, 000–000

Hybrid Partial Least Squares Regression with Multiple Functional and Scalar Predictors with Application to Renal Imaging Data

Jongmin Mun, University of Southern California, jongmin.mun@marshall.usc.edu

Jeong Hoon Jang, University of Texas Medical Branch, jejang@utmb.edu

Summary: Motivated by renal imaging data with accompanying side information, we propose a hybrid partial least squares (PLS) regression method to capture cross-modality correlations with both functional and scalar predictors. At the population level, we extend the nonlinear iterative PLS (NIPALS) algorithm to a hybrid Hilbert space, unifying functional and scalar predictors into a single vector and iteratively maximizing the empirical covariance. We establish orthogonality of the PLS scores and L_2 convergence of the PLS fitted value to the true response. We then propose a practical sample-level algorithm incorporating regularization to promote smoothness of regression coefficients and PLS directions. The resulting low-dimensional representations effectively capture within- and between-modality variation and predictor-response correlation, while preserving key PLS properties such as orthonormality of PLS directions and orthogonality of PLS scores. Moreover, the procedure is computationally efficient, requiring only the solution of linear systems at each iteration.

Key words: Dimension reduction; Functional data analysis; Multiple data modalities; Multivariate data analysis; Multivariate functional data; Partial least square

1. Introduction

Biomedical studies often collect multi-modal data. Our running example is the Emory University renal study (Chang et al., 2020; Jang, 2021), which records two renogram curves (functional data) along with side information (Euclidean vector) for each kidney. Such distinct yet related physiological signals can be represented as functional- and scalar-valued i.i.d. covariates in a linear regression model:

$$Y_i = \boldsymbol{\beta}^\top \mathbf{Z}_i + \sum_{k=1}^{K} \int \beta_k(t) X_{ik}(t) dt + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where Y_i is a scalar response, $X_{i1}(t), \ldots, X_{iK}(t) \in \mathbb{L}^2[0,1]$ are functional predictors, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$ is a Euclidean vector covariate, and ϵ_i denotes observational noise. For notational convenience, we assume throughout that the responses and predictors are centered, so the intercept term can be omitted. The key challenge here is the strong correlations between functional and scalar predictors, in addition to the ill-posedness of the infinite-dimensional slope functions $\beta_k(t)$ and the high dimensionality of the predictors $(K + p)$. This paper addresses all three issues in a unified manner by introducing a hybrid partial least squares (PLS) regression framework defined on a novel hybrid Hilbert space.

Previous works and limitations. Basis expansion with regularization of roughness addresses the ill-posedness of functional data by exploiting their smoothness (Cardot et al., 2003; Aguilera et al., 2016; Cai and Hall, 2006; Zhao et al., 2012). High dimensionality can be addressed by using derived inputs, with PCA regression as a straightforward example. We can apply PCA separately to functional (e.g., Happ and Greven 2018) and scalar predictors, followed by multivariate regression on the combined scores. While well studied for functional predictors alone (Hall and Horowitz, 2007; Reiss and Ogden, 2007; Febrero-Bande et al., 2017), PCA regression may fail to capture the core regression relationship since the derived inputs are not guided by the response. Partial least squares (PLS) regression is a powerful alternative that constructs orthogonal latent components from predictors to

maximize covariance with the response. For functional predictors, Preda and Saporta (2005) introduced PLS regression for a single predictor, with subsequent extensions incorporating basis approximations and roughness penalties (Reiss and Ogden, 2007; Aguilera et al., 2010, 2016), and multiple functional predictors (Beyaztas and Lin Shang, 2022). Theoretical and computational advances were developed by Delaigle and Hall (2012) and Saricam et al. (2022), respectively; see Febrero-Bande et al. (2017) for a review. These approaches ignore correlations between functional and scalar components, leading to multicollinearity and unreliable prediction. Multimodal correlations have mainly been explored in unsupervised settings, such as graphical models, precision matrix estimation, and joint PCA (Kolar et al., 2014; Geng et al., 2020; Jang, 2021). However, these methods are response-agnostic and may miss outcome-relevant correlations.

Our contributions. To address these gaps, we propose a hybrid PLS regression framework that unifies basis expansion, PLS, and functional-scalar correlations. Treating functional and Euclidean components as a single vector in a Hilbert space with a suitable inner product, we iteratively extract joint directions that maximize covariance with the response under a unit-norm constraint. The framework handles dense or irregular functional data, supports regularization to reduce overfitting, and comes with theoretical guarantees.

2. Hybrid Hilbert Space and Population Version PLS Algorithm

Our framework leverages that the nonlinear iterative PLS (NIPALS; Wold, 1975) algorithm, a standard PLS fitting method, operates solely via inner-product space operations. We define a Hilbert space for hybrid predictors with a complete inner product to extend the algorithm.

THEOREM 1 (Hybrid space): *An element* $h \in \mathbb{H} := (\mathbb{L}^2[0,1])^K \times \mathbb{R}^p$ *is an ordered tuple* $h = (f_1, \ldots, f_K, \mathbf{u})$*, where* $f_k \in \mathbb{L}^2[0,1]$ *for* $k = 1, \ldots, K$ *and* $\mathbf{u} \in \mathbb{R}^p$*. We define an inner*

product on $\mathbb{H}$ for any two elements $h_1 = (f_1, \ldots, f_K, \mathbf{u})$ and $h_2 = (g_1, \ldots, g_K, \mathbf{v})$ as:

$$\langle h_1, h_2 \rangle_{\mathbb{H}} := \sum_{k=1}^{K} \int_0^1 f_k(t) g_k(t)\, dt + \mathbf{u}^\top \mathbf{v}. \tag{2}$$

The inner product induces a norm $\|\cdot\|_{\mathbb{H}}$ on the space, defined as $\|h\|_{\mathbb{H}} := \langle h, h \rangle_{\mathbb{H}}^{1/2}$, and a corresponding metric $d(h_1, h_2) = \|h_1 - h_2\|_{\mathbb{H}}$. Then the hybrid inner product space $\mathbb{H}$ is a separable Hilbert space.

