World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:18, No:11, 2024

Estimation of Functional Response Model by Supervised Functional Principal Component Analysis

Hyon I. Paek, Sang Rim Kim, Hyon A. Ryu

Abstract**—**In functional linear regression, one typical problem is to reduce dimension. Compared with multivariate linear regression, functional linear regression is regarded as an infinite-dimensional case, and the main task is to reduce dimensions of functional response and functional predictors. One common approach is to adapt functional principal component analysis (FPCA) on functional predictors and then use a few leading functional principal components (FPC) to predict the functional model. The leading FPCs estimated by the typical FPCA explain a major variation of the functional predictor, but these leading FPCs may not be mostly correlated with the functional response, so they may not be significant in the prediction for response. In this paper, we propose a supervised FPCA method for a functional response model with FPCs obtained by considering the correlation of the functional response. Our method would have a better prediction accuracy than the typical FPCA method.

Keywords**—**Supervised, functional principal component analysis, functional response, functional linear regression.

I. Introduction

WE consider prediction problem of functional response by using functional linear model as follows. If both responses and regressors are functions, the required model is denoted in its simplest formulation

$$Y_i(t) = \alpha(t) + \int \psi(t,s) X_i(s)ds + \varepsilon_i(t), \qquad i = 1,2,\dots,N \quad (1)$$

where $\alpha(t)$ is an intercept term, $X(t)$ is the functional predictor process with the mean function $\mu(t)$, $Y(t)$ is the functional response process with the mean function $\nu(t)$ and $\psi(t, s)$ is bivariate slope function. $\psi(t, s)$, $X(t)$ and $Y(t)$ are assumed to be smooth and square integrable on domain T.

Compared with the classic regression problem, the main challenge in this functional linear model is that even a single functional predictor can lead to a saturated model due to its high flexibility.

A common strategy to address this problem is through the FPCA. The FPCA method estimates the functional linear model (1) in two steps: estimating the FPCs for the functional predictor; and then using several leading FPCs in the functional linear model. This topic has been extensively studied in the literature such as [1]-[3]. Reference [4] applied a functional logistic regression to predict the high-risk birth rate based on periodically stimulated fetal heart rate tracings. Reference [5] considered response model related with the integral of a functional predictor. Reference [6] applied multinomial functional regression model to predict the land usage by using with coarse resolution sensing data. Reference [7] considered partial functional linear quantile regression by using with slope function estimated FPC basis.

However, one limitation of conventional FPCA is that the estimation of FPCs of predictors is separated from the regression model used to predict the response and the leading FPCs focus on explaining the maximum variation of the functional predictor. Thus, the estimated FPCs may not have the maximum prediction power for Y and the prediction model may not be optimal. In practice, they usually use as many FPCs as possible, which results excessive variability into the model.

Our method is to use the information from the response function to estimate FPCs to improve prediction performance of FPCs for the response function Y. This method is called supervised functional principal component analysis (sFPCA).

Reference [8] introduced a supervised principal component analysis (sPCA) method for multivariate regression and [9] proposed a supervised singular value decomposition (SupSVD) model. Reference [10] considered a principal component regression in the generalized linear model by using selected principal components as new predictors.

However, above methods could not be applied to functional response model. It is not trivial to extend sPCA to the functional case. Reference [11] proposed a functional singular component analysis method to quantify the dependence of pairs of functional data (X, Y). Reference [12] extended SupSVD model to FPCA and proposed a supervised sparse functional principal component analysis (SupSFPCA). The estimation is based on the penalized likelihood function that imposes smoothness and sparsity penalty on PC loadings. Reference [13] composed the integrated residual sum of squares which makes use of the association between the functional response and the predictors and obtained supervised principal components that minimize the integrated residual sum of squares to estimate functional linear model.

In this paper, we proposed a framework that uses the information of the response function to improve the prediction performance of the estimated FPCs. If we ignore the response function in estimation FPCs of the predictor to predict the response, then we might not estimate FPCs practically significant to prediction. Therefore, we proposed an estimation method of functional response model based on supervised FPCA with considering the correlation between the response function and FPCs in FPC estimation step.

