Statistics and Its Interface Volume 15 (2022) 261–265

# Rejoinder of "Estimation of Hilbertian varying coefficient models"*

Young Kyung Lee, Byeong U. Park†,
Hyerim Hong, and Dongwoo Kim

We thank all discussants for their insightful comments on our paper. They highlight some important points of the paper that need more detailed discussion and clarification, and also make valuable suggestions for future study. We categorize the issues raised by the discussants into three groups: extensions to other models and data types; statistical inference; real data example. Below, we briefly address them.

## 1. EXTENSIONS TO OTHER MODELS AND DATA TYPES

Du mentions a possible extension to models with general interaction terms, which turn out to be additive regression models with bivariate component maps. In fact, the estimation of a bivariate additive regression model with a Euclidean response was studied by Lee (2017). More general cases with multivariate additive regression models are covered by a recent study, Jeon et al. (2021), where additive models with a Hilbertian response and finite-dimensional Hilbertian or manifold-valued predictors are considered. Huang, Wang and Zhou also toss a question about including the interaction between temperature and cloudiness to better explain electricity consumption pattern in the real data example (Section 4.3). There might be a certain level of such interaction. But, we simply neglected it in the analysis of the electricity consumption data since our main motivation for the data example was to see how the nominal variable, indicating weekdays or weekend, improves the accuracy of predicting electricity load trajectory during a day. Adding interaction terms such as $\mathbf{f}_{12,1}(X_1, X_2)$ and $Z \odot \mathbf{f}_{12,2}(X_1, X_2)$ to the model (4.3) may improve the prediction accuracy further depending on how strong the interaction between $X_1$ and $X_2$ is.

Du brings up another extension to models with functional predictors, which is a very challenging problem. This is also commented by Huang, Wang and Zhou regarding the real data example in the paper. Basically, one can extend the methodology to the case in a way similar to the one discussed in Section 6 of Jeon and Park (2020) for additive models with infinite-dimensional predictors. One can obtain a version of Theorem 1 for the existence of the resulting estimator and backfitting algorithm. However, there are at least two difficulties in the extension. One is in the implementation of the extended method. Our method requires Bochner integration with respect to a reference measure on the infinite-dimensional space that embodies functional data. For Euclidean spaces, Lebesgue measures take the role of a reference measure. As far as we know, there is no such thing for infinite-dimensional Hilbert or Banach spaces. Another difficulty stems from the fact that the nonparametric estimation of a function basically deteriorates as the dimension of the space on which the function is seated gets large. Additive models and varying coefficient models are two important classes of structured models that enable one to circumvent the curse of dimensionality. However, if the component maps themselves are defined on high-dimensional spaces, then one still faces the dimensionality problem with structured models. Functional data are infinite-dimensional, in which case we do not even think it is possible to find a consistent estimator. The three recent works (Sun et al., 2018; Reimherr et al., 2018; Cui et al., 2020) on function-to-function regression mentioned by Du do not have the difficulty. Among them, Sun et al. (2018) and Cui et al. (2020) are about parametric functional models. The other one, Reimherr et al. (2018), is about a nonparametric functional model that can be written as

$$\mathrm{E}\big(Y(\cdot)|X(\cdot)\big) = \int_0^1 g(\cdot, s, X(s))\, ds, \tag{1}$$

where $g$ is the unspecified target of estimation. Adapting the latter problem to our Hilbertian framework, $\mathbf{Y} = Y(\cdot)$ is a Hilbertian response and the target $\mathbf{g}(\cdot,\cdot)$ defined by $\mathbf{g}(s,u) = g(\cdot, s, u)$ is a bivariate Hilbertian map. Thus, the nonparametric Hilbertian map $\mathbf{g}$ in the model (1) is actually seated on the two-dimensional space, not on the infinite-dimensional space where $\mathbf{X} \equiv X(\cdot)$ takes values. In general, for functional or infinite-dimensional predictors, there should be a dimension reduction stage before implementing a nonparametric technique. The latter approach was actually taken by many researchers (Han et al., 2018; Park et al., 2018, e.g.).

*Research of Young Kyung Lee was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIP) (No. NRF-2021R1A2C1003920). Research of Byeong U. Park was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2019R1A2C3007355).
†Corresponding author.

