# A sparsity regularization for functional linear discriminant analysis

Juhyun Park [1], Jeongyoun Ahn [2] & Yongho Jeon [3]

[1] *Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K., juhyun.park@lancaster.ac.uk*
[2] *Department of Statistics, University of Georgia, Athens GA 30602, U.S.A. jyahn@uga.edu*
[3] *Department of Applied Statistics, Yonsei University, Seoul 03722, South Korea, yhjeon@yonsei.ac.kr*

**Résumé.** Pour le problème de la classification fonctionnelle, il est bien connu que l'analyse discriminante linéaire fonctionnelle peut obtenir une classification parfaite si la dimensionnalité infinie est bien exploitée. Néanmoins, les données fonctionnelles étant intrinsèquement infinies, la prise en compte de la réduction de dimensions joue un rôle crucial dans sa réalisation. Les techniques standard de réduction des dimensions basées sur l'analyse fonctionnelle en composantes principales et les méthodes des moindres carrés partiels sont facilement disponibles à cette fin. D'autre part, il est de plus en plus nécessaire d'intégrer *l'explicabilité* dans la formulation de l'analyse statistique, ce qui tend à favoriser une solution simple et peu dense. Cette considération est bien développée pour les données de dimension finie avec une pénalité de type lasso ($\ell_1$), mais sa contrepartie de dimension infinie ($L^1$) est rarement étudiée pour les données fonctionnelles. Dans cet article, nous reformulons l'analyse discriminante linéaire fonctionnelle en tant que problème de régularisation avec pénalité appropriée. L'utilisation de la formule de pénalisation présente l'avantage supplémentaire de pouvoir incorporer certaines contraintes structurelles dans des coefficients fonctionnels, tels que la faible densité et la régularité, comme nous le souhaitons. En particulier, nous proposons une analyse discriminante linéaire fonctionnelle régularisée avec une pénalité de fragmentation fonctionnelle de $L^1$. Nous démontrons que notre formulation a une solution bien définie et une propriété de dispersion fonctionnelle souhaitable dans le sens de la sélection du domaine. De plus, notre solution converge vers un classificateur optimal. Des études numériques sont incluses pour évaluer les performances des échantillons finis et les comparer aux méthodes existantes.

**Mots-clés.** Explicabilité de l'apprentissage automatique, Modèles non-paramétriques, Parcimonie et grande dimension, Problèmes inverses et parcimonie, Sélection du domaine.

**Abstract.** For functional classification problem, it is well known that functional linear discriminant analysis can achieve a perfect classification, if the infinite-dimensionality is well exploited. Nevertheless, as functional data are inherently infinite-dimensional, consideration of dimension reduction plays a crucial role in its realization. Standard dimension reduction techniques based on functional principal component analysis or partial least squares methods are readily available for this purpose. On the other hand, there is an increasing need to incorporate *interpretability* within the formulation of our statistical analysis, which tends to favour a simple and *sparse* solution. Such consideration is well developed for finite dimensional data with lasso type ($\ell_1$) penalty but its infinite dimensional counterpart ($L^1$) is rarely studied for functional data. In this article, we reformulate the functional linear discriminant analysis as a regularization problem with appropriate penalty. An added advantage of using penalty formulation is the possibility of embedding some structural constraints in functional coefficient such as sparsity or smoothness as we desire. In particular, we propose a regularized functional linear discriminant analysis with $L^1$ functional sparsity penalty. We demonstrate that our formulation has a well defined solution and has a desirable functional sparsity property in the sense of domain selection. In addition, our solution is shown to converge to an optimal classifier. Numerical studies are included to assess finite sample performance and compare with existing methods.

**Keywords.** Domain selection, Functional sparsity, Interpretability, Inverse problems and regularization, Non-parametric models.

# 1 Summary

One of the common ways to collect high dimensional data is through automatic collection of records from continuous monitoring. Examples include spectrometric data in chemometrics and environmental data of pollutants from air quality monitoring. Dealing with such types of complex high dimensinal data requires us to consider the underlying structural constraints of the data. Incorporating this informaion in our analysis is important in developing parsimonious statistical models. In this regard, such types of data are better viewed as functional data in the sense that the underlying variable is of functional nature. Although functional data are inherently infinite dimensional, the possibility of utilising structural constraints of the functions such as continuity or smoothness offers an efficient way to deal with high dimensional problems. Many novel techniques are developed under functional data analysis framework. See Ramsay and Silverman (2005); Ferraty and Vieu (2006); Hsing and Eubank (2015) for more details.

