf{Y},1}$, respectively. The step specific to PLSR consists in the prediction of $\mathbf{Y}$ from the "$\mathbf{X}$-latent variable" $\mathbf{t}_1$. This is obtained by first reconstituting $\mathbf{Y}$ from its latent variable as:

$$\widehat{\mathbf{Y}}_1 = \mathbf{u}_1 \mathbf{c}_1^T, \tag{74}$$

and then rewriting Eq. (74) as:

$$\widehat{\mathbf{Y}}_1 = \mathbf{t}_1 b_1 \mathbf{c}_1^T \tag{75}$$

with

$$b_1 = \mathbf{t}_1^T \mathbf{u}_1. \tag{76}$$

The scalar b_1 is the slope of the regression of $\widehat{\mathbf{Y}}_1$ on $\mathbf{t}_1$ (recall that because $\mathbf{Y}$ and $\mathbf{X}$ are centered the regression equation requires only the slope and so there is no intercept in the equation). Eq. (75) shows that $\widehat{\mathbf{Y}}_1$ is obtained as a linear regression from the latent variable extracted from $\mathbf{X}_0$. The regression weight for the example is $b_1 = 3.39$.

Matrices $\widehat{\mathbf{X}}_1$ and $\widehat{\mathbf{Y}}_1$ are then subtracted from the original $\mathbf{X}_0$ and original $\mathbf{Y}_0$ respectively to give *deflated* $\mathbf{X}_1$ and $\mathbf{Y}_1$:

$$\mathbf{X}_1 = \mathbf{X}_0 - \widehat{\mathbf{X}}_1, \tag{77}$$

and

$$\mathbf{Y}_1 = \mathbf{Y}_0 - \widehat{\mathbf{Y}}_1. \tag{78}$$

4.1.2. Next step

The first set of latent variables has now been extracted. Matrices $\mathbf{X}_1$ and $\mathbf{Y}_1$ now become the input matrices for the next iteration and play the roles of $\mathbf{X}_0$ and $\mathbf{Y}_0$, respectively [*cf.*, Eq. (65)]. From the SVD of matrix $\mathbf{R}_2 = \mathbf{X}_1^T \mathbf{Y}_1$, we get $\mathbf{w}_2$, $\mathbf{c}_2$, $\mathbf{t}_2$ and b_2 and the new deflated matrices $\mathbf{X}_2$ and $\mathbf{Y}_2$.

4.1.3. Last step

The iterative process continues until $\mathbf{X}$ is completely decomposed into L components (where L is the rank of $\mathbf{X}$). When this is done, the weights (i.e., all the $\mathbf{w}_\ell$'s) for $\mathbf{X}$ are stored in the J by L matrix $\mathbf{W}$ (whose ℓth column is $\mathbf{w}_\ell$). For our example $\mathbf{W}$ is:

$$\mathbf{W} = \begin{bmatrix} -0.43 & -0.04 & -0.01 & -0.71 & -0.04 & 0.25 & -0.23 & -0.18 \\ 0.20 & -0.36 & -0.04 & 0.09 & -0.28 & -0.05 & -0.67 & -0.27 \\ 0.10 & 0.45 & -0.28 & -0.03 & -0.17 & 0.22 & 0.09 & -0.05 \\ -0.03 & 0.12 & -0.34 & -0.25 & 0.37 & -0.53 & 0.06 & -0.53 \\ 0.00 & 0.40 & 0.50 & -0.30 & -0.50 & -0.32 & 0.00 & 0.02 \\ -0.41 & 0.14 & 0.09 & 0.01 & 0.32 & -0.13 & -0.09 & 0.45 \\ 0.09 & -0.04 & -0.50 & -0.07 & -0.49 & -0.39 & 0.19 & 0.25 \\ 0.16 & -0.09 & 0.47 & 0.05 & 0.10 & -0.33 & -0.09 & -0.08 \\ 0.07 & 0.35 & -0.20 & 0.03 & 0.19 & -0.24 & -0.62 & 0.37 \\ -0.41 & 0.20 & 0.10 & 0.48 & -0.06 & -0.22 & 0.05 & -0.36 \\ 0.16 & -0.41 & 0.09 & -0.27 & 0.19 & -0.31 & 0.23 & 0.24 \\ -0.59 & -0.36 & -0.12 & 0.15 & -0.27 & -0.15 & -0.03 & 0.12 \end{bmatrix}. \tag{79}$$

The latent variables of $\mathbf{X}$ are stored in matrix $\mathbf{T}$. Because $\mathbf{T}$ shows how the brain activity relates to each of the observations, $\mathbf{T}$ has I rows and L columns. For our example $\mathbf{T}$ is:

$$\mathbf{T} = \begin{bmatrix} 0.41 & 0.10 & 0.29 & 0.06 & -0.51 & -0.29 & 0.39 & 0.35 \\ 0.11 & 0.33 & 0.50 & -0.11 & 0.71 & -0.07 & 0.00 & 0.04 \\ 0.33 & 0.24 & 0.00 & -0.21 & -0.31 & 0.09 & -0.57 & -0.50 \\ 0.28 & -0.03 & -0.55 & -0.35 & 0.19 & 0.33 & 0.48 & -0.08 \\ -0.15 & 0.41 & -0.52 & 0.57 & 0.09 & -0.24 & -0.14 & 0.16 \\ 0.22 & -0.65 & 0.13 & 0.54 & 0.13 & 0.25 & -0.10 & -0.09 \\ -0.45 & 0.09 & 0.14 & -0.12 & -0.22 & 0.61 & -0.16 & 0.45 \\ -0.57 & -0.03 & 0.17 & 0.05 & -0.15 & -0.16 & 0.40 & -0.58 \\ -0.19 & -0.47 & -0.17 & -0.43 & 0.07 & -0.53 & -0.29 & 0.24 \end{bmatrix}. \tag{80}$$

The weights for $\mathbf{Y}$ are stored in $\mathbf{C}$. Because $\mathbf{C}$ weights the variables of $\mathbf{Y}$, $\mathbf{C}$ has K rows and L columns. For our example $\mathbf{C}$ is:

$$\mathbf{C} = \begin{bmatrix} -0.71 & -0.73 & -0.37 & -0.97 & 0.34 & 0.71 & -0.81 & -0.93 \\ 0.70 & 0.69 & 0.93 & 0.24 & -0.94 & -0.71 & 0.59 & -0.36 \end{bmatrix}. \tag{81}$$