The rest of the paper is organized as follows. A review of

Hyon I Paek, Sang Rim Kim, and Hyon A Ryu are with Faculty of Applied Mathematics, Kim Chaek University of Technology, Pyongyang, Democratic People's Republic of Korea (e-mail: phi92618@star-co.net.kp).

typical FPCA is outlined in Section II. Main contents of our method are described in Section III. Section IV gives concluding.

II. Estimation of Functional Response Linear Model Based on FPCA

We first review the classical FPCA method for estimating the functional linear model (1), which is also called unsupervised FPCA method. The error functions ε_i are assumed to be iid, square integrable, mean zero and independent of the regressors X_i, which are iid with the same distribution as a square integrable random function X in L^2.

We first show how to reduce the estimation of the kernel to the case of the zero intercept function α. Taking the expectations of both sides of (1), we obtain

$$\mu_Y(t) = \alpha(t) + \int \psi(t,s)\mu_X(s)ds$$

which leads to

$$Y_i(t) - \mu_Y(t) = \int \psi(t,s)(X_i(s) - \mu_X(s))ds + \varepsilon_i(t)$$

We assume that the population mean functions μ_Y and μ_X can be estimated well enough. In the case of fully observed functions, or functions transformed to functional objects, e.g. by spline smoothing of individual trajectories, we can use the estimators

$$\hat{\mu}_Y = \frac{1}{N}\sum_{i=1}^{N} Y_i(t), \quad \hat{\mu}_X = \frac{1}{N}\sum_{i=1}^{N} X_i(t)$$

We replace the curves X_i and Y_i by $X_i^c(t) = X_i(t) - \hat{\mu}_X$ and $Y_i^c(t) = Y_i(t) - \hat{\mu}_Y$ to obtain an estimator $\hat{\psi}$ in the model

$$Y_i^C(t) = \int \psi(t,s) X_i^C(s)ds + \varepsilon_i(t), \; i=1, 2, \ldots, N. \quad (2)$$

The intercept function is then estimated by $\hat{\alpha}(t) = \hat{\mu}_Y(t) - \int \hat{\psi}(t,s)\hat{\mu}_X(s)ds$. We therefore assume that the data follow the population model

$$Y(t) = \int \psi(t,s) X(s)ds + \varepsilon(t) \quad (3)$$

and that $EX = 0$, which implies $EY = 0$.

Our objective is to derive an estimator of the kernel ψ in (3). Firstly, we should find a suitable expression for the kernel ψ. Since we assume zero mean functions, we have the expansions

$$X(s) = \sum_{i=1}^{\infty} \xi_i \upsilon_i(s), \quad Y(t) = \sum_{j=1}^{\infty} \varsigma_j u_j(t), \quad (4)$$

where the υ_i are the FPCs of X and the u_j are the FPC's of Y.

Assuming that $\int \psi^2(t,s)dtds < \infty$, we will show that

$$\psi(t,s) = \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} \frac{E[\xi_\ell \zeta_k]}{E[\xi_\ell^2]} u_k(t)\upsilon_\ell(s). \quad (5)$$

Since the FPCs of a square integrable random function form a basis in $L^2([0,1])$, the bivariate functions $\{\upsilon_i(s)u_j(t), 0 \le s,t \le 1, i,j \ge 1\}$ form a basis in $L^2([0,1]\times[0,1])$. Therefore, the kernel ψ has the expansion

$$\psi(t,s) = \sum_{k,\ell=1}^{\infty} \psi_{k\ell} u_k(t)\upsilon_\ell(s)$$

where we can show that $\psi_{k\ell} = E[\xi_\ell \zeta_k]/E[\xi_\ell^2]$.(cf. [3])

Since $E[\xi_\ell^2] = \lambda_\ell$, setting $\sigma_{\ell k} = E[\xi_k \zeta_\ell]$, (5) and its estimator lead to the expressions

$$\psi(t,s) = \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} \frac{\sigma_{\ell k}}{\lambda_\ell} u_k(t)\upsilon_\ell(s)$$

$$\hat{\psi}(t,s) = \sum_{k=1}^{q}\sum_{\ell=1}^{p} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \hat{u}_k(t)\hat{\upsilon}_\ell(s). \quad (6)$$

The parameter $\sigma_{\ell k} = E[\xi_k \zeta_\ell]$ can be estimated as follows.

