Discussion*(as Proposer) of Functional Models for Time-varying Random Objects

Shahin Tavakoli
University of Warwick, Coventry, UK
s.tavakoli@warwick.ac.uk

October 30, 2019

1 Overview

I congratulate Dr Dubey and Professor Müller for their inspiring and thought-provoking paper. The paper considers time-varying random objects, which fall into the remit of functional data analysis (FDA). FDA is interested in the analysis of data points that are complex, such as curves, images, shapes, trees, movies, spectra, sounds or covariance matrices/operators (e.g. Lu et al. 2014). The space Ω (following the paper's notation) in which such data points lie falls (roughly) into 4 categories:

1. Ω is a (separable) Hilbert space: distances between points, moving along specific directions, and inner-products are defined globally.
2. Ω is a connected Riemannian Manifold: distances between points are defined globally, but moving along specific directions and inner-products are only defined locally,
3. Ω is a Banach space: distances between points and moving along a specific direction are defined globally, but there is no inner-product,
4. Ω is a Metric space: only distances between points are defined.

The backbone of FDA is functional principal component analysis (fPCA Hsing & Eubank 2015), which is analogous to PCA: if $X \in H$ is a random element of a real Hilbert space H with inner product $\langle \cdot, \cdot \rangle$, norm $\|\cdot\|$, $\mathbb{E}\|X\|^2 < \infty$, and assuming $\mathbb{E}\, X = 0$ for simplicity, fPCA solves the following iterative problem. For $k = 1, 2, \ldots$, solve

$$v_k \in \arg \max_{v \in H : \|v\|=1} \operatorname{Var}(\langle v, X \rangle) \tag{1}$$

$$\text{such that} \quad \operatorname{cov}(\langle v_k, X \rangle, \langle v_j, X \rangle) = 0, \quad j = 1, \ldots, k-1, \tag{2}$$

where the covariance constraint is omitted for $k = 1$. The v_ks are in the *same space as* X ($v_k \in H$) and called principal component (PC) loading hereafter. The PC scores $\xi_k = \langle v_k, X \rangle$, sometimes just called PCs or just scores, are *real valued*. Defining $u \otimes v$ by $(u \otimes v)f = \langle f, v \rangle\, u$, for $u, v, f \in H$, letting $C = \mathbb{E}[X \otimes X]$ be the covariance operator of X, and writing its spectral decomposition as $C = \sum_{k \geq 1} \lambda_k \varphi_k \otimes \varphi_k$, where the φ_ks are orthonormal eigenvectors of C, with associated eigenvalues λ_k, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, a standard result (Hsing & Eubank 2015) is that the PC loadings satisfy $v_k = \varphi_k, \forall k$. Furthermore,

$$X = \sum_{k \geq 1} \xi_k \varphi_k, \tag{3}$$

*published in 'Functional models for time-varying random objects', Dubey, P. and Müller, H.-G., *Journal of the Royal Statistical Society (Series B)* 2020, 82(2), pp. 309-312.

with convergence in expected squared norm. Keeping only the first K terms in the sum in (3) yields $\sum_{k=1}^{K} \xi_k \varphi_k$, which is the solution to the problem of finding subspace of dimension K onto which the orthogonal projection X of has maximal expected squared norm, or the best approximation of X in expected squared norm. An important special case is $H = L^2([0,1], \mathbb{R})$, the Hilbert space of square integrable functions on $[0,1]$, and $X(\cdot) \in L^2([0,1], \mathbb{R})$ is mean-square continuous (*continuity is key* here; this assumption seems to have been omitted in the paper), then (3) and the spectral decomposition of C hold with stronger conditions (e.g. $t \mapsto \varphi_k(t)$s are continuous), and they are called the Karhunen–Loève expansion of X and Mercer's Theorem, respectively. fPCA is popular because the PC scores can be plotted (e.g. pairs plots) for exploratory data analysis, and the PC loadings provide interpretation of PC scores (or of the modes of variations of X), and typically $X \approx \sum_{k=1}^{K} \xi_k \varphi_k$ is very good for $K = 2, 3$.

In all of the data categories 1. to 4. mentioned above, we can compute distances between any two points. If "moving along a specific direction" and inner-products are (locally) defined, then one can perform (local) fPCA. However in the absence of an inner-product, fPCA is not an option, as it hinges on linear projections. Even if we are given φ_ks, (3) is meaningless if moving along a direction is not defined. Hence for data points in a general metric space, it is not clear how to get PC scores, nor what "modes of variations" means.