Cheng and Dai discuss a number of recent works on regression with density-valued and manifold-valued data. These areas are also evolving rapidly in recent years due to today's new data environments. As for density-valued random objects, we note that the transformation approach discussed by Cheng has been applied to additive regression by Han et al. (2020), for example. In fact, the space of density functions is a Hilbert space under the Aitchison geometry (Aitchison, 1986) and thus the case is covered by the general Hilbertian framework of Jeon and Park (2020). As for manifold-valued responses, the four works (Chau and von Sachs, 2020; Cornea et al., 2017; Dai and Muller, 2018; Dai et al., 2020) mentioned by Dai are not for structured nonparametric regression. The first study on structured nonparametric regression for mainfold-valued responses is done by Lin et al. (2021), which proposes an intrinsic group additive regression model for responses taking values in the space $\mathcal{S}^+$ of symmetric positive-definite matrices. The latter work exploits the abelian group structure of $\mathcal{S}^+$ endowed by the Log-Cholesky (Lin, 2019) and Log-Euclidean (Arsigny et al., 2007) metrics. The group additive model is then transformed into an additive model on tangent spaces via the Riemannian logarithmic map. The extension is considered to pave a way for a general approach to manifold additive modeling. Cheng raises an identification issue for the linear and nonlinear parts in the model (2.6). We note that they are well identified under the constraints (2.5), as demonstrated by Proposition 1 and the discussion that follows.

## 2. STATISTICAL INFERENCE

Du raises a very interesting question on how to make statistical inference for the model studied. It is common in nonparametric inference that the estimation of unknown quantities in the asymptotic distribution of an estimator is more complicated and difficult than the original problem. In the current problem, the asymptotic distribution demonstrated in Theorem 3 involves $\boldsymbol{\beta}_j$, which do not have a closed form, the covariance operator $\mathcal{C}_j$ and the second Fréchet derivatives of the component maps $\mathbf{f}_{j,k}$. To use the asymptotic distribution for inference, these unknown quantities need to be estimated, which makes the problem even harder since the estimation of the derivatives of a function, for example, is more difficult than that of the original function. As Du suggests, one might think of a bootstrap procedure to estimate the distributions of the estimators of the component or regression maps. This is doable in the following way. In the construction of a confidence set, for example, for a component map $\mathbf{f}_{j,k}$ at a level $(1-\alpha)$ for $0<\alpha<1$, one basically seeks for the quantile $c_\alpha$ such that

$$\mathrm{P}\left(\int_0^1 \|\hat{\mathbf{f}}_{j,k}(x_j)\ominus \mathbf{f}_{j,k}(x_j)\|_{\mathbb{H}}^2\,dx_j \;\le\; c_\alpha\right)=1-\alpha.$$

One may approximate $c_\alpha$ by $\hat{c}_\alpha$ such that

$$\mathrm{P}\left(\int_0^1 \|\hat{\mathbf{f}}^*_{j,k}(x_j)\ominus \hat{\mathbf{f}}_{j,k}(x_j)\|_{\mathbb{H}}^2\,dx_j \;\le\; \hat{c}_\alpha \,\middle|\, \mathbf{X}_i,\mathbf{Y}_i, 1\le i\le n\right)$$
$$=1-\alpha,$$

where $\hat{\mathbf{f}}^*_{j,k}$ is the bootstrap version of $\hat{\mathbf{f}}_{j,k}$ constructed from a bootstrap sample $\{(\mathbf{X}^*_i,\mathbf{Y}^*_i):1\le i\le n\}$ of $\{(\mathbf{X}_i,\mathbf{Y}_i):1\le i\le n\}$. A confidence set for $\mathbf{f}_{j,k}$ at level $(1-\alpha)$ is then given by $\left\{\mathbf{g}_{j,k}:\int_0^1\|\hat{\mathbf{f}}_{j,k}(x_j)\ominus \mathbf{g}_{j,k}(x_j)\|_{\mathbb{H}}^2\,dx_j \;\le\; \hat{c}_\alpha\right\}$. Two immediate questions with such a confidence set are whether the bootstrap procedure works theoretically and how to visualize the confidence set. These issues deserve further study.

