Second International Workshop on
Functional and Operatorial Statistics.
Santander, June 16-18, 2011

Structured penalties for generalized functional linear models (GFLM)

Jaroslaw Harezlak[#,*], and Timothy W. Randolph[†]

[#]Indiana University School of Medicine
Division of Biostatistics
Indianapolis, IN 46202
[†]Fred Hutchinson Cancer Research Center
Biostatistics and Biomathematics
Seattle, WA 98109
[*]harezlak@iupui.edu

Abstract

GFLMs are often used to estimate the relationship between a predictor function and a response (e.g. a binary outcome). This manuscript provides an extension of a method recently proposed for functional linear models (FLM) - PEER (partially empirical eigenvectors for regression) to GFLM. The PEER approach to FLMs incorporates the structure of the predictor functions via a joint spectral decomposition of the predictor functions and a penalty operator into the estimation process via a generalized singular value decomposition. This approach avoids the more common two-stage smoothing basis approach to estimating a coefficient function. We apply our estimation method to a magnetic resonance spectroscopy data with binary outcomes.

1. Introduction

The coefficient function, β, in a GFLM represents the linear relationship between a transformed mean of the scalar response, y, and a predictor, x, formally written as $g(E[y]) = \int x(t)\beta(t)\,dt$, where $g(\cdot)$ is a so called link function. The problem typically involves a set of n responses $\{y_i\}_{i=1}^n$ corresponding to a set of observations $\{x_i\}_{i=1}^n$, each arising as a discretized sampling of an idealized function; i.e., $x_i \equiv (x_i(t_1), ..., x_i(t_p))$, for some, $t_1, ..., t_p$, of $[0, 1]$. We assume the predictors have been sampled densely enough to capture a spatial predictor structure and thus $p >> n$.

Classical approaches (see for example, Crambes et.al., 2009 and Hall et.al., 2007) to the ill-posed problem of estimating β use either the eigenvectors determined by the predictors (e g. principal components regression - PCR) or methods based on a projection of the predictors onto a pre-specified basis and then obtaining an estimate from a generalized linear model formed by the transform coefficients. These methods, however, do not provide an analytically tractable way of incorporating the predictor's functional structure directly into the GFLM estimation process.

Here, we extend the framework developed in Randolph et al. (2011) which exploits the analytic properties of a joint eigen-decomposition for an operator pair—a penalty operator, L, and the operator determined by the predictor functions, X. More specifically, we exploit an eigenfunction basis whose functional structure is inherited by both L and X. As this basis is algebraically determined by the shared eigenproperties of both operators, it is neither strictly empirical (as with principal components) nor completely external to the problem (as in the case of B-spline regression models). Consequently, this approach avoids a separate fitting or smoothing step. We refer to this approach as PEER (partially empirical eigenvector regression) and here provide an overview of PEER as developed for FLMs and then describe the extension to GFLMs.

2. Overview of PEER

We consider estimates of the coefficient-function β arising from a squared-error loss with quadratic penalty. These may be expressed as

$$\tilde{\beta}_{\alpha,L} = \arg\min_{\beta}\{||y - X\beta||^2_{\mathbb{R}^n} + \alpha||L\beta||^2_{L^2}\}, \tag{1}$$

where L is a linear operator.

Within this classical formulation, PEER exploits the *joint* spectral properties of the operator pair (X, L). This perspective allows the estimation process to be guided by an informed construction of L. It succeeds when structure in the generalized singular vectors of the pair (X, L) is commensurate with the appropriate structure of β. How L imparts this structure via the GSVD is detailed in Randolph et al. (2011), and so the discussion here is restricted to providing the notation necessary for the GFLM setting.

A least-squares solution, $\hat{\beta}$, satisfies the normal equations $X'X\beta = X'y$. Estimates arise as minimizers, $\hat{\beta} = \arg\min_{\beta}||y - X\beta||^2$, but there are infinitely many such solutions and so regularization is required. The least-squares solution with a minimum norm is provided by the singular value decomposition (SVD): $X = UDV'$ where the left and right singular vectors, u_k and v_k, are the columns of U and V, respectively, and $D = \text{diag}\{\sigma_k\}_{k=1}^{p}$, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$ ($r = \text{rank(X)}$, $\sigma_r \approx 0$). The minimum-norm solution is $\hat{\beta}_+ = X^\dagger y = \sum_{\sigma_k \neq 0}(1/\sigma_k)u_k' y\, v_k$, where $X^\dagger$ denotes the Moore-Penrose inverse of X: $X^\dagger = VD^\dagger U'$, where $D^\dagger = \text{diag}\{1/\sigma_k \text{ if } \sigma_k \neq 0;\ 0 \text{ if } \sigma_k = 0\}$.

