Comments on: Dynamic Relations for Sparsely Sampled Gaussian Processes*

Hervé Cardot

Received: date / Accepted: date

First of all, I would like to congratulate H-G. Müller and W. Yang for this very stimulating work which has potential applications in many domains. Dynamics in functional data analysis is essential and this makes clear that FPCA is a nice tool to address this issue. Being able to estimate and expand derivatives of a stochastic process when observing only a few point of the trajectories is really a challenging, I would say almost incredible, issue that is addressed with brio in this paper. Note that I am really impressed too by the performances of the package PACE since computations in such a sparse context need to be developed by experts.

My main comment deals with the expansion of the ν-th order derivative of $X(t)$ proposed in equation (3) by M-Y and I would like to suggest another point of view based on a direct representation of the derivatives themselves. When the target is the ν-th order derivative the process $X(t)$, I wonder if it would not be more interesting to perform directly the Karhunen-Loève expansion of $X^{(\nu)}(t)$. As a matter of fact the nice properties of expanding the process into *optimal deterministic orthonormal* basis is generally lost since the functions $\phi_k^{(\nu)}(t)$ are generally not orthogonal anymore (except for the special case of stationary processes for which the Fourier basis are eigenvectors of the covariance operator A_G). Maybe even more delicate is that one may need a large number of functions $\phi_k^{(\nu)}(t)$ to capture some important modes of variability of the ν-th order derivative process and there is no clear way to check, when considering equation (3) of M-Y truncated as in (10), what amount of the total variance is captured by a representation which only considers K components. Thus interpretation in terms of main modes of variability may be more difficult.

A direct calculus (see *e.g.* Loève, 1963, Chapter X) gives a simple link between the covariance function of $X^{(\nu_2)}(t)$ and $X^{(\nu_1)}(s)$ and the covariance function $G(t,s)$,

$$\operatorname{cov}\left(X^{(\nu_2)}(t), X^{(\nu_1)}(s)\right) = \frac{\partial^{\nu_2+\nu_1}}{\partial t^{\nu_2}\partial s^{\nu_1}} G(t,s). \tag{1}$$

* This comment refers to the invited paper available at: http://dx.doi.org/10.1007/s11749-006-122-1

Hervé Cardot
Institut de Mathématiques de Bourgogne, UMR CNRS 5584, Université de Bourgogne, 9 Av. Alain Savary, 21078 Dijon, France
E-mail: Herve.Cardot@u-bourgogne.fr

This means that it is possible to derive the basis functions of the Karhunen-Loève expansion of $X^{(\nu)}$ provided that we are able to estimate the following partial derivative $\frac{\partial^{2\nu}}{\partial t^\nu \partial s^\nu} G(t,s)$ of the covariance function of $X(t)$. Estimating this bivariate function and its trace norm, which is the total variance of $X^{(\nu)}$, can be done by considering a modification of the estimator proposed in (8) by M-Y. Then one can get directly the representation, $\operatorname{cov}\left(X^{(\nu)}(t), X^{(\nu)}(s)\right) = \sum_{k\geq 1} \lambda_{\nu,k}\ \phi_{\nu,k}(t)\ \phi_{\nu,k}(s)$, where now $\phi_{\nu,1}, \phi_{\nu,2}, \ldots$, are orthonormal eigenfunctions associated to the eigenvalues sorted in decreasing order, $\lambda_{\nu,1} \geq \lambda_{\nu,2} \geq \ldots \geq 0$, of the covariance operator which is the integral operator with kernel function $\frac{\partial^{2\nu}}{\partial t^\nu \partial s^\nu} G(t,s)$. The Karhunen-Loève representation of $X^{(\nu)}$ is then given by

$$X^{(\nu)}(t) = \mu^{(\nu)}(t) + \sum_{k\geq 1} \xi_k^{(\nu)} \phi_{\nu,k}(t), \tag{2}$$

where the principal components score, $\xi_k^{(\nu)} = \int_{\mathcal{T}} (X^{(\nu)}(t) - \mu^{(\nu)}(t))\phi_{\nu,k}(t)dt$, are uncorrelated with $\operatorname{var}(\xi_k^{(\nu)}) = \lambda_{\nu,k}$ for $k = 1, 2, \ldots$ When observing sparse trajectories, this alternative decomposition will also enjoy a simple BLUP formula, as in equation (9) of M-Y, since now

$$\mathrm{E}\left[\xi_{ik}^{(\nu)} | \mathbf{Y}_i\right] = \mathbf{c}_{ik}^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \tag{3}$$

where $\mathbf{c}_{ik} = \operatorname{cov}(\mathbf{Y}_i, \xi_{ik}^{(\nu)})$ also satisfies, with $\mathbf{t}_i = (T_{i_1}, \ldots, T_{iN_i})^T$,

$$\mathbf{c}_{ik} = \int_{\mathcal{T}} \phi_{\nu,k}(s) \frac{\partial^\nu}{\partial s^\nu} G(s, \mathbf{t}_i)\ ds. \tag{4}$$

Considering, as in (14) in M-Y, the dynamics of the underlying stochastic system, we directly get with (1) that

$$\beta_{\nu_1,\nu_2}(s,t) = \frac{\partial^{\nu_2+\nu_1}}{\partial t^{\nu_2} \partial s^{\nu_1}} G(t,s) / \frac{\partial^{2\nu_1}}{\partial t^{\nu_1} \partial s^{\nu_1}} G(s,s), \tag{5}$$

which can also be approximated by truncations as in (24) in M-Y.