The latent variables of $\mathbf{Y}$ are stored in matrix $\mathbf{U}$. Because $\mathbf{U}$ shows how the behavioral variables relate to each of the observations, $\mathbf{U}$ has I rows and L columns. For our example $\mathbf{U}$ is:

$$\mathbf{U} = \begin{bmatrix} 2.16 & 0.76 & 0.60 & 0.28 & -0.30 & -0.15 & 0.07 & 0.03 \\ 1.12 & 0.76 & 0.22 & -0.18 & 0.22 & -0.02 & 0.00 & 0.00 \\ 1.41 & 0.28 & -0.06 & -0.19 & -0.02 & 0.12 & -0.10 & -0.04 \\ 0.12 & -0.81 & -0.75 & -0.28 & 0.10 & 0.04 & 0.06 & -0.01 \\ 0.11 & 0.61 & -0.27 & 0.38 & -0.02 & -0.06 & -0.01 & 0.01 \\ -0.10 & -0.86 & 0.19 & 0.22 & 0.12 & 0.09 & -0.02 & -0.01 \\ -1.43 & 0.10 & -0.02 & -0.18 & 0.11 & 0.19 & 0.00 & 0.03 \\ -1.67 & 0.25 & 0.31 & 0.11 & -0.14 & -0.08 & 0.03 & -0.04 \\ -1.71 & -1.08 & -0.23 & -0.16 & -0.08 & -0.12 & -0.03 & 0.02 \end{bmatrix}. \tag{82}$$

The loadings for $\mathbf{X}$ are stored in matrix $\mathbf{P}$. Because $\mathbf{P}$ is the projection of $\mathbf{X}$ onto the space of the latent variables $\mathbf{T}$ and describes the J voxels as explained by the latent variables, $\mathbf{P}$ has J rows and L columns. For our example $\mathbf{P}$ is:

$$\mathbf{P} = \begin{bmatrix} -2.17 & 0.06 & 0.51 & -1.56 & -0.53 & 0.31 & -0.45 & -0.18 \\ 1.76 & -1.26 & -0.47 & 0.00 & -0.67 & -0.78 & -1.40 & -0.27 \\ -0.34 & 2.28 & -1.24 & -0.13 & -0.87 & 0.56 & 0.22 & -0.05 \\ -0.40 & 1.15 & -0.86 & -0.39 & 2.04 & -1.04 & 0.30 & -0.54 \\ -0.78 & 0.62 & 2.01 & -1.28 & -0.93 & -0.68 & 0.00 & 0.02 \\ -2.40 & 0.29 & 0.61 & 0.23 & 1.12 & -0.37 & -0.34 & 0.46 \\ 0.56 & 0.90 & -2.24 & -0.79 & -0.80 & -0.65 & 0.34 & 0.26 \\ 0.98 & -1.30 & 1.98 & 0.10 & 0.87 & -0.79 & -0.17 & -0.08 \\ -0.29 & 1.66 & -0.68 & 0.17 & 0.98 & -1.15 & -1.50 & 0.38 \\ -2.53 & 0.44 & -0.05 & 1.00 & 0.24 & -0.42 & 0.22 & -0.37 \\ 1.61 & -1.69 & 0.79 & -0.60 & 1.06 & -0.44 & 0.43 & 0.25 \\ -2.39 & -1.11 & -0.79 & 0.05 & -0.55 & -0.35 & -0.11 & 0.12 \end{bmatrix}. \tag{83}$$

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

The regression weights are stored in a diagonal matrix **B**. The regression weights are used to predict **Y** from **X**, therefore, there is one b_ℓ for every pair of $\mathbf{t}_\ell$ and $\mathbf{u}_\ell$, and so **B** is an $L \times L$ diagonal matrix. For our example, **B** is:

$$\mathbf{B} = \begin{bmatrix} 3.39 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.74 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.95 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.61 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.34 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.30 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.08 \end{bmatrix}. \quad (84)$$

The predicted **Y** scores (in the form of *Z*-scores) are now given by:

$$\widehat{\mathbf{Y}} = \mathbf{TBC}^T = \mathbf{XB}_{PLS} \quad (85)$$

where, $\mathbf{B}_{PLS} = \mathbf{P}^{T+}\mathbf{BC}^T$, (where $\mathbf{P}^{T+}$ is the Moore–Penrose pseudo-inverse of $\mathbf{P}^T$). $\mathbf{B}_{PLS}$ has *J* rows and *K* columns. For our example, $\mathbf{B}_{PLS}$ is:

$$\mathbf{B}_{PLS} = \begin{bmatrix} 0.58 & -0.43 \\ 0.03 & 0.00 \\ -0.21 & 0.21 \\ 0.11 & -0.08 \\ -0.26 & 0.40 \\ 0.17 & -0.23 \\ -0.06 & 0.02 \\ -0.18 & 0.22 \\ -0.17 & 0.12 \\ -0.01 & -0.02 \\ 0.11 & -0.11 \\ 0.45 & -0.49 \end{bmatrix}. \quad (86)$$

The latent variables of **T** and **U** provide an estimate of **Y** given by $\widehat{\mathbf{Y}}$. In general, $\widehat{\mathbf{Y}}$ is not equal to **Y** (Abdi, 2010). However, in our example, the behavioral variables were perfectly predicted in the fixed effect model.

4.2. What does PLSR optimize?

Because PLSR is a predictive method, its goal is slightly different from PLSC. Specifically, PLSR finds a series of *L* latent variables $\mathbf{t}_\ell$ such that the covariance between $\mathbf{t}_1$ and **Y** is maximal and such that $\mathbf{t}_1$ is uncorrelated with $\mathbf{t}_2$ which has maximal covariance with **Y** and so on for all *L* latent variables (see, e.g., de Jong, 1993; Burnham et al., 1996; Tenenhaus, 1998; Tenenhaus and Tenenhaus, in press; Abdi, 2010; Esposito-Vinzi et al., 2010, for proofs and developments). Formally, we seek a set of *L* linear transformations of **X** that satisfies [compare with Eq. (6)]:

$$\mathbf{t}_\ell = \mathbf{Xw}_\ell \quad \text{such that} \quad \text{cov}(\mathbf{t}_\ell, \mathbf{Y}) = \max \quad (87)$$

(where $\mathbf{w}_\ell$ is the vector of the coefficients of the ℓth linear transformation and cov is the covariance computed between **t** and all columns of **Y**) under the constraints that

$$\mathbf{t}_\ell^T \mathbf{t}_{\ell'} = 0 \quad \text{when} \quad \ell \neq \ell' \quad (88)$$

and

$$\mathbf{t}_\ell^T \mathbf{t}_\ell = 1. \quad (89)$$

4.3. How good is the prediction?

4.3.1. Fixed effect model

A common measure of the quality of prediction is the Residual Estimated Sum of Squares (RESS), which is given by (Abdi, 2010):

$$\text{RESS} = \|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2 \quad (90)$$

where $\|\ \|^2$ is the square of the norm of a matrix (i.e., the sum of squares of all the elements of this matrix). The smaller the value of RESS, the better the quality of prediction (Abdi, 2010; Abdi and Williams, 2010d).

