Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

SUPPLEMENTARY MATERIAL FOR "SMOOTH PRINCIPAL COMPONENT ANALYSIS OVER TWO-DIMENSIONAL MANIFOLDS WITH AN APPLICATION TO NEUROIMAGING"

By Eardi Lila[*†], John A. D. Aston[*] and Laura M. Sangalli[†]

[*] *University of Cambridge and* [†] *Politecnico di Milano*

APPENDIX A: SURFACE FINITE ELEMENT DISCRETIZATION

A.1. Well-posedness of the estimation problem (3.8).

Proof. Proposition 1. We exploit a characterization theorem [Braess (2007), chapter 2] which states that if G is a symmetric, positive definite, bilinear form on a vector space L, and F is a linear functional on L, then v is the unique minimizer of

$$G(v,v) - 2F(v)$$

in V if and only if

$$G(v,\varphi) = F(\varphi) \qquad \text{for all } \varphi \in L. \tag{A.1}$$

Moreover, there is at most one solution to problem A.1.

The desired result follows from application of the above theorem considering the vector space $L = H^2(\mathcal{M})$, the symmetric, positive definite, bilinear form $G(f,\varphi) := \sum_{j=1}^p \varphi(p_j)f(p_j) + \lambda \int_{\mathcal{M}} \Delta\varphi\Delta f$ and the linear functional $F(f) = \sum_{j=1}^p f(p_j) \sum_{i=1}^n x_i(p_j)u_i$. Positive definitiveness of the form G, in $H^2(\mathcal{M})$, is shown by the following argument. Suppose that $G(f,f) = 0$ for some $f \in H^2(\mathcal{M})$; then $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f = 0$ and $\sum_{j=1}^p f(p_j)^2 = 0$. Each element $f \in H^2(\mathcal{M})$ can be written such that, for any $p \in \mathcal{M}$, $f(p) = \tilde{f}(p) + c$, with $\tilde{f} \in U = \{\tilde{f} \in H^2(\mathcal{M}) : \int_{\mathcal{M}} \tilde{f} = 0\}$ and c a constant. The solution of $\Delta_{\mathcal{M}}\tilde{f} = 0$ in U exists unique and is $\tilde{f} = 0$ [Dziuk and Elliott (2013)]. Thus $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f = 0$ for $f \in H^2(\mathcal{M})$ implies that $f(p) = c$, for any $p \in \mathcal{M}$, then $\sum_{j=1}^p f(p_j)^2 = pc^2$. But $pc^2 = 0$ if and only if $c = 0$, so $f(\cdot) = 0$. Consequently, G is positive definite on $H^2(\mathcal{M})$.

The estimator $\hat{f}$ is thus

$$\sum_{j=1}^p \varphi(p_j)\hat{f}(p_j) + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}\varphi\Delta_{\mathcal{M}}\hat{f} = \sum_{j=1}^p \varphi(p_j) \sum_{i=1}^n x_i(p_j)u_i \tag{A.2}$$

for every $\varphi \in H^2(\mathcal{M})$.

□

A.2. Reformulation of the estimation problem. The problem of finding $f \in H^2(\mathcal{M})$ that satisfies condition (A.2) for every $\varphi \in H^2(\mathcal{M})$ can be rewritten as the problem of finding $(\hat{f}, g) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$ that satisfies:

$$\begin{cases} \sum_{j=1}^{p} \varphi(p_j)\hat{f}(p_j) + \lambda \int_{\mathcal{M}} (\Delta\varphi) g = \sum_{j=1}^{p} \varphi(p_j) \sum_{i=1}^{n} x_i(p_j) u_i \\ \int_{\mathcal{M}} vg - \int_{\mathcal{M}} v(\Delta \hat{f}) = 0 \end{cases} \tag{A.3}$$

for all $(\varphi, v) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$. In fact, if the pair of functions $(\hat{f}, g) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$ satisfies condition (A.3) for all $(\varphi, v) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$, then $\hat{f}$ also satisfies problem (A.2). In contrast, if $\hat{f} \in H^2(\mathcal{M})$ satisfies problem (A.2), then the pair $(\hat{f}, \Delta\hat{f})$ automatically satisfies the two equations in problem (A.3). Owing to integration by part and to the fact that $\mathcal{M}$ has no boundaries, we get:

$$\int_{\mathcal{M}} (\Delta_{\mathcal{M}}\varphi) g = -\int_{\mathcal{M}} \nabla_{\mathcal{M}}\varphi \nabla_{\mathcal{M}} g$$
$$\int_{\mathcal{M}} v(\Delta_{\mathcal{M}}\hat{f}) = -\int_{\mathcal{M}} \nabla_{\mathcal{M}} v \nabla_{\mathcal{M}} \hat{f}$$

