MATHEMATICS
FOR
APPLICATIONS

SOME EQUIVALENCE RELATIONSHIPS
OF REGULARIZED REGRESSIONS

YUN ZHANG, JUILEE THAKAR, DAVID J. TOPHAM, ANN R. FALSEY, DELU ZENG
AND XING QIU

Abstract. Regularization is a powerful framework for solving ill-posed problem and preventing model overfitting in modern regression analysis. It is especially useful for high-dimensional or functional (infinite dimensional) regression models. In this paper, we construct two useful equivalence relationships for regularized regression: 1. An equivalence between regularized functional regression and regularized multivariate regression. This equivalence provides a computationally efficient way to fit the concurrent functional regression model. 2. An equivalence of penalized multivariate regression under a group of scaling transformation. This equivalence can be used to solve weighted principal component regression efficiently.

1. Introduction

Regression models are the most popular predictive modeling method for data with response and covariates. In many real-life applications such as computational biology, we often have to analyze datasets with more features (covariates) than the actual number of samples. In the regression setting, let p be the number of covariates and n be the number of samples, traditional methods based on least squares or maximum likelihood will fail when $p \approx n$ or $p > n$. In such "large p, small n" situation, regularized regression and principle component regression are commonly used methods to overcome the curse of dimensionality [7].

Let $\mathbf{y} \in \mathbb{R}^n$ be the response variable and $\mathbf{X} \in \mathrm{M}_{n \times p}$ be a collection of p independent variables. Assuming normal random errors $\boldsymbol{\epsilon}$, most commonly used regularized regression can be represented as the following penalized least square problem

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \hat{\boldsymbol{\beta}} = \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 + \mathrm{PEN}(\boldsymbol{\beta}). \tag{1.1}$$

Regularization on the regression coefficients $\boldsymbol{\beta}$ is specified through $\text{PEN}(\boldsymbol{\beta})$, also called the penalty term. The most popular choices of penalty are the L1 penalty (i.e. LASSO regression [15]) and the L2 penalty (i.e. ridge regression [6]). A combination of both L1 and L2 penalties is called the elastic-net regression [17] and has the following form

$$\text{PEN}(\boldsymbol{\beta}) = \lambda_1 \|\boldsymbol{\beta}\|_1 + \lambda_2 \|\boldsymbol{\beta}\|^2,$$

where λ_1 and λ_2 are the L1 and L2 penalty parameters that can be trained through cross-validation or generalized cross-validation.

In principle, these regularization technique can be applied in many advanced regression settings, such as functional linear regression [14] and principle component regression (PCR) [7]. However, computational efficiency and the selection of best penalty parameters are common issues for these problems. In this paper, we establish two forms of equivalence on the regularized regressions. The first one is a connection between penalized functional linear regression and multivariate regression, which can be applied to greatly reduce the computational burden in regularized concurrent functional regression; the second one reveals the equivalence relationship of penalized regressions under a family of scaling transformations, which can be used to implement a special version of PCR in which the penalty terms for different covariates (PCs) depends on the proportions of variance explained by these PCs. In addition, the second equivalence relationship also provides the answer for this problem: When the scale of the covariates are very different, what is the "fair" penalty for each term?

Aside from these practical applications, we believe our study also provides with researchers a different perspective of investigating the regression problems. That is, if the direct penalized regression problem is hard to solve, it is perhaps more efficient to reformulate the original regression problem into an equivalent but more familiar form with a known efficient solution. If such equivalence exists, all we have to do is to create an artificial "equivalent dataset" based on the original input data and the equivalence relationship, and then apply the well optimized method for the more familiar regression problem on such equivalent dataset.

2. EQUIVALENCE I: FAST ALGORITHM OF REGULARIZED FUNCTIONAL REGRESSION

Recall that Model (1.1) is an optimization problem with respect to the following objective function

$$\text{OBJ}(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) := \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \text{PEN}(\boldsymbol{\beta}).$$

We first present a lemma that describes an equivalence relationship among the standard regularized multivariate regression problems formulated by Model (1.1).

