Appendix A

Derek Beaton

Rotman Research Institute, Baycrest Health Sciences

and

ADNI *

ADNI

and

Gilbert Saporta

Conservatoire National des Arts et Metiers

and

Hervé Abdi

Behavioral and Brain Sciences, The University of Texas at Dallas

July 3, 2020

Abstract

A

Keywords: generalized singular value decomposition, latent models, genetics, neuroimaging, canonical correlation analysis

1 Appendix A

This Appendix provides additional details on PLS-CA-R and, in particular, a variety of extensions of and generalizations from PLS-CA-R. We also provide additional details on some concepts, such as the generalized singular value decomposition.

[A bit more of an introduction on what we're going to do here...]

1.1 The SVD, GSVD, and GPLSSVD

The SVD of $\mathbf{X}$ is

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{T}, \tag{1}$$

where $\mathbf{U}^{T}\mathbf{U} = \mathbf{I} = \mathbf{V}^{T}\mathbf{V}$. The GSVD of $\mathbf{X}$ is

$$\mathbf{X} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{T}, \tag{2}$$

where $\mathbf{P}^{T}\mathbf{M}\mathbf{P} = \mathbf{I} = \mathbf{Q}^{T}\mathbf{W}\mathbf{Q}$. Practically, the GSVD is performed through the SVD as $\widetilde{\mathbf{X}} = \mathbf{M}^{\frac{1}{2}}\mathbf{X}\mathbf{W}^{\frac{1}{2}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{T}$, where the generalized singular vectors are computed from the singular vectors as $\mathbf{P} = \mathbf{M}^{-\frac{1}{2}}\mathbf{U}$ and $\mathbf{Q} = \mathbf{W}^{-\frac{1}{2}}\mathbf{V}$. The relationship between the SVD and GSVD can be expressed as

$$\widetilde{\mathbf{X}} = \mathbf{M}^{\frac{1}{2}}\mathbf{X}\mathbf{W}^{\frac{1}{2}} \Longleftrightarrow \mathbf{X} = \mathbf{M}^{-\frac{1}{2}}\widetilde{\mathbf{X}}\mathbf{W}^{-\frac{1}{2}} \tag{3}$$

Similarly to the GSVD, let us refer to $\widetilde{\mathbf{Z}}_{\mathbf{X}} = \mathbf{M}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{Z}_{\mathbf{X}}\mathbf{W}_{\mathbf{X}}^{\frac{1}{2}}$ and $\widetilde{\mathbf{Z}}_{\mathbf{Y}} = \mathbf{M}_{\mathbf{Y}}^{\frac{1}{2}}\mathbf{Z}_{\mathbf{Y}}\mathbf{W}_{\mathbf{Y}}^{\frac{1}{2}}$.

The GPLSSVD makes use of the SVD wherein $\widetilde{\mathbf{Z}}_{\mathbf{R}} = \widetilde{\mathbf{Z}}_{\mathbf{X}}^{T}\widetilde{\mathbf{Z}}_{\mathbf{Y}} = (\mathbf{M}_{\mathbf{X}}^{\frac{1}{2}}\mathbf{Z}_{\mathbf{X}}\mathbf{W}_{\mathbf{X}}^{\frac{1}{2}})^{T}(\mathbf{M}_{\mathbf{Y}}^{\frac{1}{2}}\mathbf{Z}_{\mathbf{Y}}\mathbf{W}_{\mathbf{Y}}^{\frac{1}{2}}) = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{T}$. The GPLSSVD generalized singular vectors and component scores are computed as $\mathbf{P} = \mathbf{W}_{\mathbf{X}}^{-\frac{1}{2}}\mathbf{U}$ and $\mathbf{F}_{J} = \mathbf{W}_{\mathbf{X}}\mathbf{P}\boldsymbol{\Delta}$ for the J columns of $\mathbf{X}$, and $\mathbf{Q} = \mathbf{W}_{\mathbf{Y}}^{-\frac{1}{2}}\mathbf{V}$ and $\mathbf{F}_{K} = \mathbf{W}_{\mathbf{Y}}\mathbf{Q}\boldsymbol{\Delta}$ for the K columns of $\mathbf{Y}$. Like with the SVD $\mathbf{U}^{T}\mathbf{U} = \mathbf{I} = \mathbf{V}^{T}\mathbf{V}$, and like with the GSVD $\mathbf{P}^{T}\mathbf{W}_{\mathbf{X}}\mathbf{P} = \mathbf{I} = \mathbf{Q}^{T}\mathbf{W}_{\mathbf{Y}}\mathbf{Q}$.

The GPLSSVD also produces scores for the I rows of each matrix—usually called latent variables—as $\mathbf{L}_{\mathbf{X}} = \widetilde{\mathbf{Z}}_{\mathbf{X}}\mathbf{U}$ and $\mathbf{L}_{\mathbf{Y}} = \widetilde{\mathbf{Z}}_{\mathbf{Y}}\mathbf{V}$ where $\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \boldsymbol{\Delta}$. By its definition, this GPLSSVD maximization of the latent variables—i.e., $\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \boldsymbol{\Delta}$—is the "PLS correlation" decomposition.

Specifically, if $\mathbf{X}$ and $\mathbf{Y}$ were each column-wise centered and/or normalized, then GPLSSVD$(\mathbf{I}, \mathbf{Z}_{\mathbf{X}}, \mathbf{I}, \mathbf{I}, \mathbf{Z}_{\mathbf{Y}}, \mathbf{I})$ is PLS correlation (a.k.a. PLSSVD or Tucker's approach).