Definition 1 (Hybrid Predictor): For the Hilbert space $\mathbb{H}$ defined in Definition 1, The Borel σ-field on $\mathbb{H}$, denoted $\mathfrak{B}(\mathbb{H})$, is the smallest σ-field containing the class $\mathfrak{M}$ of all sets of the form $\{q \in \mathbb{H} \mid \langle q, h \rangle \in O\}$, for any $h \in \mathbb{H}$ and any open subset $O \subseteq \mathbb{R}$ (details can be found in Theorem 7.1.1 of Hsing and Eubank 2015). A hybrid predictor $W_i = (X_{i1}(t), \ldots, X_{iK}(t), \mathbf{Z}_i)$ is a measurable mapping from a probability space $(\Omega, \mathfrak{F}, P)$ into $\big(\mathbb{H}, \mathfrak{B}(\mathbb{H})\big)$.

Then the joint regression model (1) can be concisely written as

$$Y_i = \langle \beta, W_i \rangle_{\mathbb{H}} + \epsilon_i, \text{ where } \beta := \big(\beta_1(t), \ldots, \beta_K(t), \boldsymbol{\beta}\big) \in \mathbb{H}. \tag{3}$$

We now derive the population version of NIPALS in the hybrid space. At the population level, where the random objects are fully observable, we may drop the sample index i and simply work with Y and $W = (X_1, \ldots, X_K, Z_1, \ldots, Z_p)$. Throughout this section we assume $\mathbb{E}[Y] = 0 \in \mathbb{R}$ and $\mathbb{E}[W] = 0 \in \mathbb{H}$.

The core optimization problem in PLS is formulated in terms of maximizing the squared cross-covariance. At the population level, this quantity is naturally characterized by cross-covariance operators, which we define next.

Definition 2: The covariance operator of the hybrid predictor W is denoted as $\Sigma_W := \mathbb{E}[W \otimes W]$, so that for $u, v \in \mathbb{H}$, we have $\Sigma_W u = \mathbb{E}[\langle W, u \rangle_{\mathbb{H}} W]$, $\langle \Sigma_W u, v \rangle_{\mathbb{H}} = \mathbb{E}[\langle W, u \rangle_{\mathbb{H}} \langle W, v \rangle_{\mathbb{H}}]$. The cross-covariances between the response Y and the predictors are

$$\sigma_{YX} := \big(\mathbb{E}[YX_1], \ldots, \mathbb{E}[YX_K]\big) \in \mathcal{F}^K, \quad \sigma_{YZ} := \big(\mathbb{E}[YZ_1], \ldots, \mathbb{E}[YZ_p]\big)^\top \in \mathbb{R}^p.$$

The cross-covariance between Y and the hybrid predictor W is then

$$\Sigma_{YW} := \mathbb{E}[YW] = \big(\sigma_{YX},\, \sigma_{YZ}\big) \in \mathbb{H}.$$

Now we explain the NIPALS algorithm at the population level. Let $W^{[1]} := W$ and $Y^{[1]} := Y$. For any integer $l \geq 2$, $W^{[l]}$ and $Y^{[l]}$ are residuals (see line 7,8 of Algorithm 1). The PLS direction maximizes the squared covariance:

Definition 3: Hybrid PLS direction at the l-th step, denoted $\xi^{[l]}$, is defined as

$$\xi^{[l]} := \arg\max_{h\in\mathbb{H}} \mathrm{Cov}^2(\langle W, h\rangle_{\mathbb{H}}, Y),\ \ s.t.\ \|h\|_{\mathbb{H}} = 1, \langle h, \Sigma_W\, \xi^{[j]}\rangle_{\mathbb{H}} = 0,\ j < l. \tag{4}$$

This problem can be solved by an iterative algorithm, and yields orthogonal scores:

Theorem 2 (Properties of population PLS): *The solution of optimization problem* (4) *of Definition 3 is equivalent to that of population NIPALS (Algorithm 1). For every $s \geq 1$ and every $j \leq s-1$, we have $\mathbb{E}[\rho^{[j]}\rho^{[s]}] = 0$ and $\mathbb{E}[Y^{[s]}\rho^{[j]}] = 0$.*

Algorithm 1 Population NIPALS

1: $(W^{[1]}, Y^{[1]} \leftarrow (W, Y)$

2: **for** $l = 1, 2, \ldots, L$ **do**

3: $\quad \xi^{[l]} \leftarrow \arg\max_{h\in\mathbb{H}} \mathrm{Cov}^2(\langle W^{[l]}, h\rangle_{\mathbb{H}}, Y^{[l]})\ s.t.\ \|h\|_{\mathbb{H}} = 1,$ ▷ PLS direction