Lemma 2.1. *Assume $n > p$. Suppose matrix $\tilde{\mathbf{X}} \in \text{M}_{p \times p}$ and vector $\tilde{\mathbf{y}} \in \mathbb{R}^p$ satisfying*

$$\mathbf{X}'\mathbf{X} = \tilde{\mathbf{X}}'\tilde{\mathbf{X}}, \quad \mathbf{X}'\mathbf{y} = \tilde{\mathbf{X}}'\tilde{\mathbf{y}}. \tag{2.1}$$

Then a solution $\hat{\boldsymbol{\beta}}$ that minimizes $\text{OBJ}(\boldsymbol{\beta}|\tilde{\mathbf{X}}, \tilde{\mathbf{y}})$ also minimizes $\text{OBJ}(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y})$, and vice versa.

Proof.

$$
\begin{aligned}
\mathrm{OBJ}(\boldsymbol{\beta}|\mathbf{X},\mathbf{y}) &:= \|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2 + \mathrm{PEN}(\boldsymbol{\beta})\\
&= \|\mathbf{y}\|^2 - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathrm{PEN}(\boldsymbol{\beta})\\
&= \underbrace{\|\mathbf{y}\|^2 - \|\tilde{\mathbf{y}}\|^2}_{\text{constant}} + \|\tilde{\mathbf{y}}\|^2 - 2\boldsymbol{\beta}'\tilde{\mathbf{X}}'\tilde{\mathbf{y}} + \boldsymbol{\beta}'\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\boldsymbol{\beta}\\
&\quad + \mathrm{PEN}(\boldsymbol{\beta})\\
&= \mathrm{OBJ}(\boldsymbol{\beta}|\tilde{\mathbf{X}},\tilde{\mathbf{y}}) + \mathrm{Const.}
\end{aligned}
\tag{2.2}
$$

In other words, $\mathrm{OBJ}(\boldsymbol{\beta}|\mathbf{X},\mathbf{y})$ and $\mathrm{OBJ}(\boldsymbol{\beta}|\tilde{\mathbf{X}},\tilde{\mathbf{y}})$ only differ by a constant; thus the two optimization problems must have the same optimizer, which is $\hat{\boldsymbol{\beta}}$. $\qquad\square$

Equation (2.2) establishes an equivalence between two regularized regression problems in the high-dimensional space. We can regard $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$ that satisfy Equation (2.1) as an artificial *equivalent dataset* for the original data $\mathbf{X}$ and $\mathbf{y}$. Since we assume $n > p$, the equivalent dataset $(\tilde{\mathbf{X}},\tilde{\mathbf{y}})$ is smaller than the original datasets $(\mathbf{X},\mathbf{y})$; thus the equivalent regression problem is computationally easier to solve. Of note, we must point out that equivalent data are not unique.

Using the idea of equivalent data, Lemma 2.1 can be extended to provide a computational efficient solution to the regularized functional concurrent regression problems.

Functional linear regression is a generalization of multivariate linear regression in which one or more of the covariates and/or response variable can be functions instead of scalar or vector observations. A functional regression model in which the functional response, denoted as $y(t)$, is linearly dependent on the functional covariates observed at the same time, denoted as $\mathbf{X}(t)$, is called the *concurrent functional linear regression model* [14]. If a functional linear regression model contains both functional and scalar/vector-valued covariates, we can always use the *constant basis system* to turn scalar/vector-valued covariates into functions. Thanks to this technique, concurrent functional linear regression model can be adapted for many different problems and is arguably the most commonly used functional regression model in practice.

Regularizations for functional regression models have been proposed for various problems, but most of them are developed for classical predictive problems (i.e. scalar response and functional covariates) [2,4,8,9,12]. For the concurrent model, general solutions to the regularizations are studied in [5,11]; however, they are very computationally intensive due to the model complexity.