$$\hat{\sigma}_{\ell k} = \frac{1}{N}\sum_{i=1}^{N} < X_i, \hat{\upsilon}_\ell >< Y_i, \hat{u}_k >$$

The truncation levels p and q can be selected using the cumulative variance percentage approach. In more detail,

$$p = \inf\{k : \frac{\sum_{j=1}^{k}\lambda_j}{\sum_{j=1}^{\infty}\lambda_j} \ge 99\%\}$$

However, this procedure's prediction performance may not be optimal in many complex problems due to some reasons. First, the prediction power of those FPCs might not coincide with the amount of variation they explain. For instance, the response variable might only depend on the 10th FPC instead of any of the first 9 FPCs. Second, given a small sample size, a large value of p introduces excessive variability into the model, making the model selection a very difficult task. Therefore, there is necessity to improve the prediction performance of the estimated FPCs for each functional predictor.

III. Method

A. Supervised FPCA

Let us assume that $E(X(t)) = 0$ and $E(Y(t)) = 0$ in the following discussion. We can centralize $X(t)$ and $Y(t)$. We estimate FPC $\upsilon_1(t)$, $\upsilon_2(t)$, …as follows. The estimate $\hat{\upsilon}_k(t)$ maximizes

$$Q(\upsilon) = \frac{\theta < \upsilon, C\upsilon > + (1-\theta)E\left[< Y, \upsilon >\right]^2}{\|\upsilon\|^2}, \quad (7)$$

subject to $\|\upsilon\| = 1, < \upsilon, \hat{\upsilon}_j >= 0$, for every $j < k$ and $0 \le \theta \le 1$. Here the norm $\|\upsilon\| = \sqrt{\|\upsilon\|^2} = \sqrt{< \upsilon, \upsilon >}$ and $<f, g>$ denotes the usual L^2 inner product $< f, g >= \int_T f(t)g(t)dt$. C denotes the empirical covariance operator: $C\upsilon = \int_T \hat{C}(\cdot, t)\upsilon(t)dt$, where the empirical covariance function $\hat{C}(s,t) = \frac{1}{n}\sum_{i=1}^{n} X_i(s)X_i(t)$, and $X_i(t)$ is an independent realization of the stochastic process $X(t)$.

Let's take a closer look at the formalization of $Q(\upsilon)$ shown in (7). The first term in the numerator, $< \upsilon, C\upsilon >$ represents the variation within the functional predictor $X(t)$ that can be explained by $\upsilon(t)$; the second part in the numerator, $E\left[< Y, \upsilon >\right]^2$ represents the dynamical correlation between υ and response function Y. The balance between these two terms is controlled by the weight parameter θ. Specifying $\theta = 1$ will give rise to unsupervised FPCA. On the other hand, specifying θ less than 1 will lead to supervised FPCA. The weight parameter θ can be treated as a tuning parameter and can be determined using cross-validation.

The main reasons behind the 'squared' covariance, the second term on the numerator in (7), are two-fold. First, we wish to keep this term, which describes the association between the estimated FPC and the response function, of the numerator in (7) positive, since the variance of the FPC scores in the first term is always positive. Second, it can also help to convert the estimation process into an eigenvalue decomposition problem.

B. Smooth Supervised FPCA

The FPCs obtained using (7) might need to be further smoothed or regularized. We define another type of norm as $\|f\|_\lambda = \sqrt{\|f\|^2 + \lambda\|D^2 f\|^2}$, in which $D^2 f = \int_T f''(t)dt$. The smooth estimate for the k-th supervised FPC is obtained by maximizing

$$Q(\upsilon) = \frac{\theta < \upsilon, C\upsilon > + (1-\theta)E\left[< Y, \upsilon >\right]^2}{\|\upsilon\|_\lambda^2}, \quad (8)$$

subject to $\|\upsilon\|_\lambda = 1, < \upsilon, \hat{\upsilon}_j >= 0$, for every $j < k$ and $0 \le \theta \le 1$. The smoothing parameter λ controls the degree of smoothness. If $\lambda = 0$, there is no penalty on the roughness of the estimated component $\hat{\upsilon}(t)$ and the smooth supervised FPCs will reduce to the regular supervised FPCs. On the other hand, a very large value of λ will force the estimated component $\hat{\upsilon}(t)$ taking a linear form.