For functional data, however, $\hat{\beta}_+$ is an unstable estimate which motivates PCR estimate: $\tilde{\beta}_{\text{PCR}} = V_d D_d^{-1} U_d' y$ where $A_d \equiv \text{col}[a_1, ..., a_d]$ denotes the first d columns of a matrix A. Another classical way to obtain a more stable estimate in terms of the ordinary singular vectors is to impose a ridge penalty, $L = I$ (see Hoerl et.al., 1970) for which the minimizing solution to (1) is

$$\tilde{\beta}_{\alpha,\mathrm{R}} = (X'X + \alpha I)^{-1}X'y = \sum_{k=1}^{r} \left(\frac{\sigma_k^2}{\sigma_k^2 + \alpha} \right) \frac{1}{\sigma_k} u_k' y \, v_k, \tag{2}$$

For a given linear operator L and parameter $\alpha > 0$, the estimate in (1) takes the form

$$\tilde{\beta}_{\alpha,L} = (X'X + \alpha L'L)^{-1}X'y. \tag{3}$$

This cannot be expressed using the singular vectors of X alone, but the generalized singular value decomposition of the pair (X, L) provides a tractable and interpretable vector expansion.

We provide here a short description of the GSVD method. Additional details are available in the Randolph et al. (2011). It is assumed that X is an $n \times p$ matrix ($n \leq p$) of rank n, L is an $m \times p$ matrix ($m \leq p$) of rank m and the null spaces of X and L intersect trivially: $\mathrm{Null}(L) \cap \mathrm{Null}(X) = \{0\}$. This condition is needed to obtain a unique solution and is natural in our applications. It is not required, however, to implement the methods. We also assume that $n \leq m \leq p$, with $m + n \geq p$, and the rank of $Z := [X'L']'$ is at least p.

Then there exist orthogonal matrices U and V, a nonsingular matrix W and diagonal matrices S and M such that

$$\begin{aligned} X &= U\underline{S}W^{-1}, & \underline{S} &= \begin{bmatrix} 0 & S \end{bmatrix}, & S &= \mathrm{diag}\{\sigma_k\} \\ L &= V\underline{M}W^{-1}, & \underline{M} &= \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix}, & M &= \mathrm{diag}\{\mu_k\}. \end{aligned} \tag{4}$$

The diagonal entries of S and M are ordered as

$$\begin{matrix} 0 \leq \sigma_1 \leq \sigma_2 \leq ...\sigma_n \leq 1 \\ 1 \geq \mu_1 \geq \mu_2 \geq ...\mu_n \geq 0 \end{matrix} \qquad \text{where} \qquad \sigma_k^2 + \mu_k^2 = 1, \quad k = 1, ..., n. \tag{5}$$

Denote the columns of U, V and W by u_k, v_k and w_k, respectively. For the majority of matrices L the generalized singular vectors u_k and v_k are not the same as the ordinary singular vectors of X. One case when they are the same is for $L = I$.

The penalized estimate is a linear combination of the columns of W and the solution to the penalized regression in (1) can be expressed as

$$\tilde{\beta}_{\alpha,L} = \sum_{k=p-n+1}^{p} \left(\frac{\sigma_k^2}{\sigma_k^2 + \alpha\mu_k^2} \right) \frac{1}{\sigma_k} u_k' y \, w_k, \tag{6}$$

We refer to any $\tilde{\beta}_{\alpha,L}$ ($L \neq I$) as a PEER (partially empirical eigenvectors for regression) estimate. The utility of a penalty L depends on whether the true coefficient function shares structural properties with this GSVD basis. With regard to this, the importance of the parameter α may be reduced by a judicious choice of L (Varah, 1979) since the terms in (6) corresponding to the vectors $\{w_k : \mu_k = 0\}$ are independent of α.