4: $\quad \rho^{[l]} \leftarrow \langle \xi^{[l]}, W^{[l]}\rangle$ ▷ PLS score

5: $\quad \delta^{[l]} \leftarrow \frac{1}{\mathbb{E}[(\rho^{[l]})^2]}\mathbb{E}[W^{[l]}\rho^{[l]}]$ ▷ Linear regression of $W^{[l]}$ on $\rho^{[l]}$

6: $\quad \nu^{[l]} \leftarrow \frac{1}{\mathbb{E}[(\rho^{[l]})^2]}\mathbb{E}[Y^{[l]}\rho^{[l]}]$ ▷ Linear regression of $Y^{[l]}$ on $\rho^{[l]}$

7: $\quad W^{[l+1]} \leftarrow W^{[l]} - \rho^{[l]}\,\delta^{[l]}$ ▷ Residualized predictor

8: $\quad Y^{[l]} \leftarrow Y^{[l-1]} - \nu^{[l-1]}\,\rho^{[l-1]}$ ▷ Residualized response

9: **Output:** PLS directions $\xi^{[1]}, \ldots, \xi^{[L]}$

We show in Theorem 3 that problem (4) and Algorithm 1 are well-defined under mild conditions by proving line 3 is well-defined. To this end, Lemmas 1 and 2 define the relevant operators, omitting the iteration index for residualized predictors, responses, and operators.

LEMMA 1: *Define the operator* $\mathcal{C}_{YW}$ *such that for any* $h = (f_1, \ldots, f_K, \mathbf{v}) \in \mathbb{H}$,

$$\mathcal{C}_{YW} h := \mathbb{E}\left[\langle W, h\rangle_{\mathbb{H}} Y\right] = \langle \Sigma_{YW}, h\rangle_{\mathbb{H}} \in \mathbb{R}.$$

The operator $\mathcal{C}_{YW}$ *is a compact operator if there exist finite constants* Q_1 *and* Q_2 *such that*

$$\max_{k=1,\ldots,K} \sup_{t\in[0,1]} \mathbb{E}[YX_k](t)^2 < Q_1 \quad \text{and} \quad \max_{r=1,\ldots,p} \mathbb{E}[YZ_r]^2 < Q_2. \tag{5}$$

LEMMA 2: *Define* $\mathcal{U} := \mathcal{C}_{WY} \circ \mathcal{C}_{YW} : \mathbb{H} \to \mathbb{H}$ *as an operator such that, for* $h \in \mathbb{H}$,

$$\mathcal{U}h = \mathcal{C}_{WY}(\mathcal{C}_{YW}h) = \mathcal{C}_{YW}(\langle \Sigma_{YW}, h\rangle_{\mathbb{H}}) = \Sigma_{YW} \langle \Sigma_{YW}, h\rangle_{\mathbb{H}} = (\Sigma_{YW} \otimes \Sigma_{YW})h.$$

In other words, $\mathcal{U} = \Sigma_{YW} \otimes \Sigma_{YW}$. *Then* $\mathcal{U}$ *is a self-adjoint and positive-semidefinite operator. Under the conditions of Lemma 1,* $\mathcal{U}$ *is a compact operator.*

By the Hilbert-Schmidt theorem (e.g., Theorem 4.2.4 in Hsing and Eubank, 2015), Lemma 2 guarantees the existence of a complete orthonormal system of eigenfunctions $\{\xi_{(u)}\}_{u\in\mathbb{N}}$ of $\mathcal{U}$ in $\mathbb{H}$ such that $\mathcal{U}\xi_{(u)} = \kappa_{(u)}\xi_{(u)}$, where $\{\kappa_{(u)}\}_{u\in\mathbb{N}}$ are the corresponding sequence of eigenvalues that goes to zero as $u \to \infty$, that is, $\kappa_{(1)} \geq \kappa_{(2)} \geq \cdots \geq 0$. The following theorem presents ensures that line 3 of Algorithm 1 is well-defined under mild conditions.

THEOREM 3 (Tucker's criterion): *Under the conditions of Lemma 1, the constrained maximum* $\max_{\|\xi\|_{\mathbb{H}}=1} \mathrm{Cov}^2\left(\langle W, \xi\rangle_{\mathbb{H}}, Y\right)$ *is attained by the eigenfunction associated with the largest eigenvalue of the operator* $\mathcal{U}$.

Finally, the following theorem states that the fitted values from hybrid PLS converge to those of ordinary linear regression in the mean-squared sense. This shows that hybrid PLS, as a dimension-reduction technique, can effectively capture the predictive power of a full linear model when a sufficient number of components is used.

THEOREM 4: *Let* $Y_{\mathrm{PLS},L} = \sum_{l=1}^{L} \nu^{[l]}\rho^{[l]}$ *be the Hybrid PLS fitted value with* L *components. Then, the Hybrid PLS fitted value converges to* Y *in the mean-squared sense:*

$$\lim_{L\to\infty} \mathbb{E}\left[\|Y_{\mathrm{PLS},L} - Y\|_2^2\right] = 0.$$

3. Proposed PLS Algorithm

We propose a sample version of hybrid NIPALS (Algorithm 1) that efficiently computes PLS components and scores for multiple dense or irregular functional predictors and scalar predictors. Our algorithm incorporates regularization to exploit structural relationships, prevent overfitting, and improve interpretability. Each iteration involves two subroutines: regularized component estimation (Section 3.2.1) and residualization (Section 3.2.2) After iteration terminates the hybrid regression coefficient is estimated (Section 3.3).