4.3.2. Random effect model

The performance of PLSR (with respect to inference to the population) is done through cross-validation techniques such as the "leave-one-out" procedure (Wold et al., 2001, also called the jackknife; see also Abdi and Williams, 2010b). In this procedure, each observation is removed in turn from both **X** and **Y**, and a PLSR model is re-computed for each of the remaining observations. Then $\mathbf{B}_{PLS}$ is used to predict the left-out observation of **Y** from its **X** values. The predicted observations are stored in $\tilde{\mathbf{Y}}$. The quality of prediction is measured in a way similar to RESS, and is called the Predicted Residual Estimated Sum of Squares (PRESS). Formally PRESS is obtained as (Abdi, 2010):

$$\text{PRESS} = \|\mathbf{Y} - \tilde{\mathbf{Y}}\|^2. \quad (91)$$

The smaller PRESS is, the better the prediction.

An obvious problem with PLSR is to find the optimum number of latent variables for best predicting **Y**. Because neuroimaging data typically have more variables than observations (a problem known as the $N \ll P$ problem), PLSR will over-fit the data (i.e., a very good fixed effect model will do very poorly with new observations). If there is over-fitting, the quality of prediction will first decrease and then increase as more latent variables are used for the prediction of new observations. The number of latent variables at which PRESS increases gives an indication of the optimum number of latent variables to be retained (Abdi, 2010; Abdi and Williams, 2010d).

The next step in the random effect model is to determine the stability of the voxels, (i.e., to decide which voxels reliably depict brain-behavior relationships). For this step, standard errors and confidence intervals of the corresponding PLSR parameters are derived directly from the data through resampling techniques such as the bootstrap (Abdi, 2010; Abdi et al., 2009a).

4.4. Plotting the latent variables

The latent variables of **X** (i.e., **T**) can be plotted against each other to show how they are related to the observations. Sometimes, the weights of **Y** (i.e., **C**) are also plotted along with **T** in order to see how the latent variables of **X** relate to the **Y** variables. In our example, the first latent variable of **X** separates the NC group from the clinical groups, while the second latent variable separates participants NC_3 and PD_3 from the other participants. The second latent variable also differentiates between the words recalled and the reaction times. This may explain why NC_3 and PD_3 are different from the other participants (Fig. 15a). This plot shows the difference between observations based on the part of brain activity that predicts the behavioral measures.

The weights of **X** (i.e., **W**) can be plotted against each other in order to show how **X** variables combine to form the latent variables. In neuroimaging, this plot is not very informative because of the large number of voxels. Instead, the loadings of **X** (i.e., **P**) can be projected on a glass brain to show which regions of the brain best predict the behavior or experimental design.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

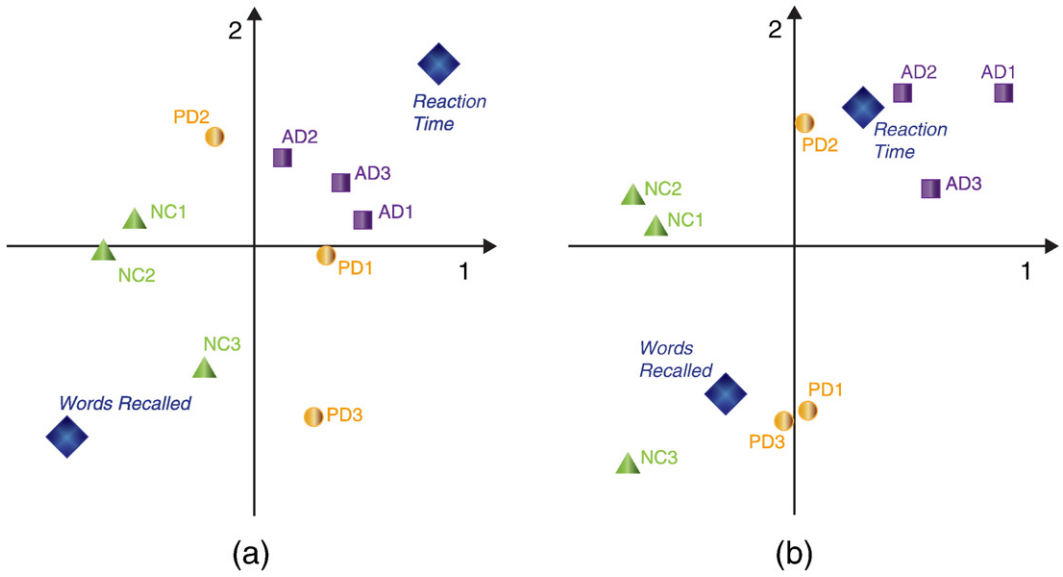

Fig. 15. Mini-example—PLSR: (a) Latent variables of **X** (i.e., **T**)—1 (horizontal) and 2 (vertical); and (b) Latent variables of **Y** (i.e., **U**)—1 (horizontal) and 2 (vertical).

The latent variables of **Y** (i.e., **U**) can be plotted against each other along with the weights of **Y** (i.e., **C**) to see how the latent variables of **Y** relate to the original **Y** variables. In our example, the first latent variable of **Y** separates the NC from the clinical groups, while the second latent variable separates participants NC_3, PD_1 and PD_2 from the other participants. As seen with the plot of the latent variables of **X** (i.e., **T**), the second latent variable also separates the words recalled and reaction time (Fig. 15b).

4.5. Applications

Because it is an iterative procedure, computing latent variables with PLSR takes longer than with PLSC. Nevertheless, PLSR has been used successfully in domains such as chemometrics because it can analyze data with strongly correlated, noisy, and numerous independent variables and also simultaneously model several dependent variables (Wold et al., 2001). The predictive aspect of PLSR makes it a very useful tool to model relationships between large data matrices of brain activity and behavior and to subsequently predict behavior from brain activity.

4.5.1. Shape modeling

One application of PLSR in neuroimaging is the model-based segmentation of sub-cortical brain structures by Rao et al. (2008). Here the goal is to use brain structures that are easy to segment in order to facilitate the segmentation of brain structures that are hard to segment. *Statistical shape modeling*, as this procedure is called, uses PLSR to analyze the shapes of structures in the brain and to describe variations between groups of participants, such as normal controls and clinical populations.