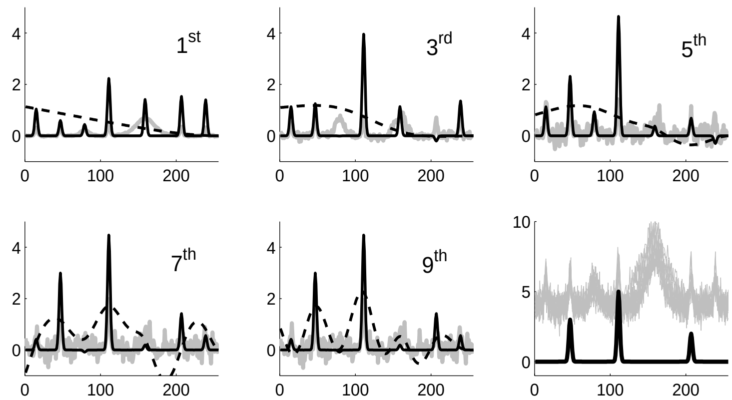

Figure 1: **Partial sums of penalized estimates.** The first five odd-numbered partial sums from (6) for three penalties, L: 2nd-derivative (dotted black), ridge (solid gray), targeted (solid black); see text. The last panel exhibits β (solid black) and several predictors, x_i (light gray), from the simulation.

2.1 Structured and targeted penalties

A *structured penalty* refers to a second term in (1) that involves an operator chosen to encourage certain functional properties in the estimate. Here we give examples of such penalties. If we begin with some knowledge about the subspace of functions in which the informative signal resides, then we can define a penalty based on it. For example, suppose $\beta \in \mathcal{Q} := \text{span}\{q_j\}_{j=1}^d$ for some $q_j \in L^2(\Omega)$. Set $Q = \sum_{j=1}^d q_j \otimes q_j$ and consider the orthogonal projection $P_\mathcal{Q} = QQ^\dagger$. Define $L_\mathcal{Q} = I - P_\mathcal{Q}$, then $\beta \in \text{Null}(L_\mathcal{Q})$ and $\tilde{\beta}_{\alpha, L_\mathcal{Q}}$ is unbiased.

Figure 1 illustrates the estimation process with plots of some partial sums from equation (6) for three estimates. The ridge estimate is, naturally, dominated by the leading eigenvectors of X. The second-derivative penalized estimate is dominated first by low-frequency structure. The targeted PEER estimate shown here begins with the largest peaks corresponding the largest GSV components, but quickly converges to the informative features.

2.2 Analytical properties

For a general linear penalty operator L, the analytic form of the estimate and its basic properties of bias, variance and MSE are provided in Randolph et al. (2011). Any direct comparison between estimates using different penalty operators is confounded by the fact there is no simple connection between the generalized singular values/vectors and the ordinary singular values/vectors. Therefore, Randolph et al. (2011) first considered the case of targeted or projection-based penalties. Within this class, a parameterized family of estimates is comprised of *ordinary* singular values/vectors. Since the ridge and PCR estimates are contained in (or a limit of) this family, an analytical comparison with

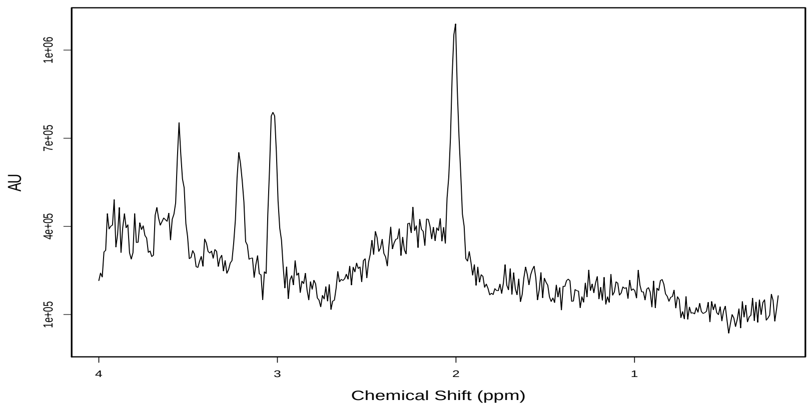

Figure 2: **A sample magnetic resonance spectroscopy (MRS) spectrum displaying brain metabolite levels in one frontal gray matter (FGM) voxel.**

some targeted PEER estimates is possible.