More than often, highly complex models are not the most practical ones for real-world applications. In our study, we chose to focus on a simple but very pragmatic family of the concurrent model, namely, the *constant coefficient* models. Specifically, we focus on the following regularized regression model

$$
\mathbf{y}(t) = \mathbf{X}(t)\boldsymbol{\beta} + \boldsymbol{\epsilon}(t), \qquad \hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \|\mathbf{y}(t) - \mathbf{X}(t)\boldsymbol{\beta}\|^2_{\mathcal{H}^n} + \mathrm{PEN}(\boldsymbol{\beta}).
\tag{2.3}
$$

Here $\mathbf{y}(t) = (y_1(t), \ldots, y_n(t))' \in \mathcal{H}^n$ is the collection of observed functional responses and

$$\mathbf{X}(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) & \ldots & x_{1p}(t) \\ x_{21}(t) & \ldots & \ldots & x_{2p}(t) \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1}(t) & x_{n2}(t) & \ldots & x_{np}(t) \end{pmatrix} \in \mathcal{H}^{n \times p}$$

is the collection of functional covariates. In this model, the regression coefficients $\boldsymbol{\beta}$ is a constant vector over time, which is clearly a special case of functional regression models with time-varying coefficients because we can represent the constant $\boldsymbol{\beta}$ as $\boldsymbol{\beta}(t) = \boldsymbol{\beta} \cdot 1(t)$. $\mathcal{H}$ is a suitable Hilbert space of functions, such as $L^2([t_0, t_1])$, for $t \in [t_0, t_1]$ equipped with inner product

$$\langle x(t), y(t) \rangle_{\mathcal{H}} := \int_{t_0}^{t_1} x(s)y(s)dt$$

and $\mathcal{H}^n := \bigoplus_{i=1}^n \mathcal{H}_i$ is the direct sum of n copies of $\mathcal{H}_i \equiv \mathcal{H}$, with the following induced inner product and squared norm

$$\langle \mathbf{x}(t), \mathbf{y}(t) \rangle_{\mathcal{H}} := \sum_{i=1}^n \langle x_i(t), y_i(t) \rangle_{\mathcal{H}}, \qquad \|\mathbf{y}(t)\|_{\mathcal{H}^n}^2 := \sum_{i=1}^n \langle y_i(t), y_i(t) \rangle_{\mathcal{H}}.$$

Another example is a Sobolev space [1, 14] $W^{2,2}([t_0, t_1])$ equipped with the following inner product

$$\langle x(t), y(t) \rangle_{\mathcal{H}} := \int_{t_0}^{t_1} x(s)y(s)dt + \lambda_r \int_{t_0}^{t_1} D^2 x(s) D^2 y(s)dt. \qquad (2.4)$$

Apparently, in Equation (2.4), the first term is the standard L^2-inner product. The second part is the *roughness penalty* that is used in penalized spline regressions.

Now, we can establish an equivalence relationship for model (2.3) similar to that presented in Lemma 2.1. Denote

$$\Sigma := \langle \mathbf{X}(t), \mathbf{X}(t) \rangle \in M_{p \times p}, \qquad \Sigma_{jk} := \sum_{i=1}^n \langle x_{ij}(t), x_{ik}(t) \rangle_{\mathcal{H}},$$

$$\boldsymbol{\gamma} := \langle \mathbf{X}(t), \mathbf{y}(t) \rangle \in \mathbb{R}^p, \qquad \gamma_j := \sum_{i=1}^n \langle x_{ij}(t), y_i(t) \rangle_{\mathcal{H}}.$$

By applying eigen-decomposition, the inner product matrix Σ can be expressed as

$$\Sigma = T \cdot \Lambda \cdot T', \quad \text{with} \quad T \in O(\mathbb{R}^p), \quad \Lambda \in \text{diag}(p),$$

where $O(\mathbb{R}^p)$ is an orthogonal group in a p-dimensional real space, and $\text{diag}(p)$ is a $p \times p$ diagonal matrix. Next, we define

$$\tilde{\mathbf{X}} := \Lambda^{1/2} \cdot T' \in M_{p \times p},$$
$$\tilde{\mathbf{y}} := (\tilde{\mathbf{X}}')^{-1} \cdot \boldsymbol{\gamma} = \Lambda^{-1/2} \cdot T' \cdot \boldsymbol{\gamma} \in \mathbb{R}^p, \qquad (2.5)$$

and $\tilde{\boldsymbol{\beta}}$ the solution of the following optimization problem

$$\tilde{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg \min} \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}} \boldsymbol{\beta}\|_{\mathbb{R}^p}^2 + \text{PEN}(\boldsymbol{\beta}). \qquad (2.6)$$

Theorem 2.2. *The functional model defined by Equation* (2.3) *and the multivariate model defined by Equation* (2.6) *are equivalent in the sense that the optimizers of these two models are identical:* $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$.

