

myPLS - Partial least squares analysis for medical image processing

Daniela Zöller* ^{1,2}, Valeria Kebets* ^{2,3}, Dimitri Van De Ville ^{1,2}

* equal contribution

¹ Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

² Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland

³ Department of Electrical and Computer Engineering, N.1 Institute for Health and Memory Networks Program, Centre for Sleep and Cognition, Clinical Imaging Research Centre, National University of Singapore, Singapore

Table of Contents

Introduction	2
Principles of PLS analysis	2
Types of PLS analyses	4
Behavioral PLS	4
Behavioral PLS with grouping	4
Behavioral PLS without grouping	5
Contrast PLS	5
Contrast PLS, no behavioral data	6
Contrast PLS, with behavioral data	6
Procrustes transform options (pls_opts.boot_procrustes_mod)	7
Grouping options	7
Grouping for PLS analysis (input.grouping)	7
Grouping for permutation testing (pls_opts.grouped_perm)	7
Grouping for bootstrapping (pls_opts.grouped_boot)	8
Grouping for plots (save_opts.grouped_plots)	8
Loadings or Saliences?	8
Visualisation	8
References	9

Introduction

Unsupervised learning methods such as Partial Least Squares (PLS) can allow to overcome the limitations that arise with classification when classes are not well defined. PLS is a data-driven multivariate statistical technique that aims to extract relationships between two data matrices (McIntosh et al., 2004). PLS has previously been used to link neural variability with age (Garrett et al., 2010), or atrophy to symptoms in Parkinson's disease (Zeighami et al., 2019). Here, we present a toolbox that deploys Behavior PLS, which aims to maximize the covariance between neuroimaging and behavioral data by deriving latent components (LCs) that are optimally weighted linear combinations of the original variables.

In the following paragraphs, we will first outline the general principles of PLS and give an overview over the different types of analyses that can be done with the toolbox, followed by a description of the most important options to consider when setting up your analysis and some guidelines regarding the interpretation.

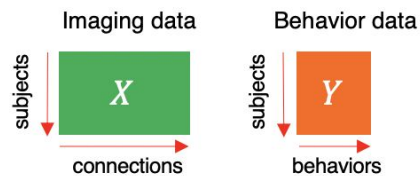
The aim of this document is to give a general overview on the different possibilities of myPLS, and to provide some broad guidelines for the setup of your analysis and for the interpretation of the results. It is **not** a step-by-step description of all available options. For a detailed overview on all available options, please have a look at the example scripts.

Principles of PLS analysis

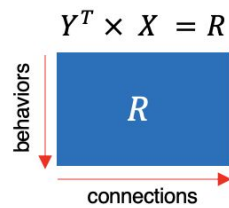
The myPLS Toolbox is based on PLS correlation (PLSC) described in Krishnan et al. (2011). In PLSC analysis, imaging data (a vectorized connectivity matrix in this example) is stored in matrix X , while matrix Y contains the behavior data. A cross-covariance matrix R is first computed between X and Y . Singular value decomposition is then applied on R , resulting in 3 low-dimensional matrices: U , S and V . U and V are singular vectors (called behavioral and imaging saliences that reflect the contribution of the original variables on the latent components), while S is a diagonal matrix containing the singular values. Subject-specific imaging (L_x) and behavioral scores (L_y) are then computed for each latent component by projecting X and Y onto their respective saliences V and U . Imaging and behavioral structure coefficients (or "loadings"; Courville and Thompson, 2001) can also be obtained by computing Pearson's correlations between the original data (X and Y) and subject-specific scores (L_x , L_y).