4.5.2. Prediction

PLSR has also been used to predict behavior from neuroimaging data. For example, Giessing et al. (2007) used PLSR to predict behavioral effects of nicotine from neural activity. The predictors were the contrast *f*MRI images between scans under nicotine and scans under placebo for each participant. The dependent variable was the residual term from a linear regression model that predicted the difference in reaction time to validly and invalidly cued targets in a visual discrimination task under the influence of nicotine, from the same difference in reaction time under placebo. PLSR showed that brain activity predicted reaction times of participants during the placebo condition.

PLSR has also been used to predict long term functional outcome of traumatic diffuse axonal injury using Diffusion Tensor Tractography (Wang et al., 2008). The authors used diffusion tensor imaging measurements to predict the outcome of brain injury assessed after six months with the Glasgow Outcome Scale–Extended. PLSR proved to be a reliable quantitative method to predict long term outcomes of diffuse axonal injury.

4.5.3. Classification

PLSR has also been applied as a Discriminant Analysis (PLS-DA) tool for classification experiments. To do this, PLS-DA uses a dummy code as **Y** [see, e.g., Eq. (34)] because the main idea is to classify observations into different groups. PLS-DA has been used to classify different types of dementia (Gottfries et al., 1995). However, the predictors in this case were not brain images but descriptor variables obtained at the time of examination (e.g., Albumin ratio, confusional symptoms, gender, vascular disease). In another study, Lehmann et al. (2007) used PLS-DA to classify various stages of Alzheimer's disease using EEG. The authors compared different types of classification algorithms and found that PLS-DA and machine learning techniques showed analogous performance.

5. Software

PLS methods necessitate sophisticated computations and therefore their application depends on the availability of software.

For neuroimaging, a special toolbox written in MATLAB (by McIntosh, Chau, Lobaugh, & Chen) dedicated to PLSC and a tutorial are freely available from www.rotman-baycrest.on.ca:8080. These programs constitute the standard for PLSC for neuroimaging.

For PLSR there are several available choices. Interested readers can download a set of MATLAB programs (with tutorial) from the senior author's home page (www.utdallas.edu/~herve). In addition, a set of MATLAB scripts implementing the examples used in this paper will also be made available from this home page. Also, a public domain set of MATLAB programs is available from the home page of the *N*-Way project (www.models.kvl.dk/source/nwaytoolbox/) along with tutorials and examples. The statistic toolbox from MATLAB includes a function to perform PLSR. The public domain program R implements PLSR through the package PLS (Mevik and Wehrens, 2007). For neuroimaging, SPM, has recently (2002) integrated a PLS module (as part of the MM toolbox written by Ferath Kherif). The

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

general purpose statistical packages SAS, SPSS, and XLSTAT (which has, by far the most extensive implementation of PLS methods) can be also used to perform PLSR. In chemistry and sensory-evaluation, two main programs are used: the first one called SIMCA-P was developed originally by Wold (who also pioneered PLSR), the second one called the UNSCRAMBLER was first developed by Martens who was another pioneer in the field. And finally, a commercial MATLAB toolbox has also been developed by EIGENRESEARCH.

6. Related methods

A complete review of the connections between PLS and the other statistical methods is, clearly, out of the scope of an introductory paper (see, however, Burnham et al., 1996; Tenenhaus, 1998; Tenenhaus and Tenenhaus, in press; Esposito-Vinzi et al., 2010, for an overview), but some directions are worth mentioning. PLSC uses the SVD in order to analyze the information common to two or more tables, and this makes it closely related to several other SVD (or eigen-decomposition) techniques with similar goals. The closest technique is obviously inter-battery analysis (Tucker, 1958) which uses the same SVD as PLSC but on non-structured matrices. Canonical correlation analysis (also called simply canonical analysis, or canonical variate analysis, see Gittins, 1985; Mardia et al., 1979) is also a related technique that seeks latent variables with largest *correlation* instead of PLSC's criterion of largest *covariance*. Under the assumptions of normality, analytical statistical tests are available for canonical correlation analysis but cross-validation procedures analogous to PLSC could also be used.

In addition, several multi-way techniques encompass as a particular case data sets with two tables. The oldest and most well known technique is multiple factor analysis which integrates different tables into a common PCA by normalizing each table with its first singular value (Abdi and Valentin, 2007a; Escofier and Pagès, 1990). A more recent set of techniques is the STATIS family which uses a more sophisticated normalizing scheme whose goal is to extract the common part of the data (see Abdi and Valentin, 2007b, for an introduction). This approach has been used to identify patterns in *f*MRI data (e.g., Abdi et al., 2009a; Kherif et al., 2003; Shinkareva et al., 2006; Shinkareva et al., 2008). Closely related techniques (which have not been used so far for neuroimaging) comprise common component analysis (Mazerolles et al., 2006) which seeks a set of factors common to a set of data tables, and co-inertia analysis which could be seen as a generalization of Tucker's (1958) inter-battery analysis (see e.g., Thioulouse et al., 2003; Dray et al., 2003; Chessel and Hanafi, 1996; Thioulouse et al., 2003, for recent developments). Also, mean-centered task PLSC is obviously related to PLS-DA because their goals are very similar. Mean-centered task PLSc is also closely related to barycentric discriminant analysis (Abdi and Wiliams, 2010a) because, as previously mentioned, these two techniques compute the SVD of the same matrix.

PLSR is strongly related to regression-like techniques which have been developed to cope with the multi-colinearity problem. These include principal component regression, ridge regression, redundancy analysis (also know as PCA on instrumental variables Rao, 1964; van den Wollenberg, 1977; Tyler, 1982), and continuum regression (Stone and Brooks, 1990) which provides a general framework for these techniques.

7. Conclusion

Partial Least Squares (PLS) methods analyze data from multiple modalities collected on the same observations. We have reviewed two particular PLS methods: Partial Least Squares *Correlation* or PLSC and Partial Least Squares *Regression* or PLSR. Both PLSC and PLSR can be used to study brain activity, behavior, and experimental design. PLSC is *relational* and analyzes the shared information between two or more sets of variables. In contrast, PLSR is *directional* and predicts a set of dependent variables from a set of independent variables or predictors. PLSC is more commonly used in neuroimaging as compared to PLSR.

The two methods differ also in the type of data matrices that are analyzed. Typically data matrices analyzed by PLSC have a complex and specific structure whereas the data matrices analyzed by PLSR tend to be simple matrices. PLSC is also very versatile and has been adapted to several different situations such as multi-block analysis. PLSR, being an iterative process, requires long computation time which can be prohibitive for some applications with very large data sets.