3.1 *Preliminary step: finite-basis approximation*

Let $\{b_m(t)\}$ be a twice-differentiable basis of $\mathbb{L}^2([0,1])$ whose second derivatives are also linearly independent, for example, cubic B-splines, Fourier, or orthonormal cubic polynomials. Fourier coefficients of functional predictors $X_{ij}(t)$, functional regression coefficients $\beta_j(t)$, PLS directions, $\xi_j(t)$, and residualization coefficients $\delta_j(t)$ (with iteration indices suppressed) with respect to $\{b_m(t)\}$ are denoted as $\{\theta_{ijm}\}, \{\eta_{jm}\}, \{\gamma_{jm}\}$, and $\{\pi_{jm}\}$. For practicality, we truncate the expansion at a moderate M (e.g., 15-20) to capture most functional variation, while smoothness is handled via penalization (Section 3.2.1). The truncated expansion for the functional data is denoted as are denoted as $\widetilde{X}_{ij}(t) := \sum_{m=1}^{M} \theta_{ijm} b_m(t)$, but for the other funtions, we denote

$$\beta_j(t) := \sum_{m=1}^{M} \eta_{jm} b_m(t), \xi_j(t) := \sum_{m=1}^{M} \gamma_{jm} b_m(t), \delta_j(t) = \sum_{m=1}^{M} \pi_{jm} b_m(t), \tag{6}$$

since these functions are estimated, using a hat over the tilde notation would be cumbersome. Let Let $\boldsymbol{\theta}_{ij}$, $\boldsymbol{\eta}_j$, $\boldsymbol{\gamma}_j$, and $\boldsymbol{\pi}_j$ denote these coefficients as M-dimensional vectors. The truncations imply that all computations in this paper are carried out entirely within the subspace

$$\widetilde{\mathbb{H}} := \operatorname{span}\big(b_1(t), \ldots, b_M(t)\big)^K \times \mathbb{R}^p \subset \mathbb{H}. \tag{7}$$

The ith hybrid predictor, projected on $\widetilde{\mathbb{H}}$, is represented by the tuple $\widetilde{W_i} := (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{iK}, \boldsymbol{Z}_i)$. For notational convenience, we stack the predictor coefficient vectors as

$$\Theta_j := (\boldsymbol{\theta}_{1j}, \ldots, \boldsymbol{\theta}_{nj})^\top \in \mathbb{R}^{n\times M}, \quad \Theta := (\Theta_1, \ldots, \Theta_K, \mathbf{Z}) \in \mathbb{R}^{n\times(MK+p)}. \tag{8}$$

Let us denote the response vector as $\mathbf{y} := (y_1, \ldots, y_n)^\top$. Let $B, B'' \in \mathbb{R}^{M\times M}$ denote the Gram matrices of the basis functions and their second derivatives, with entries

$$B_{m,m'} := \int_0^1 b_m(t)\, b_{m'}(t)\, dt, \quad B''_{m,m'} := \int_0^1 b''_m(t)\, b''_{m'}(t)\, dt, \tag{9}$$

for $m, m' = 1, \ldots, M$. We then define the block-diagonal matrices

$$\mathbb{B} := \mathrm{blkdiag}(B, \ldots, B, I_p), \quad \mathbb{B}'' := \mathrm{blkdiag}(B'', \ldots, B'', I_p), \tag{10}$$

Then the full data for the hybrid PLS problem at the l-th iteration can be represented by the tuple

$$(\mathbb{B}, \mathbb{B}'', \Theta, \mathbf{y}) \in \mathbb{R}^{(MK+p)\times(MK+p)} \times \mathbb{R}^{(MK+p)\times(MK+p)} \times \mathbb{R}^{n\times(MK+p)} \times \mathbb{R}^n, \tag{11}$$

with the index l omitted for brevity.

Remark 1: Different bases could be used for each functional predictor to handle multiple dense or irregular functional predictors. We adopt a common basis for simplicity. The definitions of $\mathbb{B}$ and $\mathbb{B}''$ remain general enough to accommodate distinct bases if needed.

3.2 *Iterative steps*

The iterative process presented here yields an orthonormal hybrid basis that effectively captures the predictor-response relationships. It proceeds through two intermediate steps: the estimation of the PLS component direction (Section 3.2.1) and residualization (Section 3.2.2).

3.2.1 *Iterative step 1: regularized estimation of PLS component direction.* If the PLS direction estimates $\hat{\xi}_j$ are non-smooth, they produce a nonsmooth regression coefficient $\hat{\beta}$, since it is a linear combination of the PLS directions (Section 3.3). Non-smooth $\hat{\beta}$ can cause

overfitting and reduce interpretability. To address this, we propose a smoothness-enforcing estimation procedure by penalizing the roughness of each $\hat{\xi}_j$ via its integrated squared second derivative:

$$\mathrm{PEN}(\hat{\xi}_j) := \int_0^1 \{\hat{\xi}_j''(t)\}^2 dt. \tag{12}$$

Instead of solely maximizing the squared empirical covariance $\widehat{\mathrm{Cov}}^2(\langle \widetilde{W}, \xi\rangle_{\mathbb{H}}, Y)$ as in the classical NIPALS algorithm, we add a roughness penalty to simultaneously control the complexity of the estimated functional components. Our estimation procedure at the l-th iteration, iteration index omitted and assuming the observations have been residualized in previous steps, solves the following optimization problem:

$$\hat{\xi} := \arg\max_{\xi \in \widetilde{\mathbb{H}}} \ \widehat{\mathrm{Cov}}^2\big(\langle \widetilde{W}, \xi\rangle_{\mathbb{H}}, Y\big) \ s.t. \ \|\xi\|_{\mathbb{H}} + \sum_{j=1}^{K} \lambda_j \,\mathrm{PEN}(\xi_K) = 1. \tag{13}$$