1.2   From GSVD triplet to GPLSSVD sextuplet

We introduce an extension of the GSVD triplet for PLS, called the "GPLSSVD sextuplet". The GPLSSVD sextuplet helps us in two ways: (1) it simplifies some of the notation and decomposition concepts used here, and (2) it provides the basis for generalization of cross-decomposition methods (e.g., canonical correlation, reduced rank regression) that we expand upon in the Appendix.

The "GPLSSVD sextuplet" takes the form of GPLSSVD$(\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y})$ and decomposes the matrix $\mathbf{Z_R} = (\mathbf{M_X^{\frac{1}{2}}}\mathbf{Z_X})^T \mathbf{M_Y^{\frac{1}{2}}}\mathbf{Z_Y}$ as

$$\mathbf{Z_R} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^T \text{ with } \mathbf{P}^T\mathbf{W_X}\mathbf{P} = \mathbf{I} = \mathbf{Q}^T\mathbf{W_Y}\mathbf{Q}. \tag{4}$$

From this decomposition, the GPLSSVD creates latent variables as

$$\mathbf{L_X} = \mathbf{Z_X}\mathbf{W_X}\mathbf{P} \text{ and } \mathbf{L_Y} = \mathbf{Z_Y}\mathbf{W_Y}\mathbf{Q} \text{ where } \mathbf{L_X}^T\mathbf{L_Y} = \boldsymbol{\Delta}. \tag{5}$$

When $\mathbf{X}$ and $\mathbf{Y}$ are column-wise centered normalized and when all constraint matrices are equal to the identity matrix—i.e., GPLSSVD$(\mathbf{I}, \mathbf{Z_X}, \mathbf{I}, \mathbf{I}, \mathbf{Z_Y}, \mathbf{I})$—the GPLSSVD implements the "PLS correlation" decomposition (Krishnan et al. 2011, Bookstein 1994, McIntosh et al. 1996), also known as PLSSVD (Tenenhaus 1998), co-inertia (Dolédec & Chessel 1994, Dray (2014)), and originally as Tucker's interbattery factor analysis (Tucker 1958). Finally, we also introduce a small modification of the "triplet" and "sextuplet" notations as a "quadruplet" and a "septuplet" that indicate the desired rank to be returned by the GSVD or GPLSSVD. For example, if we want only one component from either approach we would indicate the desired rank to return as GSVD$(\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, 1)$ and GPLSSVD$(\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y}, 1)$. Both the GSVD and GPLSSVD in these cases would return only one set of singular vectors, generalized singular vectors, and component scores, and one singular value; for GPLSSVD it would return only one pair of latent variables.

We present Algorithm 2 as a "generalized partial least squares regression" by way of the GPLSSVD sextuplet. We discuss the generalized aspects of the algorithm in more detail in Section ??.

Though we have presented PLS-CA-R as a generalization of PLS-R that accomodates

virutally any data type (by way of CA), the way we formalized PLS-CA-R—in Section **??** and describe its algorithm in Algorithm 2—leads to further variants and broader generalizations; generalizations that span various PLS, CA, and related approaches, several typical PLS algorithms, a variety of optimizations (e.g., canonical correlation), and ridge-like regularization.

2   PLS algorithms

PLS correlation decomposition is a symmetric method where neither data table plays a privileged (or predictive) role, and is performed with a single pass of the singular value decomposition (SVD). PLS canonical decomposition is also symmetric, but makes use of the SVD iteratively and deflates both data tables in each iteration.

There exist three commonly used PLS algorithms: (1) PLS "correlation" decomposition (Krishnan et al. 2011, Bookstein 1994, McIntosh et al. 1996)—also known as PLSSVD (Tenenhaus 1998) or Tucker's Interbattery Factor Analysis (Tucker 1958), (2) PLS "canonical" decompostion (Tenenhaus 1998), and

2.1   GPLS algorithms

In general there exist three primary PLS algorithms: PLS correlation decomposition (Bookstein 1994, Ketterlinus et al. 1989) generally more known in neuroimaging (McIntosh et al. 1996, McIntosh & Lobaugh 2004, Krishnan et al. 2011) which has numerous alternate names such as PLS-SVD and Tucker's interbattery factor analysis (Tucker 1958) amongst others (see also Beaton et al. 2016), PLS regression decomposition (cf. Section **??** and also Algorithm 2) and the PLS canonical decomposition (Tenenhaus 1998, Wegelin et al. 2000), which is a symmetric method with iterative deflation (i.e., it has features of both PLS-C and PLS-R). Given the way in which we formalize PLS-CA-R—as a generalized PLS-R—here we show how PLS-CA-R provides the basis of generalizations of these three algorithms, as well as further optimizations, similar to Borga et al. (1992), Indahl et al. (2009), and de Micheaux et al. (2019) but we do so in a more comprehensive way that incorporates more methods than other unification strategies, and we also do so in a way

that accomodates multiple data types. We refer to the three techniques under the umbrella of generalized partial least squares (GPLS) as GPLS-COR, GPLS-REG, and GPLS-CAN, for the "correlation", "regression", and "canonical" decompositions respectively. GPLS-COR and GPLS-CAN are symmetric decomposition approaches where neither $\mathbf{Z_X}$ nor $\mathbf{Z_Y}$ are privileged. GPLS-REG is an asymmetric decomposition approach where $\mathbf{Z_X}$ is privileged. We present the GPLS-COR, GPLS-REG, and then GPLS-CAN algorithms with their respective optimizations. We do so in the previously mentioned order because GPLS-COR is used as the basis of all three algorithms and GPLS-CAN shares features and concepts with both GPLS-COR and GPLS-REG. For all of these we rely on the formlization of PLS-CA-R we established in Section **??**—specifically for various mixed data types under the χ^2 model (as used in CA).