The relationship between PLSC and PLSR are also explored in Burnham et al. (1996) and, recently Gidskehaug et al. (2004) proposed to integrate these two approaches into a new predictive approach called BRIDGE-PLS. In practice, the two techniques are likely to give similar conclusions because the criterion they optimize are quite similar.

Acknowledgments

We would like to thank two anonymous reviewers, Betty Edelman, Cheryl Grady, and Ashwati Krishnan for help and comments on previous versions of this paper.

Appendix A. List of notations

Acronyms		
PCA	Principal components analysis	
PLSC	Partial least squares correlation	
PLSR	Partial least squares regression	
PLS-DA	Partial least squares discriminant analysis	
PLS-PM	Partial least squares path modeling	
SVD	Singular value decomposition	
General Notations		
$\|\ \|$	norm of a matrix	
$\mathbf{1}$	$I \times 1$ vector of ones	
$\mathbf{I}$	identity matrix	
I	the number of observations or rows of $\mathbf{X}$ and $\mathbf{Y}$	
I_n	the number of observations $\mathbf{X}_n$	
J	the number of columns of $\mathbf{X}$	
J_t	the number of columns of $\mathbf{X}_t$	
K	the number of columns of $\mathbf{Y}$	
N	the number of *a priori* sub-matrices of $\mathbf{X}$ (or $\mathbf{Y}$)	
T	the number of scans in spatio-temporal neuroimaging methods, such as EEG, *f*MRI, and NIRS	
T	transpose	
$\mathbf{X}$	the matrix of brain activity (or independent variables)	
$\mathbf{X}_n$	submatrix of $\mathbf{X}$	(Eq. (11))
$\mathbf{X}_t$	an $I \times J_t$ matrix representing a single scan in spatio-temporal neuroimaging	
$\mathbf{Y}$	the matrix of behavioral or design variables (or dependent variables)	
$\mathbf{Y}_n$	submatrix of $\mathbf{Y}$	(Eq. (12))
Singular Value Decomposition (SVD)		
$\mathbf{\Delta}$	diagonal matrix of the singular values (singular values are akin to standard deviations)	(Eq. (1))
δ_ℓ	ℓth singular value	
$\mathbf{U}$	left singular vectors (or saliences) of $\mathbf{Z}$. They represent the design or behavioral profiles that best characterize $\mathbf{R}$.	(Eq. (1))
$\mathbf{u}_\ell$	ℓth left singular vector	
$\mathbf{V}$	right singular vectors (or saliences) of $\mathbf{Z}$. They represent the images that best characterize $\mathbf{R}$.	
$\mathbf{v}_\ell$	ℓth right singular vector	
$\mathbf{Z} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T = \sum_\ell^L \delta_\ell \mathbf{u}_\ell \mathbf{v}_\ell^T$	singular value decomposition of $\mathbf{Z}$	(Eq. (1))

(*continued on next page*)