Partial least squares analysis

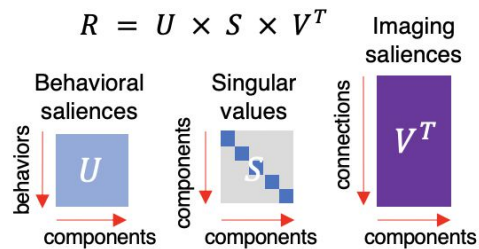
Original data



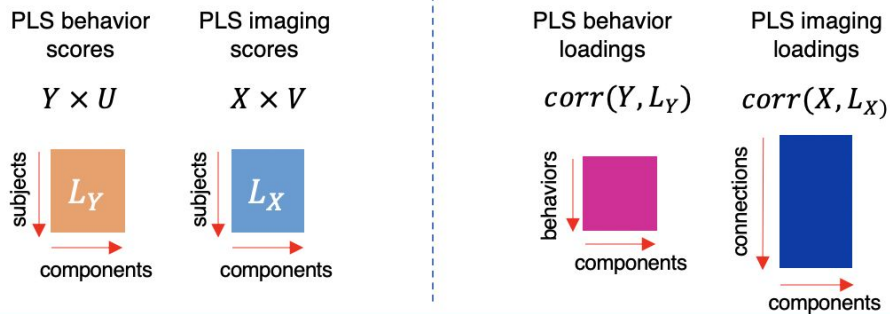
Cross-covariance matrix



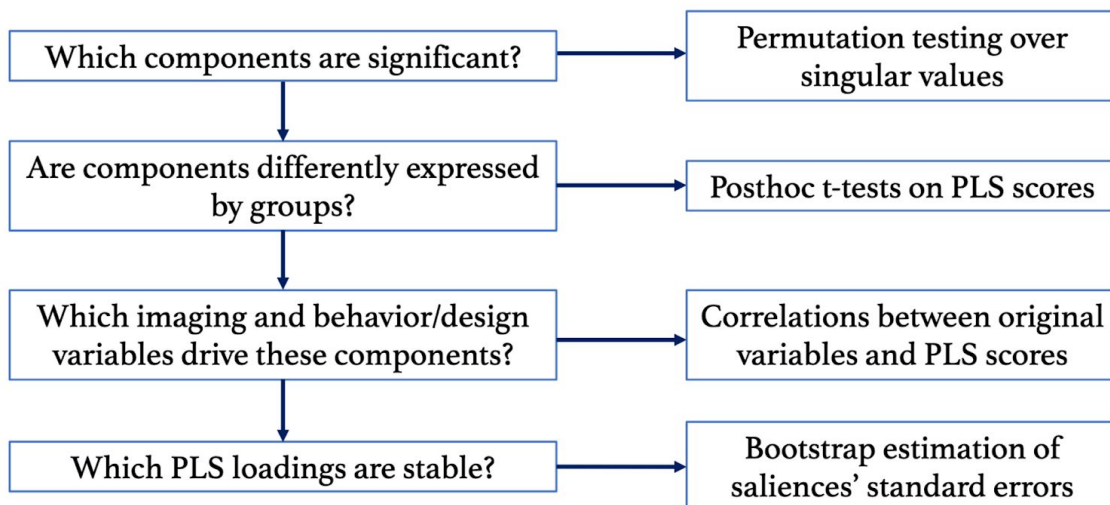
Singular value decomposition



PLS scores & loadings



PLS interpretation



Types of PLS analyses

myPLS allows to run behavioral PLS (as for example in Kebets et al., 2019) or contrast PLS (as for example in Zöllner et al., 2017).

Behavioral PLS

The goal of grouped behavioral PLS (see also Krishnan et al., 2011) is to identify common or differing effects of covariance in different groups *disregarding* absolute group differences. Thus, behavioral variables and imaging variables are usually normalized *within* each group. Then, the covariance matrix R is computed separately for each group, and concatenated before singular value decomposition. In this way, it is possible to detect differences in the brain-behavior relationships between the groups. However, the absolute differences in behavior are disregarded.

If one does not want to disregard absolute group differences in behavioral and imaging variables, it is possible to compute behavioral PLS across all subjects (i.e., omitting the grouping information). For more details on this approach, we refer the interested reader to Kebets et al., 2019, where this PLS variant was used to investigate the alignment of psychiatric diagnostic categories with the covariance between RSFC and transdiagnostic psychiatric symptoms.

Behavioral PLS with grouping

The Toolbox options for behavior PLS with grouping are:

`behav_type = 'behavior'`

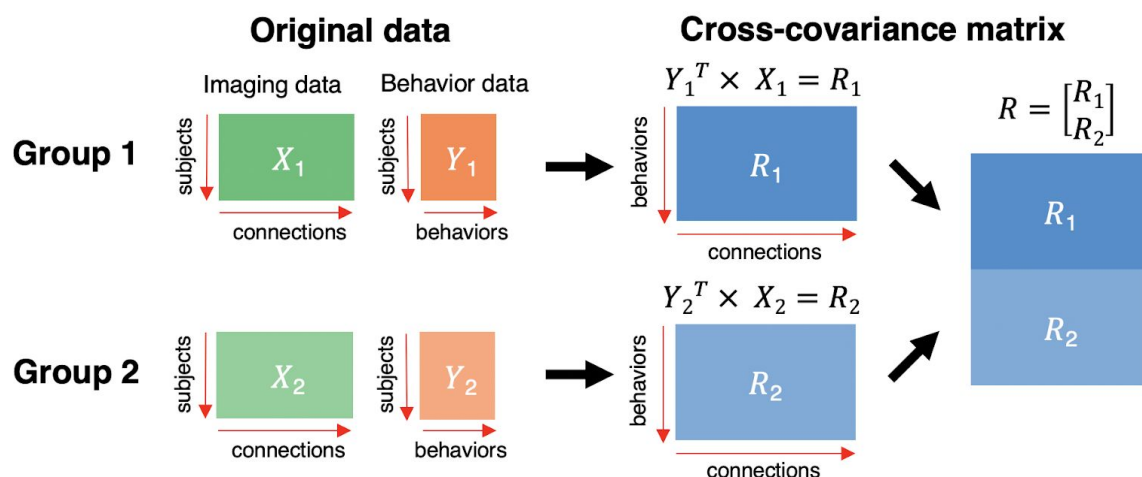
`grouped_PLS = 1`

Usually variables will be normalized within every group prior to grouped behavioral PLS:

`normalization_behav = 2`

`normalization_img = 2`

Behavioral PLS with grouping



Behavioral PLS without grouping

The Toolbox options for behavioral PLS without grouping are:

`behav_type = 'behavior'`

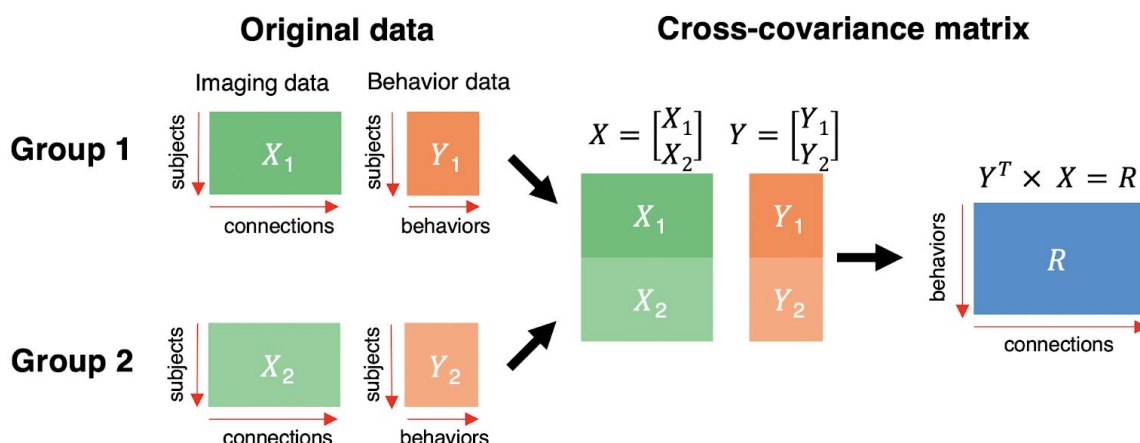
`grouped_PLS = 0`

Usually variables will be normalized across all subjects for ungrouped PLS:

`normalization_behav = 1`

`normalization_img = 1`

Behavioral PLS without grouping



Contrast PLS

If grouping information is disregarded in behavioral PLS, absolute differences in the behavior and/or in the imaging data could be mixed with continuous effects of brain-behavior covariance, which could be undesired for some research questions.

Contrast PLS allows to disentangle *absolute differences* between groups, effects of covariance that are *common* to both groups, and covariance effects that are *specific* for each group.

In contrast PLS, the grouping information is included into the behavior data matrix Y as a categorical variable. This approach is closely linked to *contrast task PLS* defined in (Krishnan et al., 2011). However, here, we compute the contrast between diagnostic groups rather than between task conditions and the myPLS toolbox allows to combine contrast and behavioral data. For more details on this approach, we refer the interested reader to (Zöllner et al., 2017), where this PLS variant was used to investigate developmental changes of BOLD variability in patients at risk for psychosis in comparison to healthy control subjects. Contrast PLS without behavior is closely related to linear discriminant analysis (LDA). For details on the relationship between PLS and LDA, we refer the interested reader to (Brereton and Lloyd, 2014).