C. Computational Details

Let's give the computational details on how to estimate the smooth supervised FPC $\upsilon(t)$ given a set of value for (θ, λ). To simplify the computation, we use the same B-spline basis functions to represent FPC $\upsilon_j(t)$, the functional predictor $X_i(t)$ and the functional response $Y_i(t)$. Let $\Phi(t) = (\varphi_1(t), \varphi_2(t), \ldots, \varphi_M(t))^T$, we can rewrite $(X_1(t), X_2(t), \ldots, X_n(t))^T = S\Phi(t)$, where S is an $n \times M$ coefficient matrix. Similarly, we rewrite $(Y_1(t), Y_2(t), \ldots, Y_n(t))^T = R\Phi(t)$, where R is an $n \times M$ coefficient matrix. In addition, we represent $\upsilon(t) = \sum_{m=1}^{M} \beta_m \phi_m(t) = \boldsymbol{\beta}^T \Phi(t)$ and $u(t) = \sum_{m=1}^{M} \gamma_m \phi_m(t) = \boldsymbol{\gamma}^T \Phi(t)$, in which $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ denote the coefficient vectors $(\beta_1, \beta_2, \ldots, \beta_M)^T$ and $(\gamma_1, \gamma_2, \ldots, \gamma_M)^T$. Then the empirical covariance function can be expressed as

$$\hat{C}(s,t) = \frac{1}{n}\Phi(s)^T \mathbf{S}^T \mathbf{S}\Phi(t)$$

Therefore, the first term in the numerator of (7) is given by

$$< \upsilon, C\upsilon >= \frac{1}{n}\boldsymbol{\beta}^T \mathbf{W}\mathbf{S}^T\mathbf{S}\mathbf{W}\boldsymbol{\beta}, \quad (9)$$

where $\mathbf{W}$ is an $M \times M$ matrix with elements $w_{ij} = < \varphi_i(t), \varphi_j(t) >$.

For the second term in the numerator, we first derive the form of $< Y_i, \upsilon >$. For each $Y_i(t)$, $< Y_i, \upsilon >$ is written as

$$< Y_i, \upsilon >= \boldsymbol{\beta}^T \mathbf{W}\mathbf{R}_i^T = \boldsymbol{\beta}^T \mathbf{W}_i,$$

where $\mathbf{R}_i$ is the i-th row of the coefficient matrix $\mathbf{R}$, and $\mathbf{W}_i = \mathbf{W}\mathbf{R}_i^T$. Finally, the dynamical correlation term between Y and the FPC $\upsilon(t)$ is written as

$$E\left[< Y, \upsilon >\right]^2 = \frac{1}{n}\sum\left[< Y_i, \upsilon >\right]^2 = \frac{1}{n}\boldsymbol{\beta}^T \mathbf{W}\mathbf{R}^T\mathbf{R}\mathbf{W}\boldsymbol{\beta} \quad (10)$$

For the denominator part in (8), the norm of $\upsilon(t)$ is given by

$$\|\upsilon\|_\lambda^2 = \boldsymbol{\beta}^T \mathbf{W}\boldsymbol{\beta} + \lambda\boldsymbol{\beta}^T \mathbf{D}\boldsymbol{\beta} = \boldsymbol{\beta}^T \mathbf{G}\boldsymbol{\beta} \quad (11)$$

where $\mathbf{D}$ denotes a $M \times M$ matrix with element $d_{ij} = < D^2\varphi_i(t), D^2\varphi_j(t) >$ and $\mathbf{G} = \mathbf{W} + \lambda\mathbf{D}$.

Putting (9), (10) and (11) together, $Q(\upsilon)$ in (8) is given by

$$Q(\upsilon) = \frac{\boldsymbol{\beta}^T \mathbf{V}\boldsymbol{\beta}}{\boldsymbol{\beta}^T \mathbf{G}\boldsymbol{\beta}}$$

where

$$\mathbf{V} = \frac{1}{n}(\theta\mathbf{W}\mathbf{S}^T\mathbf{S}\mathbf{W} + (1-\theta)\mathbf{W}\mathbf{R}^T\mathbf{R}\mathbf{W}).$$