3. Extension to GFLM

Generalization of PEER to the GFLM setting proceeds via replacement of the continuous responses $y_1, \ldots, y_n$ by responses coming from a general exponential family whose expectations $g(\mu_i)$ are linearly related to a functional predictor X_i. We specifically focus here on the binary responses and logistic regression setting. We replace the least squares criterion by a likelihood function appropriate for the member of the exponential family distribution and find the estimate of β by minimizing the following expression:

$$\tilde{\beta}_{\alpha,L} = \arg\min_{\beta}\{\sum_i l(g(y_i), X_i\beta) + \alpha||L\beta||^2_{L^2}\}, \tag{7}$$

where $l(\cdot)$ is the log-likelihood function. The fitting procedure for PEER in GFLM setting is a modification of an iteratively reweighted least squares (IRLS) method. In a similar spirit to the BLUP and REML estimation of the tuning parameter in the linear mixed model equivalent setting, we select the tuning parameter using the penalized quasi-likelihood (PQL) method associated with the generalized linear mixed models. REML criterion is preferred here, since it has been been shown to outperform the GCV method (see Reiss and Ogden, 2007).

4. Application to a magnetic resonance spectroscopy data

We apply the GFLM-PEER method to study the relationship of the magnetic resonance spectroscopy (MRS) data and neurocognitive impairment arising from the HIV Neuroimaging Consortium (HIVNC) study (see Harezlak et al., 2011 for the study description). In particular, we are interested in studying the relationship of the metabolite level concentrations in the brain and classification of the patients into neurologically asymptomatic (NA) and neurocognitively impaired (NCI) groups. The predictor functions come in the form of spectra for each studied voxel in the brain (see Figure 2). Our method provides promising results when compared to the more established functional regression methods which do not take into account the external pure metabolite spectra profiles . We also obtain interpretable functional regression estimates that do not rely on a two–step procedure estimating the metabolite concentrations first and then using them as predictors in a logistic regression model.

5. Discussion

Estimation of the coefficient function β in a generalized functional linear model requires a regularizing constraint. When the data contain natural spatial structure (e.g., as derived from the physics of the problem), then the regularizing constraint should acknowledge this. In the FLM case, exploiting properties of the GSVD provided a new analytically-rigorous approach for incorporating spatial structure into functional linear models. In the GFLM case, we extend the IRLS procedure to take into account the penalty operator.

A PEER estimate is intrinsically based on GSVD factors. This fact guides the choice of appropriate penalties for use in both FLM and GFLM. Heuristically, the structure of the penalty's least-dominant singular vectors should be commensurate with the informative structure of β. The properties of an estimate are determined jointly by this structure and that in the set of predictors. The structure of the generalized singular functions provides a mechanism for using a priori knowledge in choosing a penalty operator allowing, for instance, one to target specific types of structure and/or avoid penalizing others. The effect a penalty has on the properties of the estimate is made clear by expanding the estimate in a series whose terms are the generalized singular vectors/values for X and L.

References

[1] Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing spline estimators for functional linear regression, *Ann. Statist.*, **37**, 35–72.

[2] Hall, P., and Horowitz, J.L. (2007). Methodology and convergence rates for functional linear regression, *Ann. Statist.*, **35**, 70–91.

[3] Harezlak, J., Buchthal, S., Taylor, M., Schifitto, G., Zhong, J., Daar, E., and et al., (2011). Persistence of HIV- associated cognitive impairment, inflammation and neuronal injury in the HAART era. *AIDS*, in press, doi: 10.1097/QAD.0b013e3283427da7.

[4] Hoerl, A.E., and Kennard, R.W., (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55-67.

[5] Randolph, T., Harezlak, J. and Feng, Z. (2011). Structured penalties for functional linear models: partially em- pirical eigenvectors for regression, unpublished manuscript.

[6] Reiss, P.T. and Ogden, R.T. (2007) Functional principal component regression and functional partial least squares, *JASA*, **102**, 984-986.

[7] Varah, J.M. (1979) A practical examination of some numerical methods for linear discrete ill-posed problems, *SIAM Review*, **21**, no. 1, 100-111.