Contrast PLS, no behavioral data

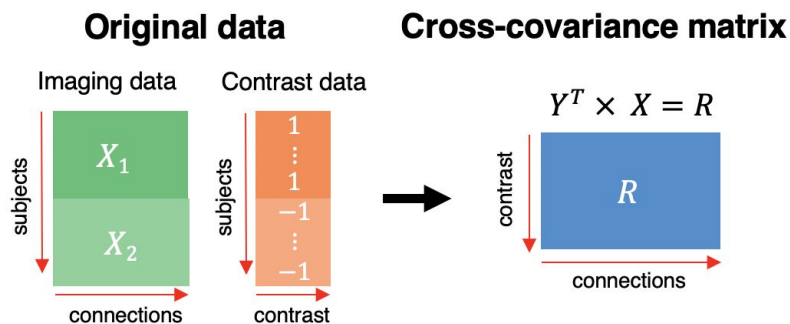
The Toolbox options for contrast PLS without behavior are:

- `behav_type` = 'contrast'
- `grouped_PLS` = 0

Variables have to be normalized across all subjects for contrast PLS:

- `normalization_behav` = 1
- `normalization_img` = 1

Contrast PLS, no behavior



Contrast PLS, with behavioral data

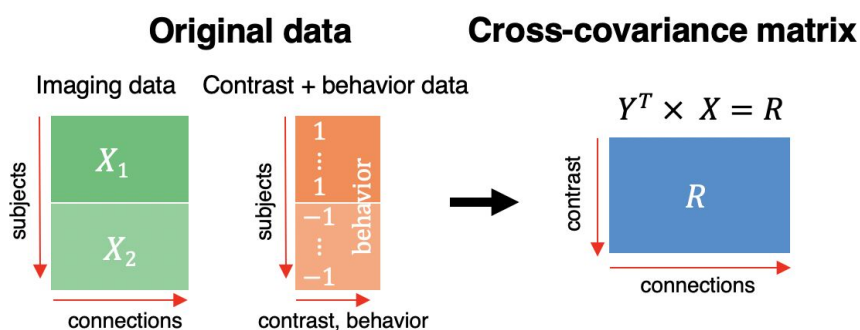
The Toolbox options for contrast PLS including behavior are:

- `behav_type` = 'contrastBehav' (for contrast and whole-group behavior effects)
or `behav_type` = 'contrastBehavInteract' (to include also contrast*behavior
- this option was used in Zöllner et al., 2017)
- `grouped_PLS` = 0

Variables have to be normalized across all subjects for contrast PLS:

- `normalization_behav` = 1
- `normalization_img` = 1

Contrast PLS, with behavior



Procrustes transform options (pls_opts.boot_procrustes_mod)

When using permutation testing or bootstrapping, resampling can cause axis rotation or reflection, whereby the order of the components extracted with each permutation might change (i.e. axis rotation) or the sign of the saliences for each bootstrap sample might change (i.e. axis reflection). Procrustes rotation is used to correct for this (see McIntosh and Lobaugh, 2004 for details). The default option (=1) is a standard rotation computed on U (behavior saliences) only.

In some cases, the bootstrap-estimated variance of behavior saliences can be very close to zero. In this case, we recommend using the second rotation option (=2), whereby the rotation is computed on both U and V, and then averaged across the two.

Don't hesitate to contact us if this happens to you !

Grouping options

Grouping for PLS analysis (input.grouping)

Subjects can be assigned to one or multiple groups for the PLS analysis:

- If one group is indicated, PLS will aim to maximize covariance across all subjects. The cross-covariance matrix will be computed across all subjects. Singular value decomposition (SVD) will produce one set of behavioral saliences and one set of imaging saliences across all subjects.
- If multiple groups are indicated, PLS will aim to maximize covariance **within each group**. In this case, the **cross-covariance matrix will be computed for each group** separately, and then covariance matrices of all groups will be concatenated together. The singular value decomposition (SVD) will produce a set of **separate behavioral saliences for each group** (but only one set of imaging saliences for all groups).

Grouping for permutation testing (pls_opts.grouped_perm)

Subjects can be assigned to one or multiple groups for permutation testing:

- 0 = subjects' behavior data will be randomly shuffled **across all subjects** during permutation testing.
- 1 = subjects' behavior data will be randomly shuffled **within each group** during permutation testing.