In this work, we are interested in a classification problem for functional data as curves. Denote the functional predictor by $X$, observable on $\mathcal{I}$ and the class label by $Y$. Assume that $X$ is a member of two possible groups ($Y = 0$ or $Y = 1$). Suppose that $X(t)|_{Y=0}$ and $X(t)|_{Y=1}$ for $t \in \mathcal{I}$ are square integrable stochastic processes with mean function $\mu_0(t)$ and $\mu_1(t)$ respectively, and have common covariance function $\gamma(s,t) = \mathrm{cov}(X(s), X(t))$. Let $\pi_0 = \Pr(Y = 0)$, $\pi_1 = \Pr(Y = 1)$ and $\mu(t) = \mu_1(t) - \mu_0(t)$. We seek a classification rule that depends on the linear map with unknown direction $\beta$

$$F_\beta(X) = \int_{\mathcal{I}} X(t)\beta(t)\, dt = \langle X, \beta \rangle \,. \tag{1}$$

This defines a functional linear classification problem, where we wish to determine $\beta(\cdot)$ in such a way that the linear map yields a *good class separation*. Although simple, the linear classifier can achieve a perfect classification if the infinite-dimensionality is well exploited (Delaigle and Hall, 2012; Berrendero et al., 2018), a distinctive feature compared to finite dimensional problems. Delaigle and Hall (2012) demonstrate such phenomenon with a simple linear centroid classifier and suggest a practical representation using components obtained from functional principal component analysis and partial least squares. Kraus and Stefanucci (2017) propose alternative regularisation methods to compute the representation.

While *optimal* performance is an important criterion to consider, the increasing impact of statistical analysis on modern scientific investigations has created the need of careful consideration of *interpretability* of the outcomes of our analysis. Particular instances of interpretability are formulated under the *sparsity* regularization. For example, James et al. (2009) and Zhou et al. (2013) propose a lasso-type sparsity regularization for interpretability under regression setting and Tian and James (2013) under classification setting, advocating a simpler form of solution that contains zero regions. Although the development still relies on the discrete concept of sparsity whose property is not directly applicable to infinite dimensional setting (Kneip et al., 2016; Roche, 2018), these works demonstrate the importance of such consideration in infinite-dimensional problems. Functional formulation under non-parametric models is relatively scarce. Few exceptions are Wang and Kai (2015), Tu et al. (2018), Lin et al. (2017) and Hall and Hooker (2016), all of which are concerned with functional linear regression problems.

In this work, we seek an alternative approach to functional linear classification with a direct estimation method to address dimensionality, optimality and interpretability. We specifically utilize the results in Delaigle and Hall (2012) and reformulate it as a regularization problem with an appropriate choice of penalty function. An advantage of using penalty formulation is the possibility of incorporating various structural constraints in functional data such as functional sparsity and smoothness as we desire.

So far the regularization methods have been used mostly for either smoothness ($\int(f^{(m)})^2$) in regression (e.g. Kneip et al. (2016)) or invertibility ($\int f^2$) (Kraus and Stefanucci, 2017) in classification, with $L^2$-type penalty. The idea of sparsity as variable selection with $\ell_1$-type penalty is actively developed in discrete high dimensional setting, but much less for the infinite-dimensional functional setting. It is much more difficult to extend it with a fundamentally discrete notation of sparsity. We argue that the counterpart of the $\ell_1$ penalty is the functional $L^1$ penalty, but the latter is hardly used in statistical inference framework. In our formulation, we extend the usual regularization methods with differentiable $L^2$-type penalty to non-differentiable $L^1$ penalty function to address functional sparsity and domain selection, demonstrating its utility in infinite-dimensional problems arising in functional data analysis.

# References


Berrendero, J. R., Cuevas, A., and Torrecilla, J. L. (2018). On the use of reproducing kernel hilbert spaces in functional classification. *Journal of the American Statistical Association*, 113(523):1210–1218.

Delaigle, A. and Hall, P. (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society, Series B*, 74(2):267–286.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer series in statistics. Springer.

Hall, P. and Hooker, G. (2016). Truncated linear models for functional data. *Journal of the Royal Statistical Society: Series B*, 78(3):637–653.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley series in probability and statistics. Wiley.

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

Kneip, A., Poß, D., and Sarda, P. (2016). Functional linear regression with points of impact. *Annals of Statistics*, 44(1):1–30.

Kraus, D. and Stefanucci, M. (2017). Classification of functional fragments by regularized linear classifiers with domain selection. *ArXiv e-prints*.

Lin, Z., Cao, J., Wang, L., and Wang, H. (2017). Locally sparse estimator for functional linear regression models. *Journal of Computational and Graphical Statistics*, 26(2):306–318.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer series in statistics. Springer, 2nd. edition.

Roche, A. (2018). Variable selection and estimation in multivariate functional linear regression via the lasso. Technical report, hal-01725351.

Tian, T. S. and James, G. M. (2013). Interpretable dimension reduction for classifying functional data. *Computational Statistics and Data Analysis*, 57:282–296.

Tu, C. Y., Park, J., and Wang, H. (2018). Estimation of functional sparsity in nonparametric varying coefficient models for longitudinal data analysis. *Statistica Sinica*. in press.

Wang, H. and Kai, B. (2015). Functional sparsity: Global versus local. *Statistica Sinica*, 25:1337–1354.

Zhou, J., Wang, N.-Y., and Wang, N. (2013). Functional linear model with zero-value coefficient function at sub-regions. *Statistica Sinica*, 23:25–50.