The bootstrapping idea described in the above paragraph can be also applied to testing problems. This is in line with a suggestion made by Zhou, Guo and Zhang. Specifically, for testing $H_0:\mathbf{f}_{j,k}=\mathbf{0}$, one basically wants to find $\gamma_\alpha$ such that

$$\mathrm{P}_{H_0}\left(\int_0^1 \|\hat{\mathbf{f}}_{j,k}(x_j)\|_{\mathbb{H}}^2\,dx_j \;>\; \gamma_\alpha\right)=\alpha,$$

where $\mathrm{P}_{H_0}$ is the probability measure corresponding to the null hypothesis $H_0$. A bootstrap estimate $\hat{\gamma}_\alpha$ of $\gamma_\alpha$ may be defined by

$$\mathrm{P}\left(\int_0^1 \|\hat{\mathbf{f}}^*_{j,k}(x_j)\|_{\mathbb{H}}^2\,dx_j \;>\; \hat{\gamma}_\alpha \,\middle|\, \mathbf{X}_i,\mathbf{Y}_i, 1\le i\le n\right)=\alpha,$$

where the bootstrap version $\hat{\mathbf{f}}^*_{j,k}$ is obtained from the backfitting equations at (2.12) based on a bootstrap sample $\{(\mathbf{X}^*_i,\mathbf{Y}^*_i):1\le i\le n\}$ from an empirical model that reflects $H_0$. With such a bootstrap quantile $\hat{\gamma}_\alpha$, one rejects $H_0$ if

$$\int_0^1 \|\hat{\mathbf{f}}_{j,k}(x_j)\|_{\mathbb{H}}^2\,dx_j > \hat{\gamma}_\alpha.$$

Zhou, Guo and Zhang also comment on the connection of our model (2.1) to an ANOVA problem with density functions as data objects from $k$ sub-populations. The observations from $k$ sub-populations may be put into a single model using dummy variables. However, the ANOVA model does not fit into our framework since the mean density functions of the sub-populations are considered as constants in the space of density functions.

## 3. REAL DATA EXAMPLE

Huang, Wang and Zhou make various comments on the real data example. We respond to some of them here that are left not being addressed in the previous sections.

First, as for adding the seasonal indicator, we did not consider it at the time of writing the paper because we thought that the seasonal effect is well explained by temperature and cloudiness, the predictors we already considered in the model (4.3). In preparing this rejoinder, we checked it by fitting a model where the seasonal effect was added.

Let $X_4$ be the indicator of weekdays, $X_5$, $X_6$ and $X_7$ be the indicators of summer (June–August), fall (September–November) and winter (December–February), respectively. We fitted the model,

$$\begin{aligned}\mathrm{E}(\mathbf{Y}|\mathbf{X}) = f_{1,3}(X_1) &\oplus f_{2,3}(X_2)\\ &\oplus X_4 \odot \Big(f_{1,4}(X_1) \oplus f_{2,4}(X_2)\Big)\\ &\oplus X_5 \odot \Big(f_{1,5}(X_1) \oplus f_{2,5}(X_2)\Big)\\ &\oplus X_6 \odot \Big(f_{1,6}(X_1) \oplus f_{2,6}(X_2)\Big)\\ &\oplus X_7 \odot \Big(f_{1,7}(X_1) \oplus f_{2,7}(X_2)\Big).\end{aligned}$$

We found that the value of ASPE with the above model was 0.0081, which is worse than 0.0052 with the model (4.3). One reason for this is that, as we note above, each season in Korea has a unique climate and thus its effect is sufficiently well accounted by temperature and cloudiness. Including the additional variables in the model seemed to increase only the variability of the estimated model. Another, which also seemed to worsen the ASPE performance, is that the presence of the seasonal variables actually hinders proper tuning for the bandwidths. To see this, we first note that adding the seasonal indicators increases the dimension of the matrices $\hat{\mathbf{M}}_{jj}(x_j)$, defined at (2.10), from $2 \times 2$ to $5 \times 5$. Let $I_{k,1}$ for $1 \le k \le 4$ be the index set of $i$ for $\mathbf{Y}_i$ corresponding to weekdays in the $k$th season, where $k = 1$ for spring, $k = 2$ for summer, $k = 3$ for fall and $k = 4$ for winter. Also, let $I_{k,0}$ be the index set for weekend in the $k$th season. Put $S_{k,l}(x_j) = \sum_{i \in I_{k,l}} K_{h_j}(x_j, X_{ij})$ for $j = 1, 2$, $l = 0, 1$ and $k = 1, \ldots, 4$. Then, we find

$$\begin{aligned}&\det\Big(n \cdot \hat{\mathbf{M}}_{jj}(x_j)\Big)\\ &= \sum_{k=1}^{4} S_{k,0}(x_j) \cdot S_{k,1}(x_j) \cdot \prod_{k'=1,\neq k}^{4} \Big(S_{k',0}(x_j) + S_{k',1}(x_j)\Big).\end{aligned}$$