Proof. By construction, $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} = \Sigma$ and $\tilde{\mathbf{X}}'\tilde{\mathbf{y}} = \boldsymbol{\gamma}$. The rest is similar to the proof of Lemma 2.1. □

Theorem 2.2 provides a bridge between a relatively difficult problem (the regularized functional regression problem defined by Equation (2.3)) and a relatively easy problem (the regularized multivariate model defined by Equation (2.6)). Furthermore, it may be the case that most of the eigenvalues of the cross-inner-product matrix Σ is near zero in real data analysis [13, 16]. In this case, we may replace all small eigenvalues (e.g. $\lambda_k \leqslant 0.01 \times \mathrm{tr}(\Lambda)$, or $\leqslant 1\%$ of the sum of eigenvalues) by zero; so there are only $q \ll p$ nonzero eigenvalues remained for the equivalent data. Then in Equation (2.5), Λ is replaced by Λ_q and T is replaced by T_q, where Λ_q consists of those q nonzero eigenvalues and T_q consists of the corresponding q columns of T. In this way, we still have $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} \approx \Sigma$ and $\tilde{\mathbf{X}}'\tilde{\mathbf{y}} \approx \boldsymbol{\gamma}$; but the equivalent dataset can be significantly smaller: $\tilde{\mathbf{X}} \in M_{q \times p}$ and $\tilde{\mathbf{y}} \in \mathbb{R}^q$.

An implementation of the above equivalence can be found in the `equiv.regression()` function in R package FUNNEL [16].

3. Equivalence II: Weighted principal component regression

Principal component regression (PCR) is another popular choice for regularizing the regression problem in high-dimensional space. Instead of regressing on the original covariates with high dimensionality and possibly severe collinearity, PCR conducts principal component analysis (PCA) on the multivariate covariate space first and use the leading principal components (PCs) as the surrogate regressors. Using PCA, each PC can be expressed as a linear combination of the original covariates; in the end, the estimated regression coefficients for the PC regressors can be linearly transformed back as coefficients for the original covariates. PCR is particularly useful for overcoming the collinearity problem in the original covariates, since PCR only includes the leading PCs that contribute large proportion of variation in the high-dimensional covariate space and the PCs are orthogonal to each other by definition.

PCA does a great job in dimension reduction for multivariate data. However, in many biological applications, sample sizes are usually very limited. Even with dimension reduction, the required number of PCs for explaining a desirable amount of variation is still more than the collected sample size. Penalty terms again need to be placed in the PC regression problem. Besides, an important feature of PCs is that they are decreasingly ordered by the amount of variation explained by each PC. Regularized regression problem like Equation (1.1) penalizes each covariates equally: the penalty weights equal λ_1 collectively for the L1-norm of the regression coefficients and λ_2 for the L2-norm. With prior knowledge about the ordering of the PCs, ideally we would incorporate differential weighting scheme for the penalty, i.e. assigning less penalty to the regression coefficients corresponding to the PCs that explain more variations.

Without regularization, it is easy to incorporate weights in covariates. Let $\mathbf{W} = \mathrm{diag}(w_1, w_2, \ldots, w_p)$ be a diagonal weighting matrix. The solution can be expressed as

$$\hat{\boldsymbol{\beta}}_{\mathbf{W}} := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\mathbf{W}\boldsymbol{\beta}\|^2 = (\mathbf{W}\mathbf{X}'\mathbf{X}\mathbf{W})^{-1}\mathbf{W}\mathbf{X}'\mathbf{y}$$

$$= \mathbf{W}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{W}^{-1}\hat{\boldsymbol{\beta}}.$$

Here $\hat{\boldsymbol{\beta}}$ is $\arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$, the OLS solution of the original regression problem without weights.

Next, we are going to establish an equivalence relationship that balances the different scaling weights for the regression coefficients in the least squares part as well as the regularization part.