The GPLS-COR decomposition is the simplest GPLS technique. It requires only a single pass of the SVD—or in our case the GPLSSVD. There are no explicit iterative steps in GPLS-COR. GPLS-COR takes as input the two preprocessed matrices—$\mathbf{Z_X}$ and $\mathbf{Z_Y}$—and their respective row and column weights: $\mathbf{M_X}$ and $\mathbf{W_X}$ for $\mathbf{Z_X}$, and $\mathbf{M_Y}$ and $\mathbf{W_Y}$ for $\mathbf{Z_Y}$, where C is the desired number of components to return. GPLS-COR is shown in Algorithm 1.

Result: Generalized PLS-correlation between $\mathbf{Z_X}$ and $\mathbf{Z_Y}$
Input : $\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y}, C$
Output: $\mathbf{U}, \mathbf{V}, \mathbf{P}, \mathbf{Q}, \mathbf{F}_J, \mathbf{F}_K, \mathbf{L_X}, \mathbf{L_Y}, \mathbf{\Delta}$
GPLSSVD($\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y}, C$)

Algorithm 1: Generalized PLS-correlation algorithm. GPLS-COR is the GPLSSVD and provides the basis of other GPLS techniques. Furthermore, GPLS-COR easily allows for a variety of optmizations for examples canonical correlation, reduced rank regression (redundancy analysis), and even ridge-like regularization.

GPLS-COR maximizes the relationship between $\mathbf{L_X}$ and $\mathbf{L_Y}$ with the orthogonality constraint $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{Y},c'} = 0$ when $c \neq c'$ where $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{Y},c} = \delta_c$ and thus $\mathbf{L_X}^T \mathbf{L_Y} = \mathbf{U}^T \widetilde{\mathbf{Z}}_{\mathbf{X}}^T \widetilde{\mathbf{Z}}_{\mathbf{Y}} \mathbf{V}^T = \mathbf{U}^T \widetilde{\mathbf{Z}}_{\mathbf{R}} \mathbf{V}^T = \mathbf{U}^T \mathbf{U} \mathbf{\Delta} \mathbf{V}^T \mathbf{V}^T = \mathbf{\Delta}$. We can show this with the generalized vectors and weight as $\mathbf{L_X}^T \mathbf{L_Y} = \mathbf{P}^T \mathbf{W_X} \mathbf{Z_X}^T \mathbf{M_X}^{\frac{1}{2}} \mathbf{M_Y}^{\frac{1}{2}} \mathbf{Z_Y} \mathbf{W_Y} \mathbf{Q}^T = \mathbf{P}^T \mathbf{W_X} \mathbf{P} \mathbf{\Delta} \mathbf{Q}^T \mathbf{W_Y} \mathbf{Q} = \mathbf{\Delta}$. Furthermore, GPLS-COR (via GPLSSVD) provides all of the other outputs as previously described in

5

Section **??**. GPLS-COR—which is the GPLSSVD—provides the basis for the other two algorithms: both GPLS-REG and GPLS-CAN make use of GPLS-COR (i.e., the GPLSSVD) with rank 1 solutions iteratively.

The GPLS-REG decomposition builds off of the GPLS-COR algorithm, but does so by way of the GPLSSVD septuplet iteratively for C iterations, with only a rank 1 solution is provided for each use of the GPLSSVD. Then the two data matrices—$\mathbf{Z_X}$ and $\mathbf{Z_Y}$—are deflated for each step asymmetrically, with a privileged $\mathbf{Z_X}$. GPLS-REG is shown in Algorithm 2.

Result: Generalized PLS-regression between $\mathbf{Z_X}$ and $\mathbf{Z_Y}$

Input : $\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y}, C$

Output: $\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}, \widetilde{\mathbf{P}}, \widetilde{\mathbf{Q}}, \widetilde{\mathbf{F}}_J, \widetilde{\mathbf{F}}_K, \mathbf{L_X}, \mathbf{L_Y}, \widetilde{\mathbf{\Delta}}, \mathbf{T_X}, \widehat{\mathbf{U}}, \mathbf{B}$

for $c = 1, \ldots, C$ **do**

    GPLSSVD$(\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y}, 1)$

    $\mathbf{t_X} \leftarrow \boldsymbol{\ell}_\mathbf{X} \times ||\boldsymbol{\ell}_\mathbf{X}||^{-1}$

    $b \leftarrow \boldsymbol{\ell}_\mathbf{Y}^T \mathbf{t_X}$

    $\widehat{\mathbf{u}} \leftarrow (\mathbf{M_X}^{\frac{1}{2}} \mathbf{Z_X} \mathbf{W_X}^{\frac{1}{2}})^T \mathbf{t_X}$

    $\mathbf{Z_X} \leftarrow \mathbf{Z_X} - [\mathbf{M_X}^{-\frac{1}{2}} (\mathbf{t_X} \widehat{\mathbf{u}}^T) \mathbf{W_X}^{-\frac{1}{2}}]$

    $\mathbf{Z_Y} \leftarrow \mathbf{Z_Y} - [\mathbf{M_Y}^{-\frac{1}{2}} (b\mathbf{t_X} \widetilde{\mathbf{v}}^T) \mathbf{W_Y}^{-\frac{1}{2}}]$

end

Algorithm 2: Generalized PLS-regression algorithm. The results of a rank 1 GPLSSVD are used to compute the latent variables and values necessary for deflation of $\mathbf{Z_X}$ and $\mathbf{Z_Y}$. PLS-CA-R is a specific instance of GPLS-REG, which we defined in Section **??**.