Thus, $\hat{\mathbf{M}}_{jj}(x_j)$ is invertible if and only if the right hand side of the above equation is not zero. Note that the backfitting equation at (2.12) is well defined only if $\hat{\mathbf{M}}_{jj}(x_j)$ is invertible for all $x_j$ and for all $j$. The constraint, $\det\big(n \cdot \hat{\mathbf{M}}_{jj}(x_j)\big) \neq 0$ for all $x_j \in [0, 1]$ and $j = 1, 2$, certainly restricts the bandwidth ranges on which we search for the bandwidths $h_j$ by the CBS scheme mentioned in Section 4.2. In fact, the CBS bandwidths chosen for fitting the model (4.3) were $h_1 = 0.052$ and $h_2 = 0.2472$, while with the seasonal variables being added the CBS scheme gave $h_1 = 0.7696$ and $h_2 = 0.4513$, which seemed to give the worse ASPE performance due to excessive bias.

Huang, Wang and Zhou present a question on the difference between fitting (4.3) and two separate additive models, one for the weekday and the other for the weekend data. The two approaches are essentially the same as far as the models are concerned. The only difference is in tuning the smoothing parameters. In fitting the two separate additive models, one basically tunes four bandwidths, one for each of four associated component maps. Let $h_{1,1/2}$ and $h_{2,1/2}$ denote the two bandwidths used to fit the separate additive model for the weekday data. Likewise, let $h_{1,-1/2}$ and $h_{2,-1/2}$ be those for the weekend data. Then, the four bandwidths $h_{1,\pm 1/2}$ and $h_{2,\pm 1/2}$ are not for the individual components, $\hat{\mathbf{f}}_{1,3}$, $\hat{\mathbf{f}}_{1,4}$, $\hat{\mathbf{f}}_{2,3}$ and $\hat{\mathbf{f}}_{2,4}$, but for the four combined maps,

$$\begin{aligned}&\hat{\mathbf{f}}_{1,3} \oplus \Big(\frac{1}{2} \odot \hat{\mathbf{f}}_{1,4}\Big),\ \ \hat{\mathbf{f}}_{1,3} \ominus \Big(\frac{1}{2} \odot \hat{\mathbf{f}}_{1,4}\Big),\\ &\hat{\mathbf{f}}_{2,3} \oplus \Big(\frac{1}{2} \odot \hat{\mathbf{f}}_{2,4}\Big),\ \ \hat{\mathbf{f}}_{2,3} \ominus \Big(\frac{1}{2} \odot \hat{\mathbf{f}}_{2,4}\Big),\end{aligned}$$

respectively, under the coding scheme $Z = \pm 1/2$. For the existence of the estimators in fitting the two separate additive models, one basically needs

$$\begin{aligned}&\sum_{i=1:Z_i=1/2}^{n} K_{h_{j,1/2}}(x_j, X_{ij}) > 0,\\ &\sum_{i=1:Z_i=-1/2}^{n} K_{h_{j,-1/2}}(x_j, X_{ij}) > 0\end{aligned} \tag{2}$$

for all $x_j \in [0, 1]$ and for $j = 1, 2$. Thus, tuning the four bandwidths are to be done over a four dimensional grid that satisfies (2).

On the other hand, with the model (4.3), the two components $\mathbf{f}_{j,3}$ and $\mathbf{f}_{j,4}$ are estimated using the same bandwidth $h_j$ so that one tunes only two bandwidths $h_1$ and $h_2$. We note that, if one desires to use a different amount of smoothing for each of the component maps, $\hat{\mathbf{f}}_{1,3}$, $\hat{\mathbf{f}}_{1,4}$, $\hat{\mathbf{f}}_{2,3}$ and $\hat{\mathbf{f}}_{2,4}$, then one may employ the two-step procedure analyzed by Park et al. (2015). Now, writing (4.3) in the form of (2.10), we get

$$\mathrm{E}(\mathbf{Y}|\mathbf{X}) = \big(\mathbf{Z}^\top \odot \mathbf{f}_1(X_1)\big) \,\oplus\, \big(\mathbf{Z}^\top \odot \mathbf{f}_2(X_2)\big),$$

where $\mathbf{Z} = (1, Z)^\top$, $\mathbf{f}_1 = (\mathbf{f}_{1,3}, \mathbf{f}_{1,4})^\top$ and $\mathbf{f}_2 = (\mathbf{f}_{2,3}, \mathbf{f}_{2,4})^\top$. For the existence of the estimators in fitting the model (4.3), $\hat{\mathbf{M}}_{jj}(x_j)$ need to be invertible for all $x_j \in [0, 1]$ and for $j = 1, 2$. Note that $\hat{\mathbf{M}}_{jj}(x_j)$ is invertible if and only if

$$\begin{aligned}&n^{-1} \sum_{i=1}^{n} \big(\alpha_0 + \alpha_1 I(Z_i = 1/2)\big)^2 K_{h_j}(x_j, X_{ij}) = 0\\ &\text{implies}\quad \alpha_0 = \alpha_1 = 0.\end{aligned} \tag{3}$$