Now, asking the auxiliary function g and of the test functions v to be such that $g, v \in H^1(\mathcal{M})$, the problem of finding $\hat{f} \in H^2(\mathcal{M})$ that satisfies (A.2) for each $\varphi \in H^2(\mathcal{M})$ can be reformulated as finding $(\hat{f}, g) \in (H^1(\mathcal{M}) \cap C^0(\mathcal{M})) \times H^1(\mathcal{M})$

$$\begin{cases} \sum_{j=1}^{p} \varphi(p_j)\hat{f}(p_j) + \lambda \int_{\mathcal{M}} \nabla\varphi \nabla g = \sum_{j=1}^{p} \varphi(p_j) \sum_{i=1}^{n} x_i(p_j) u_i \\ \int_{\mathcal{M}} vg - \int_{\mathcal{M}} \nabla v \nabla \hat{f} = 0 \end{cases} \tag{A.4}$$

for all $(\varphi, v) \in (H^1(\mathcal{M}) \cap C^0(\mathcal{M})) \times H^1(\mathcal{M})$; Moreover, the theory of problems of elliptic regularity ensure that such $\hat{f}$ still belongs to $H^2(\mathcal{M})$ [Dziuk and Elliott (2013) and reference therein]. Finally the discrete estimators $\hat{f}_h, \hat{g}_h \in V \subset H^1(\mathcal{M})$ are obtained solving

$$\begin{cases} \int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \hat{f}_h \nabla_{\mathcal{M}_\mathcal{T}} \varphi_h - \int_{\mathcal{M}_\mathcal{T}} \hat{g}_h \varphi_h = 0 \\ \lambda \int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \hat{g}_h \nabla_{\mathcal{M}_\mathcal{T}} v_h + \sum_{j=1}^{s} \hat{f}_h(p_j) v_h(p_j) = \sum_{j=1}^{s} v_h(p_j) \sum_{i=1}^{n} x_i(p_j) u_i \end{cases}$$

for all $\varphi_h, v_h \in V$. A generic function in V can be written as the linear combination of the finite number of basis spanning V. This allows the solution $\hat{f}_h(p) = \boldsymbol{\psi}(p)^T \hat{\mathbf{f}}$ to be characterized by the linear system (3.12) in the original paper.

APPENDIX B: SIMULATION ON THE SPHERE

Here we present some further simulation studies on a domain $\mathcal{M}$ that is a sphere centered on the origin and with radius $r = 1$, approximated by the triangulated surface $\mathcal{M}_{\mathcal{T}}$ in Figure 1.

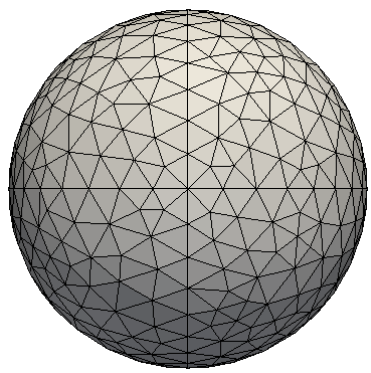

FIG 1. *The triangulated surface approximating the sphere with 488 points.*

B.1. Noisy obervations. We generate $n = 50$ smooth functions $x_1, \ldots, x_{50}$ on $\mathcal{M}_{\mathcal{T}}$ by

$$x_i = u_{i1} v_1 + u_{i2} v_2, \qquad i = 1, \ldots, n$$

where v_1 and v_2 represent the two PC functions with expressions

$$\begin{cases} v_1(x,y,z) = \frac{1}{2}\sqrt{\frac{15}{\pi}}\frac{xy}{r^2} \\ v_2(x,y,z) = \frac{3}{4}\sqrt{\frac{35}{\pi}}\frac{xy(x^2-y^2)}{r^4} \end{cases}$$

and u_{i1}, u_{i2} represent the PC scores, generated independently and distributed as $u_{i1} \sim N(0, \sigma_1^2)$, $u_{i2} \sim N(0, \sigma_2^2)$ with $\sigma_1 = 4$, $\sigma_2 = 2$. The PC functions are two components of the Spherical Harmonics basis set, so they are orthonormal on the sphere, i.e. $\int_{\mathcal{M}} v_i^2 = 1$ for $i \in \{1, 2\}$ and $\int_{\mathcal{M}} v_i v_k = 0$ for $i \neq k$ with $i, k \in \{1, 2\}$. The PC functions are plotted in Figure 2. The functions x_i are sampled at locations coinciding with the nodes of the mesh in Figure 1. At these locations, a Gaussian white noise with standard deviation $\sigma = 0.1$ has been added to the true function x_i. We are then interested in recovering the smooth PC functions v_1 and v_2 from these noisy observations.