Here, $\hat{\xi} \in \widetilde{\mathbb{H}}$ is an ordered pair expanded as:

$$\hat{\xi} = \big(\hat{\xi}_1(t), \ldots, \hat{\xi}_K(t), \hat{\boldsymbol{\zeta}}\big) = \left(\sum_{m=1}^{M} \hat{\gamma}_{1m}\, b_m(t), \ldots, \sum_{m=1}^{M} \hat{\gamma}_{Km}\, b_m(t), \hat{\boldsymbol{\zeta}}\right),$$

where $\hat{\boldsymbol{\zeta}}$ is the scalar part. The following theorem formulates the procedure obtaining this coefficients as a generalized Rayleigh quotient maximization:

THEOREM 5 (Regularized estimation of PLS component direction): *Let* $(\mathbb{B}, \mathbb{B}'', \Theta, \mathbf{y})$ *denote the data given at the* l*-th iteration, as defined in* (11)*. Let* $V := n^{-2}(\mathbb{B}\Theta^\top \mathbf{y})(\mathbb{B}\Theta^\top \mathbf{y})^\top$ *. Let* $\Lambda \in \mathbb{R}^{(MK+p)\times(MK+p)}$ *be defined as:*

$$\Lambda := \mathrm{blkdiag}(\lambda_1 I_M, \ldots, \lambda_K I_M, 0_{p\times p}), \ \textit{where } \lambda_1, \ldots, \lambda_K \geq 0. \tag{14}$$

Here, $0_{p\times p}$ *denotes the* $p\times p$ *zero matrix. The coefficients of the squared covariance maximizer defined in* (13)*, are obtained as*

$$\left(\hat{\gamma}_{11}, \ldots, \hat{\gamma}_{1M}, \ldots, \hat{\gamma}_{K1}, \ldots, \hat{\gamma}_{KM}, \hat{\boldsymbol{\zeta}}^\top\right)^\top = \arg\max_{\boldsymbol{\xi} \in \mathbb{R}^{MK+p}} \boldsymbol{\xi}^\top V \boldsymbol{\xi} \ \ s.t. \ \ \boldsymbol{\xi}^\top(\mathbb{B} + \Lambda\mathbb{B}'')\boldsymbol{\xi} = 1. \tag{15}$$

The constraint $\boldsymbol{\xi}^\top(\mathbb{B} + \Lambda\mathbb{B}'')\boldsymbol{\xi} = 1$ enforces the orthonormality of the estimated PLS component directions with respect to a modified inner product (see Section 3.4 for details). The smoothing parameter λ_k balances goodness of fit and smoothness in $\hat{\xi}_j$. Smaller λ_k

yields components that better fit the data but risks overfitting. Larger λ_k enforces greater smoothness, and in the limit $\lambda_k \to \infty$, $\hat{\xi}_j(t)$ approaches a linear form $a+bt$. In practice, both $\{\lambda_k\}$ and the number of components L can be selected via cross-validation using a predictive criterion such as mean squared error.

Computation. The generalized eigenproblem presented in theorem 5 may be computationally unstable in practice. However, by leveraging the rank-one structure of the matrix V theorem 6 derives a closed-form solution that requires only the solution of linear systems.

THEOREM 6 (Closed-form solution): *Consider the optimization problem described in theorem 5. Define the following quantities, which depend on the observed data but are not decision variables:*

$$\mathbf{u}_j := B\Theta_j^\top \mathbf{y} \in \mathbb{R}^M \quad \textit{for } j = 1, \ldots, K, \quad \textit{and} \quad \mathbf{v} := \mathbf{Z}^\top \mathbf{y} \in \mathbb{R}^p.$$

Let

$$q := \sum_{j=1}^{K} \mathbf{u}_j^\top (B + \lambda_j B'')^{-1} \mathbf{u}_j + \mathbf{v}^\top \mathbf{v}.$$

Then the unique (up to sign) solution to the regularized maximization problem is given in closed form by

$$\hat{\boldsymbol{\gamma}}_j = \frac{1}{\sqrt{q}} (B + \lambda_j B'')^{-1} \mathbf{u}_j \quad \textit{for } j = 1, \ldots, K, \quad \textit{and} \quad \hat{\boldsymbol{\zeta}} = \frac{1}{\sqrt{q}} \mathbf{v}.$$

The expressions above involve solving linear systems for the functional and scalar components separately, followed by normalization by a common factor. Although the unnormalized coefficients are obtained independently, the normalization step couples the functional and scalar parts, allowing them to influence one another. This coupling enables the procedure to capture the correlation between the functional and scalar components of the PLS direction.