The major contribution of the paper is to propose *an* answer to these problems if data points are time-varying random objects, i.e. $X(t) \in \Omega$ for $t \in [0,1]$, where (Ω, d) is a bounded metric space. To construct an alternative to fPCA, which hinges on the definition of covariance (not available in this setup), the key observation of the paper is that, for random variables $U, V \in \mathbb{R}$, (and U', V' i.i.d. copies),

$$\begin{aligned}\operatorname{cov}(U,V) &= \mathbb{E}\left[(U - \mathbb{E}\,U)(V - \mathbb{E}\,V)\right] \\ &= \frac{1}{4}\,\mathbb{E}\left[d_E^2(U,V') + d_E^2(U',V) - 2d_E^2(U,V)\right],\end{aligned}$$

where $d_E(U,V) = |U - V|$. Based on this, if U, V random object in metric space (Ω, d), the paper introduces the *metric covariance* $\operatorname{cov}_\Omega(U,V)$ between U and V, defined as

$$\operatorname{cov}_\Omega(U,V) := \frac{1}{4}\,\mathbb{E}\left[d^2(U,V') + d^2(U',V) - 2d^2(U,V)\right],$$

and defines the *metric autocovariance kernel* $c(s,t) := \operatorname{cov}_\Omega(X(s), X(t)) \in \mathbb{R}$, which summarizes the co-variations of $X(\cdot)$. This kernel is symmetric but not non-negative definite (unlike "standard" covariance kernels), however if the metric space (Ω, d) is of negative type (see definition in the paper), then there is an abstract Hilbert space $(H, \langle . \rangle)$ and an injective mapping $h : \Omega \to H$ such that

$$d(U,V) = \langle h(U) - h(V), h(U) - h(V) \rangle\,,$$

hence $\operatorname{cov}_\Omega(U,V) = \mathbb{E}\left[\langle h(U) - \mathbb{E}\,h(U), h(V) - \mathbb{E}\,h(V)\rangle\right]$, which implies that $c(s,t)$ is non-negative definite. Mercer's Theorem is now in force *provided* $c(s,t)$ is continuous, and the eigenfunction $\phi_k(\cdot) \in L^2([0,1], \mathbb{R})$ of the integral operator induced by $c(\cdot,\cdot)$ are continuous, and they can be interpreted as the modes of variation of X. A decomposition such as (3) is still not possible (and note that ϕ_k in *not* a function $[0,1] \to \Omega$), but the paper proposes two versions of PC scores:

Object FPCs: summarize $X(\cdot)$ by $\psi_\oplus^k \in \Omega, k \in \{1, \ldots, K\}$ ("scores" in the metric space Ω),

Fréchet scores: summarize $X(\cdot)$ by $\beta_k \in \mathbb{R}, k \in \{1, \ldots, K\}$ ("scores" are real-valued).

Typically one hopes that $K = 2, 3$ provides reasonable summaries of X. These hinge on computing Fréchet integrals and Fréchet means, which are minimizers of some functionals (described in the paper). This requires that Ω is a *complete* metric space (an assumption that seems to have been omitted in the paper). These new "scores" allow summarizing each time-varying object by a few objects FPCs (in Ω), or by a few Fréchet scores (in $\mathbb{R}$), which helps exploratory analysis. The ϕ_ks, Object FPCs and Fréchet scores are, in my opinion, the major contribution of the paper. Sample versions of them are proposed, and are backed-up with consistency results.

2 Some Critical Thoughts

Going back to page 2 of the paper, it is written:

> *We aim here at identifying dominant directions of variation [...] where the random objects are indexed by time and in a general metric space.*

So the goal was to (a) deal with general metric spaces, and (b) get dominant modes of variation (which hinges on measuring co-variations and maximum variations). However, all the examples considered in the paper (Ω is a space of densities, networks, or covariances) have a structure that is *much richer* than a metric space. If only a metric is available, $\mathrm{cov}_\Omega(s,t)$ and its eigenfunctions can be computed, however Object FPCs or Fréchet scores will *not* be (easily) computable because they require *solving minimization problems in the metric space* Ω.