GPLS-REG maximizes the relationship between $\mathbf{L_X}$ and $\mathbf{L_Y}$ with the orthogonality constraint $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{X},c'} = 0$ when $c \neq c'$ where $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{Y},c} = \delta_c$ which is also diag$\{\mathbf{L_X}^T \mathbf{L_Y}\} = $ diag$\{\widetilde{\mathbf{\Delta}}\}$.

The GPLS-CAN decomposition builds off of the GPLS-COR algorithm, but does so by way of the GPLSSVD septuplet iteratively for C iterations, with only a rank 1 solution is provided for each use of the GPLSSVD. Then the two data matrices—$\mathbf{Z_X}$ and $\mathbf{Z_Y}$—are deflated for each step symmetrically. GPLS-CAN is shown in Algorithm 3

6

Result: Generalized PLS-canonical between $\mathbf{Z_X}$ and $\mathbf{Z_Y}$

Input : $\mathbf{M_X}$, $\mathbf{Z_X}$, $\mathbf{W_X}$, $\mathbf{M_Y}$, $\mathbf{Z_Y}$, $\mathbf{W_Y}$, C

Output: $\widetilde{\mathbf{U}}$, $\widetilde{\mathbf{V}}$, $\widetilde{\mathbf{P}}$, $\widetilde{\mathbf{Q}}$, $\widetilde{\mathbf{F}}_J$, $\widetilde{\mathbf{F}}_K$, $\mathbf{L_X}$, $\mathbf{L_Y}$, $\widetilde{\boldsymbol{\Delta}}$, $\mathbf{T_X}$, $\mathbf{T_Y}$, $\widehat{\mathbf{U}}$, $\widehat{\mathbf{V}}$

for $c = 1, \ldots, C$ **do**

$\quad$ GPLSSVD($\mathbf{M_X}, \mathbf{Z_X}, \mathbf{W_X}, \mathbf{M_Y}, \mathbf{Z_Y}, \mathbf{W_Y}, 1$)

$\quad$ $\mathbf{t_X} \leftarrow \boldsymbol{\ell}_\mathbf{X} \times ||\boldsymbol{\ell}_\mathbf{X}||^{-1}$

$\quad$ $\mathbf{t_Y} \leftarrow \boldsymbol{\ell}_\mathbf{Y} \times ||\boldsymbol{\ell}_\mathbf{Y}||^{-1}$

$\quad$ $\widehat{\mathbf{u}} \leftarrow (\mathbf{M}_\mathbf{X}^{\frac{1}{2}} \mathbf{Z_X} \mathbf{W}_\mathbf{X}^{\frac{1}{2}})^T \mathbf{t_X}$

$\quad$ $\widehat{\mathbf{v}} \leftarrow (\mathbf{M}_\mathbf{Y}^{\frac{1}{2}} \mathbf{Z_Y} \mathbf{W}_\mathbf{Y}^{\frac{1}{2}})^T \mathbf{t_Y}$

$\quad$ $\mathbf{Z_X} \leftarrow \mathbf{Z_X} - [\mathbf{M}_\mathbf{X}^{-\frac{1}{2}}(\mathbf{t_X}\widehat{\mathbf{u}}^T)\mathbf{W}_\mathbf{X}^{-\frac{1}{2}}]$

$\quad$ $\mathbf{Z_Y} \leftarrow \mathbf{Z_Y} - [\mathbf{M}_\mathbf{Y}^{-\frac{1}{2}}(\mathbf{t_Y}\widehat{\mathbf{v}}^T)\mathbf{W}_\mathbf{Y}^{-\frac{1}{2}}]$

end

Algorithm 3: Generalized PLS-canonical algorithm. The results of a rank 1 GPLSSVD are used to compute the latent variables and values necessary for deflation of $\mathbf{Z_X}$ and $\mathbf{Z_Y}$. Note that the deflation in GPLS-CAN differs from GPLS-REG in Algorithm 2.

GPLS-CAN maximizes the relationship between $\mathbf{L_X}$ and $\mathbf{L_Y}$ with the orthogonality constraints $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{X},c'} = 0$ and $\boldsymbol{\ell}_{\mathbf{Y},c}^T \boldsymbol{\ell}_{\mathbf{Y},c'} = 0$ when $c \neq c'$ where $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{Y},c} = \delta_c$ which is also diag$\{\mathbf{L_X^T L_Y}\} = $ diag$\{\widetilde{\boldsymbol{\Delta}}\}$.

Note that across all three algorithms defined here, that the first component is identical when the same preprocessed data and constraints are provided to the GPLSSVD. In nearly all cases, subsequent components across the three algorithms differ, but also generally they do not differ substantially. The similarities can be traced back to the common maximization of $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{Y},c} = \delta_c$, where the differences can be traced back to the specific orthogonality optimizations when $c \neq c'$ where: (1) GPLS-COR in Algorithm 1 is $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{Y},c'} = 0$, (2) GPLS-REG in Algorithm 2 is $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{X},c'} = 0$, and (3) GPLS-CAN in Algorithm 3 is both $\boldsymbol{\ell}_{\mathbf{X},c}^T \boldsymbol{\ell}_{\mathbf{X},c'} = 0$ and $\boldsymbol{\ell}_{\mathbf{Y},c}^T \boldsymbol{\ell}_{\mathbf{Y},c'} = 0$.