Partial Least Squares Correlation (PLSC)		
$\mathbf{L_X}$	brain scores/latent variables in PLSC for brain activity; $\mathbf{L_X} = \mathbf{XV}$	(Eq. (4))
$\ell_{\mathbf{X},\ell}$	The ℓth column of $\mathbf{L_X}$	
$\ell^T_{\mathbf{X},\ell}\ell_{\mathbf{Y},\ell}$	scalar product of the latent variables $\ell_{\mathbf{X},\ell}$ and $\ell_{\mathbf{Y},\ell}$ (equal to the corresponding singular value δ_ℓ); Expresses the covariance between these latent variables	(Eqs. (6) and (9))
$\mathbf{L_Y}$	behavior or design scores/latent variables in PLSC for behavior or design variables; $\mathbf{L_Y} = \mathbf{YU}$	(Eq. (5))
$\ell_{\mathbf{Y},\ell}$	The ℓth column of $\mathbf{L_Y}$	
$\mathbf{M}$	matrix of means of each condition in mean-centered PLSC	(Eq. (35))
$\mathbf{R}$	matrix of correlations between the columns of $\mathbf{X}$ and $\mathbf{Y}$; $\mathbf{R} = \mathbf{Y}^T\mathbf{X} = \mathbf{U}\Delta\mathbf{V}^T$	(Eqs. (2) and (3))
$\mathbf{R}_{\text{behavior}}$	matrix of correlations between $\mathbf{X}$ and $\mathbf{Y}_{\text{behavior}}$	(Eq. (15))
$\mathbf{R}_{\text{contrast}}$	matrix of correlations between $\mathbf{X}$ and $\mathbf{Y}_{\text{contrast}}$	(Eq. (26))
$\mathbf{R}_{\text{mean-centered}}$	matrix of the deviations of the groups to their grand mean; $\mathbf{R}_{\text{mean-centered}} = \mathbf{M} - \mathbf{1}\left[\frac{1}{N}\mathbf{1}^T\mathbf{M}\right]$	(Eq. (37))
$\mathbf{R}_{\text{multi}}$	matrix of correlations between $\mathbf{X}$ and multiple $\mathbf{Y}$ matrices in multi-table PLSC	(Eq. (56))
$\mathbf{R}_{\text{seed}}$	matrix of correlations between $\mathbf{X}_{\text{seed}}$ and $\mathbf{Y}_{\text{seed}}$	(Eq. (44))
$\mathbf{X}_{\text{seed}}$	$\mathbf{X}$ matrix with the vectors of $\mathbf{Y}_{\text{seed}}$ removed for seed PLSC	
$\mathbf{Y}_{\text{behavior}}$	$\mathbf{Y}$ matrix of demographic and/or behavioral data	(Eq. (12))
$\mathbf{Y}_{\text{contrast}}$	$\mathbf{Y}$ matrix of orthonormal contrasts reflecting the experimental hypotheses	(Eq. (25))
$\mathbf{Y}_{\text{dummy coding}}$	$\mathbf{Y}$ matrix of dummy coding that codes for experimental groups or conditions	(Eq. (34))
$\mathbf{Y}_{\text{seed}}$	Seed matrix taken from $\mathbf{X}$ for seed PLSC	(Eq. (43))
Partial Least Squares Regression (PLSR)		
$\mathbf{B}$	a diagonal matrix of slopes of the predictions of $\mathbf{Y}$ from $\mathbf{T}$	(Eqs. (62) and (63))
$\mathbf{B}_{\text{PLS}}$	matrix of coefficients to predict $\mathbf{Y}$ from $\mathbf{X}$; equivalent to the regression weights of multiple regression: $\mathbf{B}_{\text{PLS}} = \mathbf{P}^{T+}\mathbf{BC}^T$	(Eq. (64))
$\mathbf{C}$	factor loadings of $\mathbf{Y}$	(Eqs. (62) and (63))
$\mathbf{C}_\ell$	ℓth weight vector of $\mathbf{R}_\ell$	(Eq. (66))
$\mathbf{c}_\ell$	ℓth weight vector of $\mathbf{Y}$	(Eq. (67))
Δ_ℓ	diagonal matrix of singular values	(Eq. (66))
$\mathbf{P}$	factor loadings of $\mathbf{X}$	(Eq. (62))
$\mathbf{P}^+$	Moore-Penrose pseudo inverse of $\mathbf{P}$	(Eq. (64))
$\mathbf{p}_\ell$	ℓth factor loading of $\mathbf{X}$ on $\mathbf{t}$; $\mathbf{p}_\ell = \mathbf{X}^T_{\ell-1}\mathbf{t}_\ell$	(Eq. (70))
$\mathbf{R}_1$	matrix of correlations of $\mathbf{X}_0$ and $\mathbf{Y}_0$; the SVD is performed on the $\mathbf{R}_1$ matrix; $\mathbf{R}_1 = \mathbf{X}_0^T\mathbf{Y}_0 = \mathbf{W}_1\Delta_1\mathbf{C}_1^T$	(Eqs. (65) and (66))
$\mathbf{T}$	matrix of latent variables that model $\mathbf{X}$ and simultaneously predict $\mathbf{Y}$; $\mathbf{T}^T\mathbf{T} = \mathbf{I}$	(Eqs. (62) and (63))
$\mathbf{t}_\ell$	ℓth latent variable of $\mathbf{X}_{\ell-1}$; $\mathbf{t}_\ell = \mathbf{X}_{\ell-1}\mathbf{w}_\ell$	(Eq. (67))
$\mathbf{u}_\ell$	ℓth latent variable of $\mathbf{Y}_{\ell-1}$; $\mathbf{u}_\ell = \mathbf{Y}_{\ell-1}\mathbf{c}_\ell$	(Eq. (72))
$\mathbf{W}_\ell$	matrix of left singular vectors of $\mathbf{R}_\ell$	(Eq. (66))
$\mathbf{w}_\ell$	ℓth weight vector of $\mathbf{X}$	
$\mathbf{X}$	matrix of brain activity; $\mathbf{X} = \mathbf{TP}^T$	(Eq. (62))
$\mathbf{X}_0$	matrix of mean-centered and normalized $\mathbf{X}$ values used in the first iteration	(Eq. (65))
$\underline{\mathbf{X}}_\ell$	matrix of deflated $\mathbf{X}$ values for the ℓth iteration	
$\widehat{\mathbf{X}}_\ell$	matrix of predicted $\mathbf{X}$ values for the ℓth iteration	
$\underline{\mathbf{Y}}$	matrix of behavior or design variables	(Eq. (62))
$\widehat{\mathbf{Y}}_\ell$	the matrix of predicted $\mathbf{Y}$ values for the ℓth iteration	(Eq. (63))
$\mathbf{Y}_0$	centered and normalized $\mathbf{Y}$ values used in the first iteration	(Eq. (65))
$\mathbf{Y}_\ell$	matrix of deflated $\mathbf{Y}$ values for the ℓth iteration	(Eq. (65))
Quality of the PLSR model		
RESS	residual estimated sum of squares; a measure of the quality of the prediction in a fixed effect model; the smaller the value of RESS the better the quality of prediction	(Eq. (90))
PRESS	residual estimated sum of squares; a measure of the quality of the prediction in a random effect model; the smaller the value of PRESS the better the quality of prediction	(Eq. (91))
$\tilde{\mathbf{Y}}$	matrix of predicted observations using the jackknife or "leave-one-out" procedure	

References

Abdi, H., 2007. Singular value decomposition (SVD) and generalized singular value decompositon (GSVD). In: Salkind, N. (Ed.), Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks (CA), pp. 907–912.

Abdi, H., 2010. Partial least square regression, projection on latent structure regression, PLS-Regression. Wiley Interdiscip. Rev. Comput. Stat. 2, 97–106.

Abdi, H., Valentin, D., 2007a. Multiple factor analysis (MFA). In: Salkind, N. (Ed.), Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks (CA), pp. 657–663.

Abdi, H., Valentin, D., 2007b. STATIS. In: Salkind, N. (Ed.), Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks (CA), pp. 955–962.

Abdi, H., Williams, L.J., 2010a. Barycentric discriminant analysis. In: Salkind, N., Dougherty, D., Frey, B. (Eds.), Encyclopedia of Research Design. Sage, Thousand Oaks (CA), pp. 64–75.

Abdi, H., Williams, L.J., 2010b. The jackknife. In: Salkind, N., Dougherty, D., Frey, B. (Eds.), Encyclopedia of Research Design. Sage, Thousand Oaks (CA), pp. 655–660.

Abdi, H., Williams, L.J., 2010c. Matrix algebra. In: Salkind, N., Dougherty, D., Frey, B. (Eds.), Encyclopedia of Research Design. Sage, Thousand Oaks (CA), pp. 761–776.

Abdi, H., Williams, L.J., 2010d. Principal components analysis. Wiley Interdiscip. Rev. Comput. Stat. 2, 433–450.

Abdi, H., Valentin, D., O'Toole, A.J., Edelman, B., 2005. DISTATIS: the analysis of multiple distance matrices. Proceedings of the IEEE Computer Society: International Conference on Computer Vision and Pattern Recognition, pp. 43–47.

Abdi, H., Dunlop, J.P., Williams, L.J., 2009a. How to compute reliability estimates and display confidence and tolerance intervals for pattern classifiers using the Bootstrap and 3-way multidimensional scaling (DISTATIS). Neuroimage 45, 89–95.

Abdi, H., Edelman, B., Valentin, D., Dowling, W.J., 2009b. Experimental design and analysis for psychology. Oxford University Press, Oxford.

Addis, D.R., McIntosh, A.R., Moscovitch, M., Crawley, A.P., McAndrews, M.P., 2004. Characterizing spatial and temporal features of autobiographical memory retrieval networks: a partial least squares approach. Neuroimage 23, 1460–1471.

Bergström, Z.M., Velmans, M., deFockert, J., Richardson-Klavehn, A., 2007. ERP evidence for successful voluntary avoidance of conscious recollection. Brain Res. 1151, 119–133.