3.2.2 *Iterative step 2: residualization via hybrid-on-scalar regression.* The l-th iteration's second step involves residualization of both predictors and responses. We first compute the

individual PLS score:

$$\hat{\rho}_i^{[l]} := \langle \widetilde{W}_i^{[l]}, \hat{\xi}^{[l]} \rangle_{\mathbb{H}}, \tag{16}$$

using the estimated PLS component direction $\widehat{\xi}^{[l]}$ obtained from Theorems 5 and 6. Since $\widetilde{W}_i^{[l]}$ are assumed to have a sample mean of zero, these PLS scores will also have a sample mean of zero. To obtain the $(l+1)$-th iteration's responses and hybrid predictors, we regress the (l)-th iteration's responses and hybrid predictors on these PLS scores by least squares and then residualize. Specifically, the $(l+1)$-th predictor is computed as a residual of hybrid-on-scalar linear regression model:

$$\widetilde{W}_i^{[l]} = \widehat{\rho}_i^{[l]} \hat{\delta}^{[l]} + \epsilon_i,$$

where $\hat{\delta}^{[l]} \in \widetilde{\mathbb{H}}$ is the regression coefficient. In the same spirit as the PLS component direction estimation step, rather than treating the hybrid object as a long vector of concatenated function evaluations at time points and scalar vectors, we employ a basis expansion approach to fit the entire hybrid object in one step. Therefore, our method is computationally efficient, and applicable for dense or irregular functional data. Consequently, $\delta^{[l]}$ is obtained by minimizing a least squares criterion:

$$\hat{\delta}^{[l]} := \arg\min_{\delta \in \widetilde{\mathbb{H}}} \sum_{i=1}^{n} \|\widetilde{W}_i^{[l]} - \hat{\rho}_i^{[l]}\, \delta\|_{\mathbb{H}}^2. \tag{17}$$

On the other hand, the $(l+1)$-th response is computed as a residual of a scalar-on-scalar linear regression model $Y_i^{[l]} = \hat{\nu}^{[l]} \widehat{\rho}_i^{[l]} + \epsilon_i$. The following theorem demonstrates that this residualization step can be performed simply, analogous to scalar PLS.

Lemma 3 (Closed-form solution): *Let us denote* $\hat{\boldsymbol{\rho}}^{[l]} := (\hat{\rho}_1^{[l]}, \ldots, \hat{\rho}_n^{[l]})^\top$. *The* $(l+1)$*-th iteration's predictors and responses are computed as*

$$\widetilde{W}_i^{[l+1]} := \widetilde{W}_i^{[l]} - \widehat{\rho}_i^{[l]} \hat{\delta}^{[l]}, \ \textit{where}\ \delta^{[l]} := \frac{1}{\|\hat{\boldsymbol{\rho}}^{[l]}\|_2^2} \sum_{i=1}^{n} \widehat{\rho}_i^{[l]} \widetilde{W}_i^{[l]},,$$

and

$$Y_i^{[l+1]} = Y_i^{[l]} - \hat{\nu}^{[l]} \widehat{\rho}_i^{[l]}, \ \textit{where}\ \hat{\nu}^{[l]} := \frac{\mathbf{y}^{[l]\top} \hat{\boldsymbol{\rho}}^{[l]}}{\|\hat{\boldsymbol{\rho}}^{[l]}\|_2^2}. \tag{18}$$

Since $\widetilde{W}_i^{[l]}$ and $Y_i^{[l]}$ are assumed to have a sample mean of zero, their respective residuals, $\widetilde{W}_i^{[l+1]}$ and $Y_i^{[l+1]}$, also maintain a zero sample mean.

3.3 *Final step: estimating the hybrid regression coefficient*

The hybrid regression coefficient β in model (3) can be written as a linear combination of PLS directions:

LEMMA 4: *Let use define* $\widehat{\iota}^{[1]} := \widehat{\xi}^{[1]}$. *For* $l \geq 2$, *we recursively define:*

$$\widehat{\iota}^{[l]} = \widehat{\xi}^{[l]} - \sum_{u=1}^{l-1} \langle \widehat{\delta}^{[u]}, \widehat{\xi}^{[l]} \rangle_{\mathbb{H}} \widehat{\iota}^{[u]}.$$

Then we have:

$$\widehat{\rho}_i^{[l]} = \langle W_i^{[l]}, \widehat{\xi}^{[l]} \rangle_{\mathbb{H}} = \langle W_i, \widehat{\iota}^{[l]} \rangle_{\mathbb{H}}.$$

Next, (18) leads to the following model:

$$Y_i = \sum_{l=1}^{L} \widehat{\nu}^{[l]} \widehat{\rho}_i^{[l]} + \epsilon_i.$$

This model lets us to express Y_i as:

$$Y_i = \sum_{l=1}^{L} \widehat{\nu}^{[l]} \langle W_i^{[l]}, \widehat{\xi}^{[l]} \rangle_{\mathbb{H}} + \epsilon_i = \Big\langle W_i, \sum_{l=1}^{L} \widehat{\nu}^{[l]} \widehat{\iota}^{[l]} \Big\rangle_{\mathbb{H}} + \epsilon_i,$$

which, given the uniqueness of β, leads to

$$\widehat{\beta} = \sum_{l=1}^{L} \widehat{\nu}^{[l]} \widehat{\iota}^{[l]} \sum_{l=1}^{L} \left(\widehat{\nu}^{[l]} - \sum_{k=l+1}^{L} \widehat{\nu}^{[k]} \langle \widehat{\delta}^{[l]}, \widehat{\xi}^{[k]} \rangle_{\mathbb{H}} \right) \widehat{\xi}^{[l]}. \tag{19}$$

Thus the number of PLS components L to be estimated can be chosen by cross-validation.