2.2   GPLS optimizations and further generalizations

From the GPLS perspective, we can better unify the wide variety of approaches with similar goals but variations of metrics, transformations, and optimizations that often appear under a wide variety of names (e.g., PLS, CCA, interbattery factor analysis, co-inertia analysis, canonical variates, PLS-CA, and so on; see Abdi et al. (2017)). The way we defined the GPLS algorithms—in particular with the constraints applied to the rows and columns of each data matrix—leads to numerous further generalizations.

For simplicity, let us first focus on Algorithm 1, and assume that $\mathbf{X}$ and $\mathbf{Y}$ are continuous data, where $\mathbf{Z_X}$ and $\mathbf{Z_Y}$ are column-wise centered and/or scaled versions of $\mathbf{X}$ and $\mathbf{Y}$. Though we have established Algorithm 1 as GPLS-COR—and more generally as the GPLSSVD—we can obtain the results of three of the most common "two-table" techniques: PLS correlation (PLSC), canonical correlation analysis (CCA), and redundancy analysis (RDA, a.k.a., reduced rank regression [RRR]). Standard PLSC is performed as GPLSSVD$(\mathbf{I}, \mathbf{Z_X}, \mathbf{I}, \mathbf{I}, \mathbf{Z_Y}, \mathbf{I})$, CCA is performed as GPLSSVD$(\mathbf{I}, \mathbf{Z_X}, (\mathbf{Z_X^T Z_X})^{-1}, \mathbf{I}, \mathbf{Z_Y}, (\mathbf{Z_Y^T Z_Y})^{-1})$, and RDA—where $\mathbf{X}$ is privileged—is performed as GPLSSVD$(\mathbf{I}, \mathbf{Z_X}, (\mathbf{Z_X^T Z_X})^{-1}, \mathbf{I}, \mathbf{Z_Y}, \mathbf{I})$. Furthermore, these three variants—PLSC, CCA, and RDA/RRR—also generalize discriminant analyses under different optimizations so long as $\mathbf{X}$ is a dummy-coded or complete disjunctive matrix to assign each observation (row) to a specific group or category (columns).

Most importantly, because of the ways we formalized the GPLS algorithms—see also Section **??**—and the variety of ways to suitably transform data (e.g., the various coding schemes we have shown) allow application of PLS-CA-R and GPLS algorithms on a variety of different problems or models such as log or power transformations and alternate choices for weights (see Eq. **??**) or models (see Eq. **??**). That means that the GPLS algorithms further generalize many approaches, especially the numerous variants of CA. Generally in the cases of strictly positive data, there may be a need to preprocess data within the family of power transformations for CA (Greenacre 2009) or alternate distance metrics such as Hellinger distance (Rao 1995, Escofier 1978). Finally, with the choices of weights can change, as they do for Hellinger CA, and for the variations of "non-symmetrical CA" (D'Ambra & Lauro 1992, Kroonenberg & Lombardo 1999, Takane et al. 1991), where both types of variants require one set of weights as $\mathbf{I}$ (akin to RDA/RRR-type optimizations with CA/χ^2 models across any of the GPLS algorithms).

2.3 Ridge-like regularization

It is also possible to apply ridge-like regularization to PLS-CA regression, correlation, and canonical decompositions. We show two possible strategies for ridge-like regularization under the data/model assumptions and the preprocessing we established in Section **??**.

The first approach is based on Takane's regularized multiple CA (Takane & Hwang 2006) and regularized nonsymmetric CA (Takane & Jung 2009). To do so, it is convenient to slightly reformulate PLS-CA-R, but we still require $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{O_X}$, $\mathbf{O_Y}$, $\mathbf{E_X}$, and $\mathbf{E_Y}$ as defined in Section **??**. First we re-define $\mathbf{Z_X} = (\mathbf{O_X} - \mathbf{E_X}) \times (\mathbf{1}^T \mathbf{X} \mathbf{1})$ and $\mathbf{Z_Y} = (\mathbf{O_Y} - \mathbf{E_Y}) \times (\mathbf{1}^T \mathbf{Y} \mathbf{1})$; which are the same as in Eq. **??** except scaled by the grand sum of its respective source data matrix. Next we define the following additional matrices: $\mathbf{D}_{\mathbf{X},I} = \mathrm{diag}\{\mathbf{X}\mathbf{1}\}$, and $\mathbf{D}_{\mathbf{Y},I} = \mathrm{diag}\{\mathbf{Y}\mathbf{1}\}$ which are diagonal matrices of the row sums of $\mathbf{X}$ and $\mathbf{Y}$, respectively and $\mathbf{D}_{\mathbf{X},J} = \mathrm{diag}\{\mathbf{1}^T\mathbf{X}\}$, and $\mathbf{D}_{\mathbf{Y},K} = \mathrm{diag}\{\mathbf{1}^T\mathbf{Y}\}$ which are the column sums of $\mathbf{X}$ and $\mathbf{Y}$. Then PLS-CA correlation, regression, and canonical decompositions replace the GPLSSVD step in Algorithms 1, 2, 3 with GPLSSVD($\mathbf{D}_{\mathbf{X},I}^{-1}, \mathbf{Z}_{\mathbf{X}}^{T}, \mathbf{D}_{\mathbf{X},J}^{-1}, \mathbf{D}_{\mathbf{Y},I}^{-1}, \mathbf{Z}_{\mathbf{Y}}^{T}, \mathbf{D}_{\mathbf{Y},K}^{-1}$). The only differences between this Takane-ian reformulation and what we originally established is that the generalized singular vectors ($\mathbf{P}$ and $\mathbf{Q}$) and the component scores ($\mathbf{F_J}$ and $\mathbf{F_K}$)

9

differ by constant scaling factors (which are the sums of $\mathbf{X}$ and $\mathbf{Y}$ for their respective scores).