Let $\boldsymbol{\delta} = \mathbf{G}^{1/2}\boldsymbol{\beta}$, maximizing $Q(\upsilon)$ is equivalent to maximizing $\boldsymbol{\delta}^T(\mathbf{G}^{-1/2})^T\mathbf{V}\mathbf{G}^{-1/2}\boldsymbol{\delta}$ subject to $\boldsymbol{\delta}^T\boldsymbol{\delta} = 1$. Then $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_p$ will be the leading p eigenvector of the matrix: $(\mathbf{G}^{-1/2})^T\mathbf{V}\mathbf{G}^{-1/2}$. Consequently, we can derive $\hat{\boldsymbol{\beta}}_j = (\mathbf{G}^{1/2})^{-1}\boldsymbol{\delta}_j$. The corresponding smooth supervised FPC is $\hat{\upsilon}_j(t) = \hat{\boldsymbol{\beta}}_j^T \Phi(t)$ for $j = 1, \ldots, p$.

$\hat{u}_i(t)$ can be estimated by applying smooth unsupervised FPCA to $Y(t)$. The estimate $\hat{u}_k(t)$ maximizes

$$Q(u) = \frac{< u, C_y u >}{\|u\|_\lambda^2}$$

subject to $\|u\| = 1, < u, \hat{u}_i >= 0$, for every $i < k$. Here C_y denotes the empirical covariance operator:

$$C_y u = \int_T \hat{C}_y(\cdot, t) u(t) dt$$

where the empirical covariance function $\hat{C}_y(s,t) = \frac{1}{n}\sum_{i=1}^n Y_i(s) Y_i(t)$, and $Y_i(t)$ is an independent realization of the stochastic process $Y(t)$. If we use the same smoothing parameter λ as in estimation of FPCs for predictor, then $Q(u)$ is simplified in matrix form as follows:

$$Q(u) = \frac{\boldsymbol{\gamma}^T \mathbf{U} \boldsymbol{\gamma}}{\boldsymbol{\gamma}^T \mathbf{G} \boldsymbol{\gamma}}$$

where

$$\mathbf{U} = \frac{1}{n} \mathbf{W} \mathbf{R}^T \mathbf{R} \mathbf{W}$$

Let $\boldsymbol{\eta} = \mathbf{G}^{1/2} \boldsymbol{\gamma}$, maximizing $Q(u)$ is equivalent to maximizing $\eta^T (\mathrm{G}^{-1/2})^T \mathrm{U} \mathrm{G}^{-1/2} \eta$ subject to $\eta^T \eta = 1$. Then $\eta_1, \ldots, \eta_q$ will be the leading q eigenvector of the matrix $(\mathbf{G}^{-1/2})^T \mathbf{U} \mathbf{G}^{-1/2}$. Consequently, we can derive $\hat{\boldsymbol{\gamma}}_i = (\mathbf{G}^{1/2})^{-1} \boldsymbol{\eta}_i$. The corresponding smooth FPC is $\hat{u}_i(t) = \hat{\boldsymbol{\gamma}}_i^T \Phi(t)$ for $j = 1, \ldots, q$.

D.Functional Regression

With the estimated leading p FPCs $\hat{\upsilon}_1(t), \hat{\upsilon}_2(t), \ldots, \hat{\upsilon}_p(t)$, we can estimate functional regression model of $Y(t)$ on $X(t)$.

$$Y(t) = \int \psi(t,s) X(s) ds + \varepsilon(t)$$

The slope function is given by

$$\hat{\psi}(t,s) = \sum_{k=1}^{q} \sum_{\ell=1}^{p} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \hat{u}_k(t) \hat{\upsilon}_\ell(s)$$

where

$$\hat{\sigma}_{\ell k} = \frac{1}{N} \sum_{i=1}^{N} < X_i, \hat{\upsilon}_\ell >< Y_i, \hat{u}_k >$$

and

$$\hat{\lambda}_\ell = \frac{1}{N} \sum_{i=1}^{N} < X_i, \hat{\upsilon}_\ell >^2$$

The number of FPCs used in functional regression, p and q can be regarded as tuning parameters. We can determine the p as follows. First, we set $p = 1$, evaluate cross-validation error whenever p increases. The experience shows that we can choose p when the cross-validation error is decreased smaller than the fixed threshold value. We can choose q in the same way.