It is still not clear to me which approach should be preferred and this probably depends on many ingredients such as the properties of the covariance function of the process $X(t)$, the order of the derivative under consideration as well as how sparse the data are. On the one hand, trying to go directly to the target $X^{(\nu)}(t)$ seems to be a desirable property and considering the ν-th order derivative of the Karhunen-Loève expansion of $X(t)$ may lead to important losses in interpretation and accuracy due to a need for a larger dimension K in order to get a finite dimension functional space that can represent well the variations of the trajectories of $X^{(\nu)}(t)$ around their mean function $\mu^{(\nu)}(t)$ without being able to determine simply what has been lost when truncating. More precisely, for each K, we always have

$$\sum_{k=1}^K \lambda_{\nu,k} = \mathrm{E} \int_{\mathcal{T}} \left(\sum_{k=1}^K \xi_k^{(\nu)} \phi_{\nu,k}(t) \right)^2 dt \geq \mathrm{E} \int_{\mathcal{T}} \left(\sum_{k=1}^K \xi_k \phi_k^{(\nu)}(t) \right)^2 dt. \tag{6}$$

Nevertheless, a precise comparison of the two different approaches does not seem to be trivial. The important drawback of the direct approach given in (2) is that it necessitates estimators of higher order derivatives of the covariance function, $\frac{\partial^{2\nu}}{\partial t^\nu \partial s^\nu} G(t,s)$, and these quantities will never be estimated as accurately as $\frac{\partial^\nu}{\partial t^\nu} G(t,s)$. Its main interest, which is described in (6), is that it may need an expansion with much less terms for the same explained variance. This may be important since it is well known (see *e.g.* Dauxois *et al.* 1982) that the accuracy of the estimators of the eigenfunctions is rapidly decreasing as k increases and strongly depends on the inverse of the gap between adjacent eigenvalues.

Another point that I would address is the importance in this work of the regularity assumption of the trajectories which is hard to check visually with sparse observations. Indeed, there can be a trap in truncated decomposition (10) in M-Y and one has to be careful. The example of the brownian motion is of particular interest. All order derivatives of the eigenfunctions exist (see *e.g.* Ash and Gardner, 1975) even if the derivatives of the trajectories themselves do not exist. In such a situation equation (10), based on a finite rank expansion of the trajectories can lead to misleading interpretation and modeling. Note that when the curves are observed with fine grids some techniques are available (see *e.g.* Blanke and Vial, 2008) to determine what is the amount of regularity but this approach can not be extended in a straightforward way to sparse observations.

To finish, I would like to suggest two directions for future investigations. On the one hand, it can also be of interest to consider the dynamics of the regression function itself. A preliminary work on the estimation by projection onto deterministic basis of the derivative of the regression function in the functional linear model is proposed in Cardot and Johannes (2009). Under some hypotheses, this type of estimator is shown to attain optimal rates of convergence. Extension to functional responses deserves further investigations.

The second issue deals with additional information on the individual trajectories. When dealing with functional data one may also observe for each individual auxiliary real covariates. Taking this information into account not only to describe the conditional mean, as suggested in Section 2, but also the conditional main modes of variability can be of interest in many situations and has still not been addressed much in the literature. A rather straightforward extension of the conditional functional PCA (Cardot, 2007), which relies on a nonparametric estimation of the conditional covariance function, would be to combine the dynamic point of view and the estimation procedures with sparse data proposed in this paper. Such approach would certainly be of interest when exploring real datasets. When one has at hand many real covariates purely nonparametric approaches are less attractive and much work still needs to be done to get tractable estimators of conditional covariance functions, the starting point being probably to find parsimonious modeling of the conditional covariance.

References

Blanke, D. and Vial, C. (2008). Assessing the number of mean square derivatives of a Gaussian process. *Stochastic Process. Appl.* **118**, 1852-1869.

Cardot, H. (2007). Conditional functional principal components analysis. *Scand. J. Statist.* **34**, 317-335.

Cardot, H. and Johannes, J. (2009). Thresholding projection estimators in functional linear models. To appear in *J. Multivariate Analysis,* doi:10.1016/j.jmva.2009.03.001.

Loève, M. (1963). *Probability theory.* Third edition D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London.