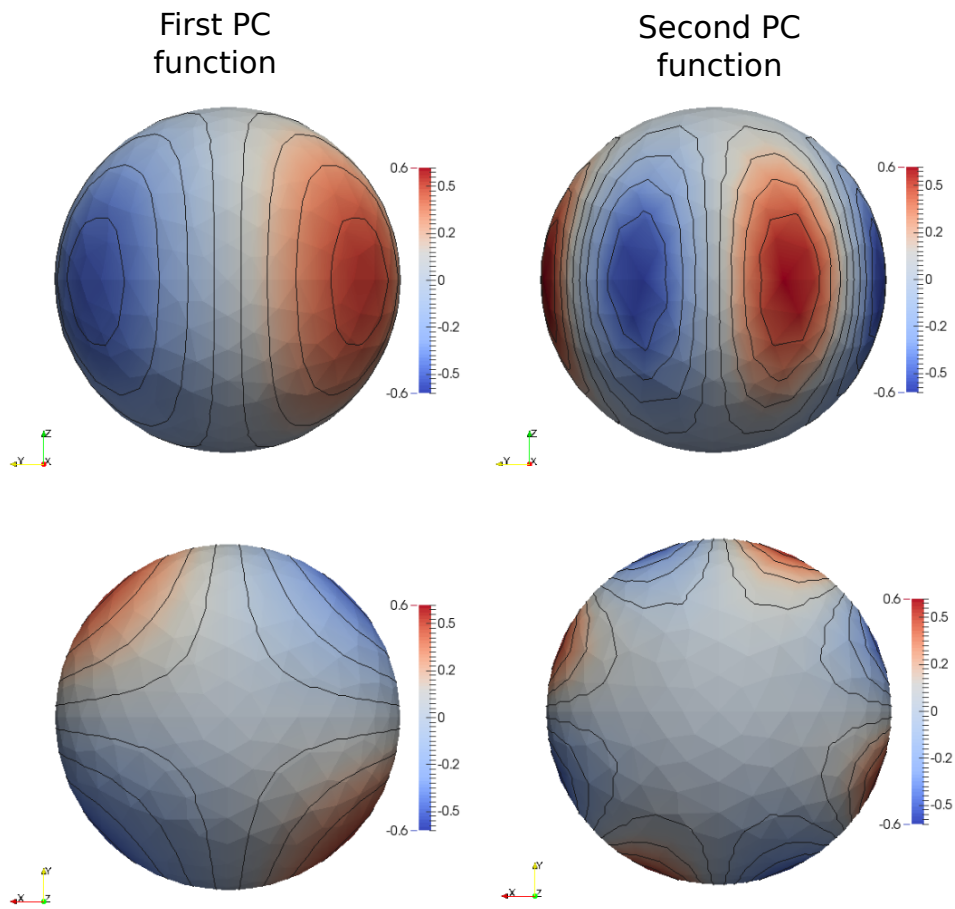

FIG 2. *From the left to the right, two views of the true first and second PC functions.*

We apply the proposed SM-FPCA method, choosing the optimal smoothing parameter λ, both with the K-fold and with GCV. We compare to the approach based on pre-smoothing followed by MV-PCA on the denoised evaluations of the functions at the locations p_j, $j = 1, \ldots, p$. In this case, the smoothing techniques used is Spherical Splines [Wahba (1981)], using the implementation in the R package *mgcv*. The smoothing parameter choice is based on the GCV criterion. We will refer to this approach as SSpline-PCA. The results are summarized in Figure 3.