Bookstein, F., 1982. The geometric meaning of soft modeling with some generalizations. In: Jöreskog, K., Wold, H. (Eds.), System under indirect observation, Volume 2. North-Holland, Amsterdam.

Bookstein, F.L., 1994. Partial least squares: a dose–response model for measurement in the behavioral and brain sciences. Psycoloquy 5.

Bookstein, F.L., Steissguth, A.P., Sampson, P.D., Conner, P.D., Barr, H.M., 2002. Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. Neuroimage 15, 233–251.

Burnham, A., Viveros, R., MacGregor, J., 1996. Frameworks for latent variable multivariate regression. J. Chemometr. 10, 31–45.

Caplan, J.B., McIntosh, A.R., De Rosa, E., 2007. Two distinct functional networks for successful resolution of proactive interference. Cereb. Cortex 17, 1650–1663.

Chessel, D., Hanafi, M., 1996. Analyse de la co-inertie de K nuages de points. Rev. Stat. Appl. 44, 35–60.

de Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. Chemom. Intell. Lab. Syst. 18, 251–263.

de Jong, S., Phatak, A., 1997. Partial least squares regression. Proceedings of the second international workshop on recent advances in total least squares techniques and error-in-variables modeling. Society for Industrial and Applied Mathematics, pp. 25–36.

de Leeuw, J., 2007. Derivatives of generalized eigen-systems with applications. Department of Statistics Papers, pp. 1–28.

Della-Maggiore, V., Sekuler, A.B., Grady, C.L., Benett, P.J., Sekuler, R., McIntosh, A.R., 2000. Corticolimbic interactions associated with performance on a short-term memory task are modified by age. J. Neurosci. 20, 8410–8416.

Dray, S., Chessel, D., Thioulouse, J., 2003. Co-inertia analysis and the linking of ecological data tables. Ecology 84, 3078–3089.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1, 54–77.

Escofier, B., Pagès, J., 1990. Multiple factor analysis. Comput. Stat. Data Anal. 18, 120–140.

Esposito-Vinzi, V., Wynne, W., Chin, W.W., Henseler, J., Wang, H. (Eds.), 2010. Handbook of Partial Least Squares: Concepts, Methods and Applications. Springer Verlag, New York.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-Mental State". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

Fujiwara, E., Schwartz, M.L., Gao, F., Black, S.E., Levine, B., 2008. Ventral frontal cortex functions and quantified MRI in traumatic brain injury. Neuropsychologia 46, 461–474.

Gidskehaug, L., Stødkilde-Jørgensen, H., Martens, M., Martens, H., 2004. Bridge-PLS regression: two-block bilinear regression without deflation. J. Chemometr. 18, 208–215.

Giessing, C., Fink, G., Rosler, F., Thiel, C., 2007. *f*MRI data predict individual differences of behavioral effects of nicotine: a partial least squares analysis. J. Cogn. Neurosci. 19, 658–670.

Gittins, R., 1985. Canonical Analysis. Springer-Verlag, New York.

Gottfries, J., Blennow, K., Wallin, A., Gottfries, C.G., 1995. Diagnosis of dementias using partial least squares discriminant analysis. Dementia 6, 83–88.

Grady, C.L., McIntosh, A.R., Beig, S., Keightley, M.L., Burian, H., Black, S.E., 2003. Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. J. Neurosci. 23, 986–993.

Greenacre, M., 1984. Theory and applications of correspondence analysis. Academic Press, London.

Iidaka, T., Anderson, N.D., Kapur, S., Cabeza, R., Craik, F.I.M., 2000. The effect of divided attention on encoding and retrieval in episodic memory revealed by positron emission tomography. J. Cogn. Neurosci. 12, 267–280.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034