3.4 *Geometric properties*

A key property of PLS is that its directions are orthonormal and its scores orthogonal across iterations. Our regularized estimates preserve this property under a modified inner product that incorporates the roughness penalty, defined as follows:

DEFINITION 4 (Roughness-sensitive inner product): Given two hybrid predictors $W_1 = (X_{11}, \ldots, X_{1K}, \mathbf{Z}_1)$ and $W_2 = (X_{21}, \ldots, X_{2K}, \mathbf{Z}_2)$, both elements of $\mathbb{H}$ as defined in Definition

Algorithm 2 Hybrid partial least squares regression

Initialize: $(\mathbb{B}, \mathbb{B}'', \Theta^{[1]}, \mathbf{y}^{[1]})$ as the data objects after basis expansion (Section 3.1).

$\widetilde{W}_1^{[1]}, \ldots, \widetilde{W}_n^{[1]}, Y_1^{[1]}, \ldots, Y_n^{[1]} \leftarrow$ predictors ans responses from $(\mathbb{B}, \mathbb{B}'', \Theta^{[1]}, \mathbf{y}^{[1]})$

for $l = 1, 2, \ldots, L$ **do**

PLS direction and score estimation (theorem 6):

$\mathbf{u}_j^{[l]} \leftarrow B\Theta_j^{[l]\top}\mathbf{y}^{[l]},\ j = 1, \ldots, K$

$\mathbf{v}^{[l]} \leftarrow \mathbf{Z}^{[l]\top}\mathbf{y}^{[l]}$

$q^{[l]} \leftarrow \sum_{j=1}^{K} \mathbf{u}_j^{[l]\top}(B + \lambda_j B'')^{-1}\mathbf{u}_j^{[l]} + \mathbf{v}^{[l]\top}\mathbf{v}^{[l]}$

$(\hat{\gamma}_{j1}^{[l]}, \ldots, \hat{\gamma}_{jM}^{[l]})^\top \leftarrow \frac{1}{\sqrt{q}}(B + \lambda_j B'')^{-1}\mathbf{u}_j^{[l]},\ j = 1, \ldots, K$

$\hat{\boldsymbol{\zeta}}^{[l]} \leftarrow \frac{1}{\sqrt{q}}\mathbf{v}^{[l]}$

$\hat{\xi}^{[l]} \leftarrow \left(\sum_{m=1}^{M} \hat{\gamma}_{1m}^{[l]}\, b_m(t), \ldots, \sum_{m=1}^{M} \hat{\gamma}_{Km}^{[l]}\, b_m(t), \hat{\boldsymbol{\zeta}}^{[l]}\right)$ ▷ PLS direction

$\widehat{\rho}_i^{[l]} \leftarrow \langle \hat{\xi}^{[l]}, \tilde{W}_i^{[l]} \rangle,\ i = 1, \ldots, n$ ▷ PLS score

Residualization (theorem 3):

$\nu^{[l]} \leftarrow \frac{\sum_{i=1}^{n} Y_i^{[l]}\widehat{\rho}_i^{[l]}}{\sum_{i=1}^{n} \widehat{\rho}_i^{[l]2}}$ ▷ Least squares estimate

$Y_i^{[l+1]} \leftarrow Y_i^{[l]} - \nu^{[l]}\widehat{\rho}_i^{[l]}\ i = 1, \ldots, n$

$\widehat{\delta}^{[l]} \leftarrow \frac{1}{\sum_{i=1}^{n} \widehat{\rho}_i^{[l]2}} \sum_{i=1}^{n} \widehat{\rho}_i^{[l]}\widetilde{W}_i^{[l]}$ ▷ Least squares estimate

$\widetilde{W}_i^{[l+1]} \leftarrow \widetilde{W}_i^{[l]} - \widehat{\rho}_i^{[l]}\widehat{\delta}^{[l]}$

Regression coefficient estimation (Section 3.3):

$\widehat{\imath}^{[1]} \leftarrow \widehat{\xi}^{[1]}$

for $l = 2, \ldots, L$ **do**

$\widehat{\imath}^{[l]} \leftarrow \widehat{\xi}^{[l]} - \sum_{u=1}^{l-1} \langle \widehat{\delta}^{[u]}, \widehat{\xi}^{[l]} \rangle \widehat{\imath}^{[u]}$

$\widehat{\beta} \leftarrow \sum_{l=1}^{L} \widehat{\nu}^{[l]}\widehat{\imath}^{[l]}$ ▷ regression coefficient estimate

Output: the regression coefficient estimate $\widehat{\beta}$

1, and a roughness penalty matrix $\Lambda = \text{blkdiag}(\lambda_1 I_M, \ldots, \lambda_K I_M, 0_{p\times p})$, the roughness-sensitive inner product between W_1 and W_2 is defined as:

$$\langle W_1, W_2 \rangle_{\mathbb{H},\Lambda} := \sum_{k=1}^{K} \int_0^1 X_{1k}(t)X_{2k}(t)\, dt + \sum_{k=1}^{K} \lambda_k \int_0^1 X_{1k}''(t)X_{2k}''(t)\, dt + \mathbf{Z}_1^\top \mathbf{Z}_2. \tag{20}$$

Based on this inner product, the following theorem states that the PLS component directions estimated from theorem 5 are orthonormal.

THEOREM 7 (Orthonormality of estimated PLS component directions): *The PLS component directions* $\widehat{\xi}^{[1]}, \widehat{\xi}^{[2]}, \ldots, \widehat{\xi}^{[L]}$*, estimated via theorem 5 with a roughness penalty matrix* $\Lambda = \text{blkdiag}(\lambda_1 I_M, \ldots, \lambda_K I_M, 0_{p\times p})$*, are mutually orthonormal with respect to the inner product* $\langle \cdot, \cdot \rangle_{\mathbb{H},\Lambda}$*. That is,*

$$\langle \widehat{\xi}^{[l_1]}, \widehat{\xi}^{[l_2]} \rangle_{\mathbb{H},\Lambda} = \mathbb{1}(l_1 = l_2), \quad l_1, l_2 = 1, \ldots, L.$$

The next theorem states that the vectors of estimated PLS scores for different iteration numbers are mutually orthogonal.