We can regularize PLS-CA-R in the same way as Takane's RMCA. We require (1) a ridge parameter which we refer to as λ and (2) variants of $\mathbf{D}_{\mathbf{X},I}$, $\mathbf{D}_{\mathbf{X},J}$, $\mathbf{D}_{\mathbf{Y},I}$, and $\mathbf{D}_{\mathbf{Y},K}$ that we refer to as $\mathbb{D}_{\mathbf{X},I} = \mathbf{D}_{\mathbf{X},I} + [\lambda \times (\mathbf{Z}_{\mathbf{X}}\mathbf{Z}_{\mathbf{X}}^{T})^{+}]$, $\mathbb{D}_{\mathbf{Y},I} = \mathbf{D}_{\mathbf{Y},I} + [\lambda \times (\mathbf{Z}_{\mathbf{Y}}\mathbf{Z}_{\mathbf{Y}}^{T})^{+}]$, $\mathbb{D}_{\mathbf{X},J} = \mathbf{D}_{\mathbf{X},J} + [\lambda \times \mathbf{Z}_{\mathbf{X}}^{T}(\mathbf{Z}_{\mathbf{X}}\mathbf{Z}_{\mathbf{X}}^{T})^{+}\mathbf{Z}_{\mathbf{X}}]$, and $\mathbb{D}_{\mathbf{Y},K} = \mathbf{D}_{\mathbf{Y},K} + [\lambda \times \mathbf{Z}_{\mathbf{Y}}^{T}(\mathbf{Z}_{\mathbf{Y}}\mathbf{Z}_{\mathbf{Y}}^{T})^{+}\mathbf{Z}_{\mathbf{Y}}]$. When $\lambda = 0$ then $\mathbb{D}_{\mathbf{X},I} = \mathbf{D}_{\mathbf{X},I}$, $\mathbb{D}_{\mathbf{Y},I} = \mathbf{D}_{\mathbf{Y},I}$, $\mathbb{D}_{\mathbf{X},J} = \mathbf{D}_{\mathbf{X},J}$, $\mathbb{D}_{\mathbf{Y},K} = \mathbf{D}_{\mathbf{Y},K}$. We obtain regularized forms of PLS-CA for the correlation, regression, and canonical decompositions if we simply replace the GPLSSVD step in each algorithm with GPLSSVD($\mathbb{D}_{\mathbf{X},I}^{-1}, \mathbf{Z}_{\mathbf{X}}^{T}, \mathbb{D}_{\mathbf{X},J}^{-1}, \mathbb{D}_{\mathbf{Y},I}^{-1}, \mathbf{Z}_{\mathbf{Y}}^{T}, \mathbb{D}_{\mathbf{Y},K}^{-1}$). As per Takane's recommendation (Takane & Hwang 2006), λ could be any positive value, though integers in the range from 1 to 20 provide sufficient regularization, especially as λ increases.

But the Takane-ian approach may not be feasible when I, J, and/or K are particularly large because the various crossproduct and projection matrices require a large amount of memory and/or computational expense. So we now introduce a "truncated" version of the Takane regularization which is more computationally efficient, and analogous to the regularization procedure of Allen (Allen 2013, Allen et al. 2014). We re-define $\mathbb{D}_{\mathbf{X},I} = \mathbf{D}_{\mathbf{X},I} + (\lambda \times \mathbf{I})$ and $\mathbb{D}_{\mathbf{Y},I} = \mathbf{D}_{\mathbf{Y},I} + (\lambda \times \mathbf{I})$ and then also $\mathbb{D}_{\mathbf{X},J} = \mathbf{D}_{\mathbf{X},J} + (\lambda \times \mathbf{I})$ and $\mathbb{D}_{\mathbf{Y},K} = \mathbf{D}_{\mathbf{Y},K} + (\lambda \times \mathbf{I})$ where $\mathbf{I}$ are identity matrices (1s on the diagonal) of appropriate size. Like in the previous formulation, we replace the values we have in the GPLSSVD step where GPLSSVD($\mathbb{D}_{\mathbf{X},I}^{-1}, \mathbf{Z}_{\mathbf{X}}^{T}, \mathbb{D}_{\mathbf{X},J}^{-1}, \mathbb{D}_{\mathbf{Y},I}^{-1}, \mathbf{Z}_{\mathbf{Y}}^{T}, \mathbb{D}_{\mathbf{Y},K}^{-1}$); and in this particular case, the constraint matrices are all diagonal matrices, which allows for a lower memory footprint and less computational burden.

Finally, we have two concluding remarks on ridge-like regularization. The first point is that the more simplified Takane/Allen hybrid approach to ridge-like regularization also applies much more generally to virtually any technique for the SVD or GPLSSVD. For any approach, we only require some inflation factor (λ) for the constraints alon the diagonals. The second point is that while we have presented ridge-like regularization with a single λ it is entirely possible to use different λs for each set of constraints. But even though it is possible, we do not recommend this approach, as it would require a complex grid

search over all the various λ parameters. Alternatively, if multiple λs were used, one could minimize the number of parameters to search and set some of the λs to 0 and, for example, use only one or two λ values instead of four possible λ values.