Keightley, M.L., Seminowicz, D.A., Bagby, R.M., Costa, P.T., Fossati, P., Mayberg, H.S., 2003a. Personality influences limbic–cortical interactions during sad mood induction. Neuroimage 20, 2031–2039.
Keightley, M.L., Winocur, G., Graham, S.J., Mayberg, H.S., Hevenor, S.J., Grady, C.L., 2003b. An fMRI study investigating cognitive modulation of brain regions associated with emotional processing of visual stimuli. Neuropsychologia 41, 585–596.
Kherif, F., Poline, J.-B., Mériaux, S., Benali, H., Flandin, G., Brett, M., 2003. Group analysis in functional neuroimaging: selecting subjects using similarity measures. Neuroimage 20, 2197–2208.
Köhler, S., Moscovitch, M., Wincour, G., Houle, S., McIntosh, A.R., 1998. Networks of domain-specific and general regions involved in episodic memory for spatial location and object identity. Neuropsychologia 36, 129–142.
Lehmann, C., Koenig, T., Jelic, V., Prichep, L., John, R.E., Wahlund, L.-O., Dodge, Y., Dierks, T., 2007. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity. J. Neurosci. Meth. 161, 342–350.
Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. Multivariate analysis. Academic Press, London.
Martens, H., Martens, M., 2001. Multivariate analysis of quality: an introduction. Wiley, London.
Martens, H., Naes, T., 1989. Multivariate calibration. Wiley, London.
Martinez-Montes, E., Valdés-Sosa, P.A., Miwakeichi, F., Goldman, R.I., Cohen, M.S., 2004. Concurrent EEG/fMRI analysis by partial least squares. Neuroimage 22, 1023–1034.
Mazerolles, G., Hanafi, M., Dufour, E., Bertrand, D., Qannari, E., 2006. Common components and specific weights analysis: a chemometric method for dealing with complexity of food products. Chemometr. Intell. Lab. Syst. 81, 41–49.
McIntosh, A., Gonzalez-Lima, F., 1991. Structural modeling of functional neural pathways mapped with 2-deoxyglucose: effects of acoustic startle habituation on the auditory system. Brain Res. 547, 295–302.
McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23, S250–S263.
McIntosh, A.R., Bookstein, F., Haxby, J., Grady, C., 1996. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage 3, 143–157.
McIntosh, A., Nyberg, L., Bookstein, F.L., Tulving, E., 1997. Differential functional connectivity of prefrontal and medial temporal cortices during episodic memory retrieval. Hum. Brain Mapp. 5, 323–327.
McIntosh, A.R., Lobaugh, N.J., Cabeza, R., Bookstein, F.L., Houle, S., 1998. Convergence of neural systems processing stimulus associations and coordinating motor responses. Cereb. Cortex 8, 648–659.
McIntosh, A.R., Chau, W., Protzner, A., 2004. Spatiotemporal analysis of event-related fMRI data using partial least squares. Neuroimage 23, 764–775.
Mentis, M.J., Dhawan, V., Feigin, A., Delalot, D., Zgaljardic, D., Edwards, C., Eidelberg, D., 2003. Early stage Parkinson's disease patients and normal volunteers: comparative mechanisms of sequence learning. Hum. Brain Mapp. 20, 246–258.
Menzies, L., Achard, S., Chamberlain, S.R., Fineberg, N., Chen, C.-H., delCampo, N., Sahakian, B.J., Robbins, T.W., Bullmore, E., 2007. Neurocognitive endophenotypes of obsessive–compulsive disorder. Brain 130, 3223–3236.
Mevik, B.-H., Wehrens, R., 2007. The PLS package: principal component and partial least squares regression in R. J. Stat. Softw. 18, 1–24.
Nestor, P.G., O'Donnell, B.F., Mccarley, R.W., Niznikeiwicz, M., Barnard, J., Shen, Z.J., Bookstein, F.L., Shenton, M.E., 2002. A new statistical method for testing hypotheses of neuropsychological/MRI relationships in schizophrenia: partial least squares analysis. Schizophr. Res. 53, 57–66.
Nyberg, L., McIntosh, A.R., Cabeza, R., Habib, R., Houle, S., Tulving, E., 1996. General and specific brain regions involved in encoding and retrieval of events: what, where and when. Proc. Natl. Acad. Sci. USA 93, 11280–11285.
Nyberg, L., Persson, J., Habib, R., Tulving, E., McIntosh, A.R., Cabeza, R., Houle, S., 2000. Large scale neurocognitive networks underlying episodic memory. J. Cogn. Neurosci. 12, 163–173.
Nyberg, L., Forkstam, C., Petersson, K.M., Cabeza, R., Ingvar, M., 2002. Brain imaging of human memory systems: between-systems similarities and within-system differences. Brain Res. Interact. 13, 281–292.
Price, J., Ziolko, S., Weissfeld, L., Klunk, W., Lu, X., Hoge, J., Meltzer, C., Davis, S., Lopresti, B., Holt, D., DeKosky, S., Mathis, C., 2004. Quantitative and statistical analyses of PET imaging studies of amyloid deposition in humans. Nuclear Science Symposium Conference Record: 2004 IEEE, 5, pp. 3161–3164.
Protzner, A.B., McIntosh, A.R., 2007. The interplay of stimulus modality and response latency in neural network organization for simple working memory tasks. J. Neurosci. 27, 3187–3197.
Rajah, M.N., McIntosh, A.R., 2005. Overlap in the functional neural systems involved in semantic and episodic memory retrieval. J. Cogn. Neurosci. 17, 470–482.
Rao, C., 1964. The use and interpretation of principal component analysis in applied research. Sankhya 26, 329–359.
Rao, A., Aljabar, P., Rueckert, D., 2008. Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. Med. Image Anal. 12, 55–68.
Shinkareva, S., Ombao, H., Sutton, B., Mohanty, A., Miller, G., 2006. Classification of functional brain images with a spatio-temporal dissimilarity map. Neuroimage 33, 63–71.
Shinkareva, S.V., Mason, R.A., Malave, V.L., Wang, W., Mitchell, T.M., Just, M.A., 2008. Using *f*MRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS ONE 3.
Sidtis, J.J., Strother, S.C., Rottenberg, D.A., 2003. Predicting performance from functional imaging data: methods matter. Neuroimage 20, 615–624.
Stone, M., Brooks, R.J., 1990. Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. J. R. Stat. Soc. B Methodol. 52, 237–269.
Streissguth, A., Bookstein, F., Sampson, P., Barr, H., 1993. Methods of latent variable modeling by partial least squares. The Enduring Effects of Prenatal Alcohol Exposure on Child Development. University of Michigan Press.
Takane, Y., 2002. Relationships among various kinds of eigenvalue and singular value decompositions. In: Yanai, H., Okada, A., Shigemasu, K., Kano, Y., Meulman, J. (Eds.), New developments in psychometrics. Springer Verlag, Tokyo, pp. 45–46.
Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme.
Tenenhaus, M., 1998. La regression PLS. Technip, Paris.
Tenenhaus, M., Tenenhaus, A., in press, Regularized generalized canonical correlation analysis. Psychometrika.
Thioulouse, J., Simier, M., Chessel, D., 2003. Simultaneous analysis of a sequence of paired ecological tables. Ecology 20, 2197–2208.
Tippett, W., Black, S., 2008. Regional cerebral blood flow correlates of visuospatial tasks in Alzheimer's disease. J. Int. Neuropsychol. Soc. 14, 1034–1045.
Tucker, L., 1958. An inter-battery method of factor analysis. Psychometrika 23, 111–136.
Tyler, D.E., 1982. On the optimality of the simultaneous redundancy transformations. Psychometrika 47, 77–86.
Vallesi, A., McIntosh, A.R., Alexander, M.P., Stuss, D.T., 2009. *f*MRI evidence of a functional network setting the criteria for withholding a response. Neuroimage 45, 537–548.
van den Wollenberg, A., 1977. Redundancy analysis: an alternative to canonical correlation. Psychometrika 42, 207–219.
Wang, J.Y., Bakhadirov, K., Devous Sr., M.D., Abdi, H., McColl, R., Moore, C., de la Plata, C.D.M., Ding, K., Whittemore, A., Babcock, E., Rickbeil, T., Dobervich, J., Kroll, D., Dao, B., Mohindra, N., Madden, C.J., Diaz-Arrastia, R., 2008. Diffusion tensor tractography of traumatic diffuse axonal injury. Arch. Neurol. 65, 619–626.
West, R., Krompinger, J., 2005. Neural correlates of prospective and retrospective memory. Neuropsychologia 43, 418–433.
West, R., Wymbs, N., 2004. Is detecting prospective cues the same as selecting targets? An ERP study. Cogn. Affect. Behav. Neurosci. 4, 354–363.
Williams, L.J., Abdi, H., French, R., Orange, J.B., 2010. A tutorial on Multi-Block Discriminant Correspondence Analysis (MUDICA): A new method for analyzing discourse data from clinical populations. Journal of Speech Language and Hearing Research 53, 1372–1393.
Wold, H., 1982. Soft modelling, the basic design and some extensions. In: Wold, H., Jöreskog, K.-G. (Eds.), Systems Under Indirect Observation: Causality-Structure-Prediction. Part II. North-Holland Publishing Company, Amsterdam, pp. 1–54.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Please cite this article as: Krishnan, A., et al., Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review, NeuroImage (2010), doi:10.1016/j.neuroimage.2010.07.034