THEOREM 8: *Recall from Lemma 3 that* $\widehat{\boldsymbol{\rho}}^{[l]}$ *denote the* n*-dimensional vector whose elements consist of the* l*-th estimated PLS scores* ($l = 1, \ldots, L$) *of* n *observations. The vectors* $\widehat{\boldsymbol{\rho}}^{[1]}, \widehat{\boldsymbol{\rho}}^{[2]}, \ldots, \widehat{\boldsymbol{\rho}}^{[L]}$ *are mutually orthogonal in the sense that*

$$\widehat{\boldsymbol{\rho}}^{[l_1]\top} \widehat{\boldsymbol{\rho}}^{[l_2]} = 0 \quad \text{for} \quad l_1, l_2 \in \{1, \ldots, L\},\ l_1 \neq l_2.$$

4. Application to Renal Study Data

We apply Algorithm 2 to the Emory University renal study data (Chang et al., 2020), which include 216 kidneys evaluated by diuretic renal scans. Following MAG3 injection, (i) baseline renogram (time activity) curves were recorded over 24 min, followed by (ii) post-furosemide curves from a second 20-min scan. Additional data comprise (iii) obstruction ratings from three nuclear medicine experts, (iv) eight summary statistics from baseline curves, (v) six summary statistics from post-furosemide curves, and (vi) subject age and gender. The response Y was obtained by averaging the experts' diagnoses and applying a min-max scaling and logit transformation. We treat (iv), (v), and (vi) as scalar predictors. The summary statistics (iv) and (v), while commonly used in medicine and informative about obstruction,

do not fully capture the renogram's information. Conversely, the renogram curves (i) and (ii), treated as functional predictors, contain complete but overly rich information, making it difficult to isolate the key features. Since (iv), (v) and (i), (ii) are highly correlated yet complementary, it is advantageous to include both as predictors.

We smoothed the functional predictors using a B-spline basis of order 15. We performed 100 random 70/30 train-test splits, reporting the mean and standard error of test RMSE normalized by the response range. Maximum components set to 20 and number of components are set via 5-fold CV with respect to RMSE. We compare our method (hybridpls) with two baselines: (i) penalized functional regression (pfr; Goldsmith et al., 2011) which can incorporate both functional and scalar predictors, and (ii)Principal component regression (fpcr): run both of PCA for multple functional predictors Happ and Greven (2018) and scalar PCA and run OLS on the PC scores. The results in Table 1 demonstrates that our method outperforms baseline methods.

[Table 1 about here.]

References

Aguilera, A. M., Aguilera-Morillo, M. C., and Preda, C. (2016). Penalized versions of functional PLS regression. *Chemometrics and Intelligent Laboratory Systems* **154,** 80–92.

Aguilera, A. M., Escabias, M., Preda, C., and Saporta, G. (2010). Using basis expansions for estimating functional PLS regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems* **104,** 289–305.

Beyaztas, U. and Lin Shang, H. (2022). A robust functional partial least squares for scalar-on-multiple-function regression. *Journal of Chemometrics* **36,** e3394.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34,** 2159–2179.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13,** 571–591.

Chang, C., Jang, J. H., Manatunga, A., Taylor, A. T., and Long, Q. (2020). A bayesian latent class model to predict kidney obstruction in the absence of gold standard. *Journal of the American Statistical Association* **115,** 1645–1663.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics* **40,** 322–352.

Febrero-Bande, M., Galeano, P., and González-Manteiga, W. (2017). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review* **85,** 61–83.

Geng, S., Kolar, M., and Koyejo, O. (2020). Joint Nonparametric Precision Matrix Estimation with Confounding. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pages 378–388. PMLR.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics* **20,** 830–851.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35,** 70–91.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association* **113,** 649–659.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* Wiley, 1st edition edition.

Jang, J. H. (2021). Principal component analysis of hybrid functional and vector data. *Statistics in Medicine* **40,** 5152–5173.

Kolar, M., Liu, H., and Xing, E. P. (2014). Graph estimation from multi-attribute data.

Journal of Machine Learning Research **15,** 1713–1750.

Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics & Data Analysis* **48,** 149–158.

Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association* **102,** 984–996.

Saricam, S., Beyaztas, U., Asikgil, B., and Shang, H. L. (2022). On partial least-squares estimation in scalar-on-function regression models. *Journal of Chemometrics* **36,** e3452.

Wold, HERMAN. (1975). Path models with latent variables: The NIPALS approach*. In Blalock, H. M., Aganbegian, A., Borodkin, F. M., Boudon, R., and Capecchi, V., editors, *Quantitative Sociology*, International Perspectives on Mathematical and Statistical Modeling, pages 307–357. Academic Press.

Zhao, Y., Ogden, R. T., and Reiss, P. T. (2012). Wavelet-based LASSO in functional linear regression. *Journal of Computational and Graphical Statistics* **21,** 600–617.

Table 1
Prediction performance on renal study dataset

Method	Test RMSE
pfr	0.2118 (0.0240)
fpcr	0.2732 (0.0327)
hybridpls	0.1686 (0.0205)