Let $x(s)$ denote $x(s) = \sum_{i=1}^{M} c_i \phi_i(s)$, then the estimate for response $y(t)$ is given by

$$\begin{aligned} \hat{y}_i(t) &= \int_T \sum_{k=1}^{q} \sum_{\ell=1}^{p} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \sum_{m=1}^{M} \hat{\gamma}_{km} \phi_m(t) \sum_{m=1}^{M} \hat{\beta}_{\ell m} \phi_m(s) \sum_{i=1}^{M} c_i \phi_i(s) ds = \\ &= \sum_{k=1}^{q} \sum_{\ell=1}^{p} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \sum_{m=1}^{M} \hat{\gamma}_{km} \phi_m(t) \int_T \sum_{m=1}^{M} \hat{\beta}_{\ell m} \phi_m(s) \sum_{i=1}^{M} c_i \phi_i(s) ds = \\ &= \sum_{k=1}^{q} \sum_{\ell=1}^{p} \frac{\hat{\sigma}_{\ell k}}{\hat{\lambda}_\ell} \hat{\boldsymbol{\beta}}_\ell^T \mathbf{W} \mathbf{C} \hat{\boldsymbol{\gamma}}_k^T \Phi(t) \end{aligned} \quad , \qquad (12)$$

where $\mathbf{C} = (c_1, c_2, \ldots, c_M)^T$.

IV. Conclusion

In this paper, we consider the prediction problem of response function by using FPC of functional predictor. The classic FPCA method estimates FPCs to maximize the variation of FPC scores and ignore the response function. We proposed a supervised FPCA method to estimate FPCs related to the response function to improve the prediction performance of FPCs.

Our method can be extended to the case that there are several functional predictors. In this case, it is worth to research a suitable determination method of several weight parameters for correlations between the response and FPCs of each predictor.

Acknowledgment

The authors would like to thank the reviewers for their worthy suggestions for the improvement of the manuscript. We also thank Dr. Sung Hyon Ri and Dr. Il Yong Kang, who gave us much valuable advice in the early stages of this work.

References

[1] Ramsay, J. O., Silverman, B. W., 2005. Functional Data Analysis (Second edition). New York: Springer.
[2] Ramsay, J. O., Hooker, G., Graves, S., 2009. Functional Data Analysis with R and MATLAB. Use R! Springer New York.
[3] Kokoszka, P., Reimherr, M., 2018. Introduction to functional data analysis. CRC Press.
[4] Ratcliffe, S. J., Heller, G. Z., Leader, L. R., 2002. Functional data analysis with application to periodically stimulated foetal heart rate data. ii: Functional logistic regression. Statistics in medicine 21(8), 1115–1127.
[5] Muller, H.-G., Stadtmuller, U., 2005. Generalized functional linear models. Annals of Statistics, 774–805.
[6] Cardot, H., Faivre, R., Goulard, M., 2003. Functional approaches for predicting land use with the temporal evolution of coarse resolution

remote sensing data. Journal of Applied Statistics 30(10), 1185–1199.

[7] Tang, Q.G., Cheng, L.S., 2013. Partial functional linear quantile regression. Science China Press and Springer-Verlag Berlin Heidelberg 57, 1-20.

[8] Bair, E., Hastie, T., Paul, D., Tibshirani, R., 2006. Prediction by supervised principal components. Journal of the American Statistical Association 101(473).

[9] Li, G., Yang, D., Nobel, A. B., Shen, H., 2015. Supervised singular value decomposition and its asymptotic properties. Journal of Multivariate Analysis.

[10] Kawano, S., Fujisawa, H., Takada, T., Shiroishi, T., 2018. Sparse principal component regression for generalized linear models. Computational Statistics and Data Analysis 124, 180-196

[11] Yang, W., Muller, H.-G., 2011. Functional singular component analysis. Journal of the Royal Statistical Society 3(73), 303-324.

[12] Li, G., Shen, H., Huang, J. Z., 2016. Supervised sparse and functional principal component analysis. Journal of Computational and Graphical Statistics 25(3), 859–878.

[13] Zhang, X.Y., Sun, Q., Kong, D., 2023. Supervised principal component regression for functional responses with high dimensional predictors. arXiv:2103.11567v4.