Regarding (b), notice crucially that for random elements $U, V \in H$ with finite expected square norm,

$$\mathrm{cov}_\Omega(U,V) = \mathbb{E}\left[\langle U - \mathbb{E}\,U, V - \mathbb{E}\,V\rangle\right] = \mathrm{Tr}\,\mathrm{cov}(U,V) \neq \mathrm{cov}(U,V),$$

where $\mathrm{cov}(U,V)$ is the cross-covariance operator between U and V (a trace-class operator on H). While the trace is a good measure of magnitude *for self-adjoint non-negative definite* operators, it is *not* a good one for general operators. Here is a toy example to illustrate this: let $U(t) \in \mathbb{R}, t \in [0,1/2]$, be random with mean zero and $\mathrm{cov}(U(t),U(t)) = 1$, and define

$$X(t) = \begin{cases} \begin{pmatrix} U(t) \\ 0 \end{pmatrix}, & t \in [0,1/2], \\[2em] \begin{pmatrix} 0 \\ U(t-1/2) \end{pmatrix}, & t \in (1/2,1]. \end{cases}$$

We have $\mathrm{cov}_\Omega(X(t),X(t+1/2)) = 0$ for $t \in [0,1/2]$, but the true covariance matrix is

$$\mathrm{cov}(X(t),X(t+1/2)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and is non-zero. Hence metric covariance fails to measure the perfect linear association between $X(t)$ and $X(t+1/2)$.

Let us now compare the usual PC loadings in the case where $\Omega = H$, a separable Hilbert space. Then $X : [0,1] \to H$, assuming $\int_0^1 \mathbb{E}\,\|X(t)\|^2 dt < \infty$ and glossing over measurability issues and other technicalities, the first PC loading of fPCA is the H-valued function $\varphi : [0,1] \to H$ with $\int_0^1 \|\varphi(t)\|^2 dt = 1$ that maximizes

$$\mathrm{Var}(\langle \varphi, X\rangle) = \int_0^1 \left\langle \underbrace{\mathrm{cov}(X(s),X(t))}_{\text{operator on } H} \varphi(t), \varphi(s) \right\rangle ds dt,$$

where we emphasize that H can be *any* separable Hilbert space, and $\mathrm{cov}(X(s),X(t))$ is an *operator* on H for each fixed $s,t \in [0,1]$.

The first eigenfunction of the metric autocovariance is the function $\phi : [0,1] \to \mathbb{R}$ with $\int_0^1 \phi^2(t)dt = 1$ that maximizes

$$\int c(s,t)\phi(s)\phi(t) ds dt = \int \phi(s)\phi(t)\,\mathrm{Tr}[\mathrm{cov}(X(s),X(t))] ds dt$$

So unless $H = \mathbb{R}$, the variance-maximization interpretation of fPCA is lost with Object FPCA.

Let us now look at the Fréchet scores for the case $\Omega = \mathbb{R}$. We directly get

$$\beta_k := \int d(X(t), \mu_\oplus(t))\phi_k(t) dt = \int |X(t) - \mu(t)|\phi_k(t) dt,$$

hence Fréchet scores are not directly comparable to usual PC scores if $X \in L^2([0,1], \mathbb{R})$.

Having said that, the methods proposed in the paper are universally applicable off-the-shelf, without the need to think about tailoring fPCA to each specific Ω.

3 Possible Extensions

Let us propose a two suggestions for extensions in addition to those mentioned in the paper:

1. $t \in [0,1]$ could be generalized to $t \in E$. For instance, we could have $E =$ Great-Britain, $X(t)$ is a covariance matrix at location $t \in E$, and E is equipped with a non-Euclidean metric, as in Tavakoli et al. (2019).

2. The implicit injection $h : \Omega \to H$ into an abstract Hilbert space H to work with full covariance operator $\mathcal{C} := \mathbb{E}\left[(h(U) - \mathbb{E}\, h(U)) \otimes (h(V) - \mathbb{E}\, h(V))\right]$ on H, which is related to the metric covariance by

$$\mathrm{cov}_\Omega(U, V) = \mathrm{Tr}\, \mathcal{C}.$$

   It is however unclear clear how to work with $\mathcal{C}$ since h is only defined implicitly.

To conclude, I believe that this is a great paper: it proposes novel and universally applicable approaches for exploratory data analysis for cases where fPCA is not directly applicable. It is an inspiring paper, and I am glad to propose the vote of thanks.

References

Hsing, T. & Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley.

Lu, X., Marron, J. S. & Haaland, P. (2014), 'Object-oriented data analysis of cell images', *Journal of the American Statistical Association (Applications & Case Studies)* **109**(506), 548–559.
URL: *https://doi.org/10.1080/01621459.2014.884503*

Tavakoli, S., Pigoli, D., Aston, J. A. & Coleman, J. (2019), 'A spatial modeling approach for linguistic object data: Analysing dialect sound variations across great britain', *Journal of the American Statistical Association (Theory & Methods)* **114**(527), 1081–1096.
URL: *https://doi.org/10.1080/01621459.2019.1607357*