3   Fin

While that was the bulk of this work, our secondary goal was to further generalize the PLS-CA approach and to better unify many methods under a simpler framework, specifically by way of the GPLSSVD and our three GPLS algorithms. Thus our generalizations—first established in Section ??, and expanded upon in Discussion—accomodate: almost any data type, various metrics (e.g., Hellinger distance), various optimizations (e.g., PLS, CCA, or RDA type optmizations), and even two strategies for ridge-like regularization. We have foregone any discussions of inference, stability, and resampling for PLS-CA-R because, as a generalization of PLS-R, many inference and stability approaches still apply—such as feature selection or sparsification (Sutton et al. 2018), additional regularization or sparsification approaches (Le Floch et al. 2012, Guillemot et al. 2019, Tenenhaus et al. 2014, Tenenhaus & Tenenhaus 2011), cross-validation (Wold et al. 1987, Rodríguez-Pérez et al. 2018, Kvalheim et al. 2019, Abdi 2010), permutation (Berry et al. 2011), various bootstrap (Efron 1979, Chernick 2008) approaches (Abdi 2010, Takane & Jung 2009) or tests (McIntosh & Lobaugh 2004, Krishnan et al. 2011), and other frameworks such as split-half resampling (Strother et al. 2002, Kovacevic et al. 2013, Strother et al. 2004)—and are easily adapted for the PLS-CA-R and GPLS frameworks.

References

Abdi, H. (2010), 'Partial least squares regression and projection on latent structure regression (PLS Regression)', *Wiley Interdisciplinary Reviews: Computational Statistics* **2**(1), 97–106.
   URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.51*

Abdi, H., Guillemot, V., Eslami, A. & Beaton, D. (2017), 'Canonical correlation analysis', *Encyclopedia of Social Network Analysis and Mining* pp. 1–16. Springer.

Allen, G. I. (2013), 'Sparse and Functional Principal Components Analysis', *arXiv:1309.2895 [stat]* . arXiv: 1309.2895.
URL: *http://arxiv.org/abs/1309.2895*

Allen, G. I., Grosenick, L. & Taylor, J. (2014), 'A Generalized Least-Square Matrix Decomposition', *Journal of the American Statistical Association* **109**(505), 145–159.
URL: *http://dx.doi.org/10.1080/01621459.2013.852978*

Beaton, D., Dunlop, J., Abdi, H. & Alzheimer's Disease Neuroimaging Initiative (2016), 'Partial Least Squares Correspondence Analysis: A Framework to Simultaneously Analyze Behavioral and Genetic Data', *Psychological Methods* **21**(4), 621–651.

Berry, K. J., Johnston, J. E. & Mielke, P. W. (2011), 'Permutation methods', *Wiley Interdisciplinary Reviews: Computational Statistics* **3**, 527–542.
URL: *http://wires.wiley.com/WileyCDA/WiresArticle/wisId-WICS177.html*

Bookstein, F. L. (1994), 'Partial least squares: A dose-response model for measurement in the behavioral and brain sciences.', *Psycoloquy* .

Borga, M., Landelius, T. & Knutsson, H. (1992), *A Unified Approach to PCA, PLS, MLR and CCA*.

Chernick, M. (2008), *Bootstrap methods: A guide for practitioners and researchers*, Vol. 619, Wiley-Interscience.

D'Ambra, L. & Lauro, N. C. (1992), 'Non symmetrical exploratory data analysis', *Statistica Applicata* **4**(4), 511–529.

de Micheaux, P. L., Liquet, B., Sutton, M. et al. (2019), 'Pls for big data: A unified parallel algorithm for regularised group pls', *Statistics Surveys* **13**, 119–149.

Dolédec, S. & Chessel, D. (1994), 'Co-inertia analysis: an alternative method for studying species-environment relationships', *Freshwater Biology* **31**, 177–194.

Dray, S. (2014), Analysing a pair of tables: Coinertia analysis and duality diagrams, *in* J. Blasius & G. M, eds, 'Visualization and Verbalization of Data', CRC Press, Boca Raton, pp. 289–300.

Efron, B. (1979), 'Bootstrap Methods: Another Look at the Jackknife', *The Annals of Statistics* **7**(1), 1–26. ArticleType: primary_article / Full publication date: Jan., 1979 / Copyright © 1979 Institute of Mathematical Statistics.
URL: *http://www.jstor.org.libproxy.utdallas.edu/stable/2958830*

Escofier, B. (1978), 'Analyse factorielle et distances répondant au principe d'équivalence distributionnelle', *Revue de statistique appliquée* **26**(4), 29–37.

Greenacre, M. (2009), 'Power transformations in correspondence analysis', *Computational Statistics & Data Analysis* **53**(8), 3107–3116.

Guillemot, V., Beaton, D., Gloaguen, A., Löfstedt, T., Levine, B., Raymond, N., Tenenhaus, A. & Abdi, H. (2019), 'A constrained singular value decomposition method that integrates sparsity and orthogonality', *PloS one* **14**(3), e0211463.

Indahl, U. G., Liland, K. H. & Næs, T. (2009), 'Canonical partial least squares—a unified pls approach to classification and regression problems', *Journal of Chemometrics: A Journal of the Chemometrics Society* **23**(9), 495–504.