Usually, the grouping would be the same as the one indicated for the PLS analysis. However, in some cases, one might want to compute the PLS analysis across all subjects, but to assign subjects to different groups for permutation testing. Such an example can be found in Kebets et al. 2019, where the PLS analysis was computed across healthy individuals and individuals with various psychiatric disorders. Groups were not considered for the PLS analysis, because the aim was to identify transdiagnostic brain-behavior associations. However, groups were considered for permutation testing and bootstrapping because of significant group differences for most behavioral variables.

Grouping for bootstrapping (pls_opts.grouped_boot)

Subjects can be assigned to one or multiple groups for bootstrapping:

- 0 = subjects' imaging and behavior data will be resampled with replacement across all subjects during bootstrapping.
- 1 = subjects' imaging and behavior data will be resampled with replacement **within each group** during bootstrapping.

Just as for permutations, the grouping would usually be the same as the one indicated for the PLS analysis. See example above (*Grouping for permutation testing*) for an example of a case where different group options were used for the PLS analysis and bootstrapping.

Grouping for plots (save_opts.grouped_plots)

Subjects can be assigned to different groups when they are visualized on plots (e.g. scatterplots showing imaging and behavior/design scores). These groups don't have to be the same as the groups used for the PLS analysis or for permutations and bootstrapping.

Loadings or Saliences?

Different types of weights can be used to interpret the contribution of imaging (or behavior) variables to the PLS components. We provide two options in myPLS:

- 1) PLS *saliences*, which are akin to the coefficients in PCA
- 2) PLS *loadings*, which are Pearson's correlations between the original imaging (or behavior) data and the imaging (or behavior) saliencs. They are often called structure coefficients in the multiple regression/canonical correlation analysis literature.

Both saliencs and loadings provide a weight for each imaging variable and each component, and for each behavior variable and each component.

It is recommended to use loadings when you have multicollinearity among your imaging or behavior data. Moreover, loadings might be more easy to interpret (as correlations, they range between -1 and 1) than a coefficient (abstract scale).

Bootstrapping can be used to assess the stability or robustness of saliencs/loadings, providing confidence intervals as well as a bootstrap ratio (BSR), which can be interpreted as a z-score.

Visualisation

- Behavior saliencs/loadings and 1D imaging saliencs/loadings are shown on barplots, with their bootstrap-estimated confidence intervals
- For 2D and 3D imaging saliencs/loadings, we recommend displaying BSR (i.e. z-scores) to show only the imaging variables (e.g. connections or voxels) that are significantly associated with a component

References

- Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics*, 28(4), 213–225.
- Courville T., Thompson B. (2001), 'Use of structure coefficients in published multiple regression articles: β is not enough', *Educational and Psychological Measurement*, vol. 61, issue 2, pp. 229–248.
- Garrett D.D., Kovacevic N., McIntosh A.R., Grady C.L. (2010), 'Blood oxygen level-dependent signal variability is more than just noise', *Journal of Neuroscience*, vol. 30, issue 14, pp. 4914-21.
- Kebets V., Holmes A.J., Orban C., Tang S., Li J., Sun N., Kong R., Poldrack R.A., Yeo B.T.T. (2019), 'Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology', *Biological Psychiatry*, vol. 86, issue 10, pp. 779-791.
- Krishnan, A., Williams, L.J., McIntosh, A.R., Abdi, H. (2011), 'Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review', *Neuroimage*, vol. 56, pp. 455–475.
- McIntosh, A.R., Lobaugh, N.J. (2004), 'Partial least squares analysis of neuroimaging data: Applications and advances', *Neuroimage*, vol. 23, pp. 250–263.
- Zeighami Y., Fereshtehnejad S.M., Dadar M., Collins D.L., Postuma R.B., Mišić B., Dagher A. (2019), 'A clinical-anatomical signature of Parkinson's disease identified with partial least squares and magnetic resonance imaging', *Neuroimage*, vol. 190, pp. 69-78.
- Zöllner, D., Schaer, M., Scariati, E., Padula, M. C., Eliez, S., & Van De Ville, D. (2017). Disentangling resting-state BOLD variability and PCC functional connectivity in 22q11.2 deletion syndrome. *NeuroImage*, 149(January), 85–97.
- Zöllner, D., Sandini, C., Karahanoğlu, F. I., Padula, M. C., Schaer, M., Eliez, S., & Van De Ville, D. (2019), 'Large-scale brain network dynamics provide a measure of psychosis and anxiety in 22q11.2 deletion syndrome', *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 4, issue 10, pp. 881-892.