Consider two optimization problems as follows.

- Optimization problem 1:

$$\hat{\boldsymbol{\beta}}^{(1)} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left(\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda_1 \sum_{j=1}^{p} w_j^{-1}|\beta_j| + \lambda_2 \sum_{j=1}^{p} w_j^{-2}\beta_j^2 \right).$$

- Optimization problem 2:

$$\hat{\boldsymbol{\beta}}^{(2)} = \mathbf{W} \cdot \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left(\|\mathbf{y} - \mathbf{X}\mathbf{W}\boldsymbol{\beta}\|^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right).$$

Theorem 3.1. *The above two optimization problems are equivalent:* $\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}^{(2)}$.

Proof. Let $\tilde{\boldsymbol{\beta}} = \mathbf{W}^{-1}\boldsymbol{\beta} = (w_1^{-1}\beta_1, \ldots, w_p^{-1}\beta_p)$. Obviously,

$$\|\mathbf{y} - \mathbf{X}\mathbf{W}\tilde{\boldsymbol{\beta}}\|^2 + \lambda_1 \sum_{j=1}^{p} |\tilde{\beta}_j| + \lambda_2 \sum_{j=1}^{p} \tilde{\beta}_j^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \lambda_1 \sum_{j=1}^{p} w_j^{-1}|\beta_j| + \lambda_2 \sum_{j=1}^{p} w_j^{-2}\beta_j^2.$$

$\square$

The essence of Theorem 3.1 is that for a regularization problem in which the penalties associated with all covariates are different, we can always use a scaling transformation of the covariates ($\mathbf{X} \to \mathbf{X}\mathbf{W}$) to turn it into a standard regularized regression problem in which all scaled covariates ($\mathbf{X}\mathbf{W}$) are associated with exactly the same penalty. With this theorem, researchers no longer have to write new code for elastic-net regression to deal with different penalties for different covariates.

Now we can describe an algorithm of the weighted principal component regression (WPCR) as follows.

Algorithm 3.2. *Weighted principal component regression (WPCR):*

(1) *Conduct PCA on* $\mathbf{X}$. *Denote the first* K *eigenvectors as*

$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_K],$$

where each $\mathbf{t}_k \in \mathbb{R}^n$ *is an eigenvector with unit Euclidean norm. Denote the first* K *eigenvalues as* $\boldsymbol{\mu} := \mu_1, \mu_2, \ldots, \mu_K$.

(2) *Scale these eigenvectors so that*

$$\tilde{\mathbf{T}} = [\mu_1 \mathbf{t}_1, \mu_2 \mathbf{t}_2, \ldots, \mu_K \mathbf{t}_K] = \mathbf{T}\mathbf{W}.$$

(3) *Conduct regularized regression* $\mathbf{y} = \tilde{\mathbf{T}}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$ *with* λ_1 *and* λ_2; *and then scale the coefficients to* $\tilde{\boldsymbol{\gamma}} := \mathbf{W}\boldsymbol{\gamma}$.

Therefore, $\tilde{\boldsymbol{\gamma}}$ is well-balanced regression coefficients for WPCR with regularization.

4. Discussion

Regularization techniques are widely used in high-dimensional data analysis. Many of the regularization techniques and the shrinkage effects are closely related to each other and have potential equivalence forms. For example, PCR and ridge regression [3], shrinkage effect in ridge regression and linear mixed-effect model [10]. Through well-defined equivalence relationships, difficult and complex models can be transformed to simpler models, and then be solved more efficiently by using mature algorithms proposed for those simpler models.

High-dimensional data can arise in many different ways. For example, functional data can be viewed as objects in spaces with *infinite dimensions*. Often, we want to map these data into high-dimensional data so that standard regression techniques based on linear algebra can be applied without modification. In functional data analysis [14], the most commonly used technique to achieve this goal is through discretization. However, to what level the discretization should be is always a dilemma. If the discretization is sparse, then important functional features may be lost; if the discretization is densely defined on a grid of many points, the computational burden will be very high in the high-dimensional space. In this paper, we introduce the idea of establishing an equivalence relationship through defining appropriate metrics for the different data spaces, e.g. the inner product norm for the L^2-space and the Euclidean norm for the p-dimensional real space. Using the equivalence relationship, both the information loss can be eliminated and the computational efficiency can be greatly boosted.