Ketterlinus, R. D., Bookstein, F. L., Sampson, P. D. & Lamb, M. E. (1989), 'Partial least squares analysis in developmental psychopathology', *Development and Psychopathology* **1**(4), 351–371.

Kovacevic, N., Abdi, H., Beaton, D. & McIntosh, A. R. (2013), Revisiting pls resampling: comparing significance versus reliability across range of simulations, *in* 'New Perspectives in Partial Least Squares and Related Methods', Springer, pp. 159–170.

Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. (2011), 'Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review', *NeuroImage* **56**(2), 455 – 475. Multivariate Decoding and Brain Reading.
URL: *http://www.sciencedirect.com/science/article/pii/S1053811910010074*

Kroonenberg, P. M. & Lombardo, R. (1999), 'Nonsymmetric correspondence analysis: A tool for analysing contingency tableswith a dependence structure', *Multivariate Behavioral Research* **34**(3), 367–396.

Kvalheim, O. M., Grung, B. & Rajalahti, T. (2019), 'Number of components and prediction error in partial least squares regression determined by Monte Carlo resampling strategies', *Chemometrics and Intelligent Laboratory Systems* .
URL: *http://www.sciencedirect.com/science/article/pii/S0169743918307056*

Le Floch, E., Guillemot, V., Frouin, V., Pinel, P., Lalanne, C., Trinchera, L., Tenenhaus, A., Moreno, A., Zilbovicius, M., Bourgeron, T., Dehaene, S., Thirion, B., Poline, J.-B. & Duchesnay, E. (2012), 'Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse Partial Least Squares', *NeuroImage* **63**(1), 11–24.
URL: *http://www.sciencedirect.com/science/article/pii/S1053811912006775*

McIntosh, A., Bookstein, F., Haxby, J. & Grady, C. (1996), 'Spatial Pattern Analysis of Functional Brain Images Using Partial Least Squares', *NeuroImage* **3**(3), 143–157.
URL: *http://www.sciencedirect.com/science/article/pii/S1053811996900166*

McIntosh, A. R. & Lobaugh, N. J. (2004), 'Partial least squares analysis of neuroimaging data: applications and advances', *Neuroimage* **23**, S250–S263.

Rao, C. R. (1995), 'A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance', *Qüestiió: quaderns d'estadística i investigació operativa* **19**(1).

Rodríguez-Pérez, R., Fernández, L. & Marco, S. (2018), 'Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study', *Analytical and Bioanalytical Chemistry* **410**(23), 5981–5992.
URL: *https://doi.org/10.1007/s00216-018-1217-1*

Strother, S. C., Anderson, J., Hansen, L. K., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S. & Rottenberg, D. (2002), 'The Quantitative Evaluation of

Functional Neuroimaging Experiments: The NPAIRS Data Analysis Framework', *Neu-roImage* **15**(4), 747–771.
URL: *http://www.sciencedirect.com/science/article/pii/S1053811901910341*

Strother, S., La Conte, S., Hansen, L. K., Anderson, J., Zhang, J., Pulapura, S. & Rotten-berg, D. (2004), 'Optimizing the fmri data-processing pipeline using prediction and repro-ducibility performance metrics: I. a preliminary group analysis', *Neuroimage* **23**, S196–S207.

Sutton, M., Thiébaut, R. & Liquet, B. (2018), 'Sparse partial least squares with group and subgroup structure', *Statistics in Medicine* **37**(23), 3338–3356.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7821*

Takane, Y. & Hwang, H. (2006), 'Regularized multiple correspondence analysis', *Multiple correspondence analysis and related methods* pp. 259–279.

Takane, Y. & Jung, S. (2009), 'Regularized nonsymmetric correspondence analysis', *Com-putational Statistics & Data Analysis* **53**(8), 3159–3170.
URL: *http://www.sciencedirect.com/science/article/pii/S0167947308004313*

Takane, Y., Yanai, H. & Mayekawa, S. (1991), 'Relationships among several methods of linearly constrained correspondence analysis', *Psychometrika* **56**(4), 667–684.

Tenenhaus, A., Philippe, C., Guillemot, V., Cao, K.-A. L., Grill, J. & Frouin, V. (2014), 'Variable selection for generalized canonical correlation analysis', *Biostatistics* p. kxu001.
URL: *http://biostatistics.oxfordjournals.org/content/early/2014/02/17/biostatistics.kxu001*

Tenenhaus, A. & Tenenhaus, M. (2011), 'Regularized Generalized Canonical Correlation Analysis', *Psychometrika* **76**(2), 257–284.
URL: *http://www.springerlink.com/content/8r17621w56k3025w/abstract/*

Tenenhaus, M. (1998), *La régression PLS: théorie et pratique*, Editions TECHNIP. Google-Books-ID: OesjK2KZhsAC.

Tucker, L. R. (1958), 'An inter-battery method of factor analysis', *Psychometrika* **23**(2), 111–136.
   URL: *http://www.springerlink.com/content/u74149nr2h0124l0/*

Wegelin, J. A. et al. (2000), 'A survey of partial least squares (pls) methods, with emphasis on the two-block case', *University of Washington, Tech. Rep* .

Wold, S., Esbensen, K. & Geladi, P. (1987), 'Principal component analysis', *Chemometrics and Intelligent Laboratory Systems* **2**(1), 37–52.
   URL: *http://www.sciencedirect.com/science/article/pii/0169743987800849*