The relation between the conditions (2) and (3) depends on the sizes of $h_j$, $h_{j,1/2}$ and $h_{j,-1/2}$. Recall that the baseline kernel $K$ is assumed to vanish on $\mathbb{R} \setminus (-1, 1)$ and be positive on $(-1, 1)$. Thus, $K_h(x, u) > 0$ if and only if $|u - x| < h$, so that $K_h(x, u) > 0$ implies $K_{h'}(x, u) > 0$ if $h \le h'$. In case $h_j = h_{j,1/2} = h_{j,-1/2}$, it is not difficult to see that the conditions (2) and (3) are equivalent.

About the issue of normalization, we agree with Huang, Wang and Zhou that actual hourly consumption conveys more information than its normalized version that demonstrates only consumption pattern. The reason we worked on the normalized trajectories was that actual electricity load trajectories were not available. We were provided only with relative electricity loads, which were already normalized, although not to integrate to unity but to $24 \times 1{,}000$. If actual hourly trajectories are available, we may apply the model (4.3) with $\mathbb{H} = L^2[0, 24]$. We note that the normalization does not distort prediction since we are working on the correct metric introduced in Section 4.1. Also, it is easy to convert the 'density' (intensity) of electricity consumption over time during a day to the corresponding actual electricity consumption trajectory, simply by multiplying the total daily consumption, if the latter is available.

Regarding normalization after averaging versus averaging after normalization, we would like to note that the original data are $Y_i^{\mathrm{unsmth}}$ as described in Section 4.3 where $Z_i$ are monthly averages. If one does have raw data, then one might first normalize the consumption trajectories and then average out the normalized versions over the month. Although the latter is more natural as noted by Huang, Wang and Zhou, both give a summarized pattern of electricity consumption for each month. They are different but related. The difference between the two pre-processing procedures, normalization after averaging versus averaging after normalization, is actually small since daily weekday consumption patterns are similar within each month and so are daily weekend patterns. As for the comment that the data smoothing leading to $Y_i^{\mathrm{smth}}$ may introduce bias, we would say that the pre-smoothing step is necessary to apply our method to the data. The original data are electricity consumption measured on the discrete hourly time grid with possible error. Pre-smoothing is common in functional data analysis and is typically employed to retrieve unobserved true trajectories on a continuous time scale from observations on a discrete grid with errors. It would be interesting to analyze the effect of the pre-smoothing, however.

Huang, Wang and Zhou also comment on our coding scheme for weekdays versus weekend: $Z = \pm 1/2$ instead of $Z = 1/0$. Of course, the interpretation of each component map is different if $Z$ is coded in a different way. But, in any cases, $\mathbf{f}_{1,4}$ and $\mathbf{f}_{2,4}$ represent the effects of $Z$ on the nonlinear effects of $X_j$. The nonlinear effect of $X_j$ when $Z = 1/2$ equals $\mathbf{f}_{j,3}(X_j) \oplus (1/2) \odot \mathbf{f}_{j,4}(X_j)$ and the one when $Z = -1/2$ equals $\mathbf{f}_{j,3}(X_j) \ominus (1/2) \odot \mathbf{f}_{j,4}(X_j)$, so that the difference equals $\mathbf{f}_{j,4}(X_j)$. The same is true for the case where one takes the coding scheme $Z = 0/1$. The estimated maps $\hat{\mathbf{f}}_{j,4}$ for $j = 1$ and 2 are depicted on the two bottom panels in Figure 2. Furthermore, the two top panels show the estimated maps $\hat{\mathbf{f}}_{j,3} \oplus (1/2) \odot \hat{\mathbf{f}}_{j,4}$ for $j = 1$ and 2, and the two middle demonstrate $\hat{\mathbf{f}}_{j,3} \ominus (1/2) \odot \hat{\mathbf{f}}_{j,4}$. If one changes the coding scheme to $Z^* = 0/1$ and rewrite the model (4.3) as