We also present another type of equivalence in which different weights of both L1 and L2 penalties for different covariates can be "translated" into an equivalent optimization problem in which the penalty is a constant for all covariates but the covariates are scaled. With such equivalence results, researchers can apply existing algorithms designed for constant penalty to regression problems with different penalties, thus eliminate the need for writing specific code for latter cases. An algorithm for WPCR is given as an application of this Theorem.

References

[1] R. A. Adams and J. J. Fournier, *Sobolev Spaces*, 2nd ed., Academic Press, 2003.

[2] J. A. Collazos, R. Dias and A. Z. Zambom, *Consistent variable selection for functional regression models*, Journal of Multivariate Analysis **146** (2016), 63–71.

[3] J. Friedman, T. Hastie and R. Tibshirani, *The Elements of Statistical Learning*, vol. 1, Springer Series in Statistics, New York, 2001.

[4] J. Gertheiss, A. Maity and A.-M. Staicu, *Variable selection in generalized functional linear models*, Stat **2** (2013), 86–101.

[5] J. Harezlak, B. A. Coull, N. M. Laird, S. R. Magari and D. C. Christiani, *Penalized solutions to functional regression problems*, Computational Statistics & Data Analysis **51** (2007), 4911–4925.

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

[7] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning: with Application in R*, Springer Texts in Statistics **103**, Springer, 2013.

[8] G. M. James, J. Wang and J. Zhu, *Functional linear regression that's interpretable*, The Annals of Statistics **37** (2009), 2083–2108.

[9] E. R. Lee and B. U. Park, *Sparse estimation in functional linear regression*, Journal of Multivariate Analysis **105** (2012), 1–17.

[10] Y. Muñoz Maldonado, *Mixed models, posterior means and penalized least-squares*, in: R. Javier (ed.), Optimality, The third Erich L. Lehmann symposium, Rice University, Houston, TX, USA, May 16–19, 2007, Institute of Mathematical Statistics Lecture Notes – Monograph Series **57**, 2009, 216–236.

[11] H. Matsui, S. Kawano, S. Konishi, *Regularized functional regression modeling for functional response and predictors*, Journal of Math-for-Industry **1** (2009), 17–25.

[12] H. Matsui and S. Konishi, *Variable selection for functional regression models via the L_1 regularization*, Computational Statistics & Data Analysis **55** (2011), 3304–3310.

[13] X. Qiu, S. Wu and H. Wu, *A new information criterion based on Langevin mixture distribution for clustering circular data with application to time course genomic data*, Statistica Sinica **25** (2015), 1459–1476.

[14] J. O. Ramsay, *Functional data analysis*, in: Encyclopedia of Statistical Sciences, Wiley Online Library, 2006, 8 pp.

[15] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, Series B **58** (1996), 267–288.

[16] Y. Zhang, D. J. Topham, J. Thakar and X. Qiu, *FUNNEL-GSEA: FUNctioNal ELastic-net regression in time-course gene set enrichment analysis*, Bioinformatics **33** (2017), 1944–1952.

[17] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society, Series B **67** (2005), 301–320.

Yun Zhang, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA
e-mail: `yun_zhang@urmc.rochester.edu`

Juilee Thakar, Department of Biostatistics and Computational Biology, and Department of Microbiology and Immunology, University of Rochester, Rochester, NY 14642, USA
e-mail: `juilee_thakar@urmc.rochester.edu`

David J. Topham, Department of Microbiology and Immunology, University of Rochester, Rochester, NY 14642, USA
e-mail: `david_topham@urmc.rochester.edu`

Ann R. Falsey, Department of Medicine, University of Rochester Medical Center and Rochester General Hospital, Rochester, NY 14621, USA
e-mail: `ann_falsey@urmc.rochester.edu`

Delu Zeng, Department of Information and Computational Science, School of Mathematics, South China University of Technology, Guangzhou, China
e-mail: `dlzeng@scut.edu.cn`

Xing Qiu, Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA
e-mail: `xing_qiu@urmc.rochester.edu`