$$\mathrm{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{f}^*_{1,3}(X_1) \oplus \mathbf{f}^*_{2,3}(X_2) \oplus Z^* \odot \Big(\mathbf{f}^*_{1,4}(X_1) \oplus \mathbf{f}^*_{2,4}(X_2)\Big),$$

then $\mathbf{f}^*_{j,3} \oplus \mathbf{f}^*_{j,4} = \mathbf{f}_{j,3} \oplus (1/2) \odot \mathbf{f}_{j,4}$ and $\mathbf{f}^*_{j,3} = \mathbf{f}_{j,3} \ominus (1/2) \odot \mathbf{f}_{j,4}$. Thus, the estimated effects of temperature and cloudiness for weekday and weekend are independent of coding scheme, so are their interpretation.

We close this rejoinder by giving a brief remark regarding the alternative model formulation suggested by Zhou, Guo and Zhang. We agree with them that the new formulation gives a clear picture about the model (2.1). However, we basically need the representation (2.1) to better describe the SBF method and to undertake the theoretical analysis of the method.

*Received 21 May 2021*

## REFERENCES


Aitchison, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability.* Chapman & Hall, London. MR0865647

Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal of Matrix Analysis and Applications* **29** 328–347. MR2288028

Chau, J. and von Sachs, R. (2020). Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *Journal of the American Statistical Association*, in print. MR4270027

Cornea, E., Zhu, H., Kim, P., and Ibrahim, J. G. (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society, Series B* **79** 463–482. MR3611755

Cui, X., Lin, H. and Lian, H. (2020). Partially functional linear regression in reproducing kernel Hilbert spaces. *Computational Statistics and Data Analysis.* https://doi.org/10.1016/j.csda.2020.106978. MR4101440

Dai, X., Lin, Z., and Müller, H.-G. (2020). Modeling longitudinal data on Riemannian manifolds. *Biometrics*, in print.

Dai, X. and Müller, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Annals of Statistics* **46** 3334–3361. MR3852654

Han, K., Müller, H.-G. and Park, B. U. (2018). Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple predictor functions. *Bernoulli* **24** 1233–1265. MR3706793

Han, K., Müller, H.-G. and Park, B. U. (2020). Additive functional regression for densities as responses. *Journal of the American Statistical Association* **115** 997–1010. MR4107695

Jeon, J. M. and Park, B. U. (2020). Additive regression with Hilbertian responses. *Annals of Statistics* **48** 2671–2697. MR4152117

Jeon, J. M., Park, B. U. and Van Keilegom, I. (2021). Additive regression for non-Euclidean responses and predictors responses. *Annals of Statistics*, in print.

Lee, Y. K. (2017). Nonparametric estimation of bivariate additive models. *Journal of Korean Statistical Society* **46** 339–348. MR3685573

Lin, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications* **40** 1353–1370. MR4032859

Lin, Z., Müller, H.-G. and Park, B. U. (2021). Additive models for symmetric positive-definite matrices, Riemannian manifolds and Lie groups. Submitted.

Park, B. U., Chen, C. J., Tao, W., Müller, H.-G. (2018). Singular additive models for function to function regression. *Statistica Sinica* **28** 2497–2520. MR3839871

Park, B. U., Mammen, E., Lee, Y. K. and Lee, E. R. (2015). Varying coefficient regression models: A review and new developments (with discussion). *International Statistical Review* **83** 36–76. MR3341079

Reimherr, M., Sriperumbudur, B. and Taoufik, B. (2018). Optimal prediction for additive function-on-function regression. *Electronic Journal of Statististics* **12** 4571–4601. MR3893421

Sun, X., Du, P., Wang, X. and Ma, P. (2018). Optimal penalized function-on-function regression under a reproducing kernel Hilbert space framework. *Journal of the American Statistical Association* **113** 1601–1611. MR3902232

Young Kyung Lee
Department of Information Statistics
Kangwon National University
South Korea
E-mail address: youngklee@kangwon.ac.kr

Byeong U. Park
Department of Statistics
Seoul National University
South Korea
E-mail address: bupark@snu.ac.kr

Hyerim Hong
Department of Statistics
Seoul National University
South Korea
E-mail address: hhong@snu.ac.kr

Dongwoo Kim
Department of Statistics
Seoul National University
South Korea
E-mail address: kdu91@snu.ac.kr