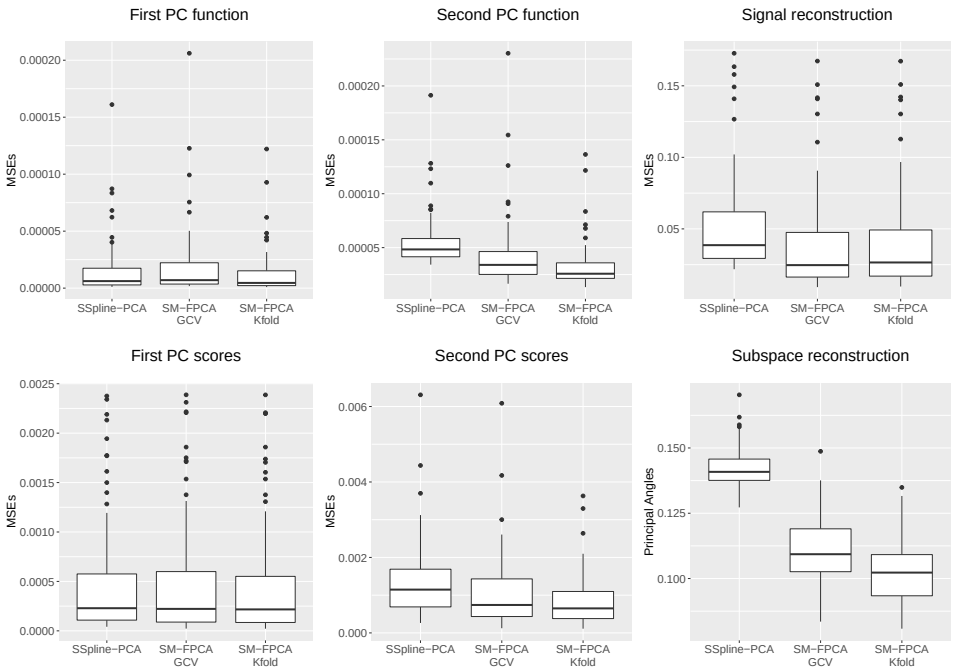

FIG 3. *Boxplots summarizing the performance of SSpline-PCA and SM-FPCA. For the SM-FPCA both GCV and K-fold has been applied for the selection of the smoothing parameter.*

The best estimates of the first two PC functions and corresponding scores are provided by the proposed SM-FPCA with selection of the smoothing parameter based on the K-fold approach. SSpline-PCA does a comparable job on the first principal component, but a significantly worst on the second. A possible explanation for this is the fact that SSpline-PCA tends to over-smooth the data, due to the low signal-to-noise setting of the simulations. This results in good performances for the first PC, but causes a loss of information that worsen the estimation of the second PC. Also the MSE on the signal reconstructions, as well as the measure based on the principal angle between the space spanned respectively by $\{v_i\}_{i=1,2}$ and the estimated PC functions $\{\hat{v}_i\}_{i=1,2}$, emphasize the good performance of the introduced algorithm.

B.2. Spatial mismatching. In this section we complement the set of simulations in the noisy setting by designing a simple simulation that shows how SM-FPCA behaves when a spatial mismatching effect is introduced.

In the motivating application to neuroimaging data, spatial mismatching is introduced by the shape registration algorithm. In this simulation, we consider a spherical domain $\mathcal{M}_{\mathcal{T}}$ and reproduce this spatial mismatching effect, that results in misalignment of the signals on this domain, by including a subject specific shift (in spherical coordinates) of the first PC function. In detail, we generate $n = 50$ smooth functions $x_1, \ldots, x_{50}$ on $\mathcal{M}_{\mathcal{T}}$ by

$$x_i = u_{i1} v_{i1}, \qquad i = 1, \ldots, n \tag{B.1}$$

where u_{i1} represent the PC scores, generated independently and distributed as $u_{i1} \sim N(0, \sigma^2)$ with $\sigma = 4$, and the functions v_{i1} represent misaligned realization of the PC function v_1. Specifically, we parametrize v_1 in spherical coordinates (θ, ϕ) and set $v_{i1}(\theta, \phi) = v_1(\theta + \theta_i, \phi + \phi_i)$, with θ_i and ϕ_i generated independently with a discrete uniform distribution on the set $\{0, 0.4\}$. In Figure 4 we show v_{i1} for the four possible realizations of shifting coefficients (θ_i, ϕ_i).

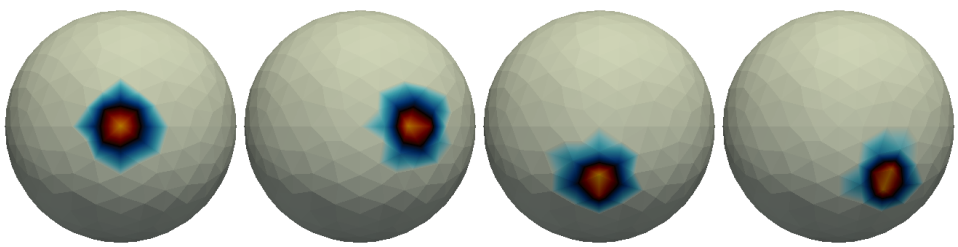

FIG 4. *A plot of the four different realizations of the misaligned PC function v_{i1}.*

The interest is to recover the structure of the only PC function v_1, from the misaligned realizations $\{x_i\}_{i=1,\ldots,n}$, ignoring the effects introduced by the shifts. To consider purely the misalignment's effect, we do not add noise to the sampled functions x_i. In fact, while the benefits of SM-FPCA in the noisy setting have already been extensively demonstrated, we aim now at considering separately the effect of a spatial mismatching on the sampled functions from the effect of the presence of noise. Pre-smoothing of the signal, as performed in SSPline-PCA, is thus unnecessary, and we compare directly MV-PCA to SM-FPCA. In fact, as already mentioned, the proposed SM-FPCA model incorporates the smoothing penalty in a more parsimonious way than the pre-smoothing approach, allowing a direct control of the smoothness of the estimated PC function. We would like to show that SM-FPCA, combined with a cross-validation approach for the choice

of the smoothing parameter λ, might help removing artefacts introduced by the spatial mismatching.

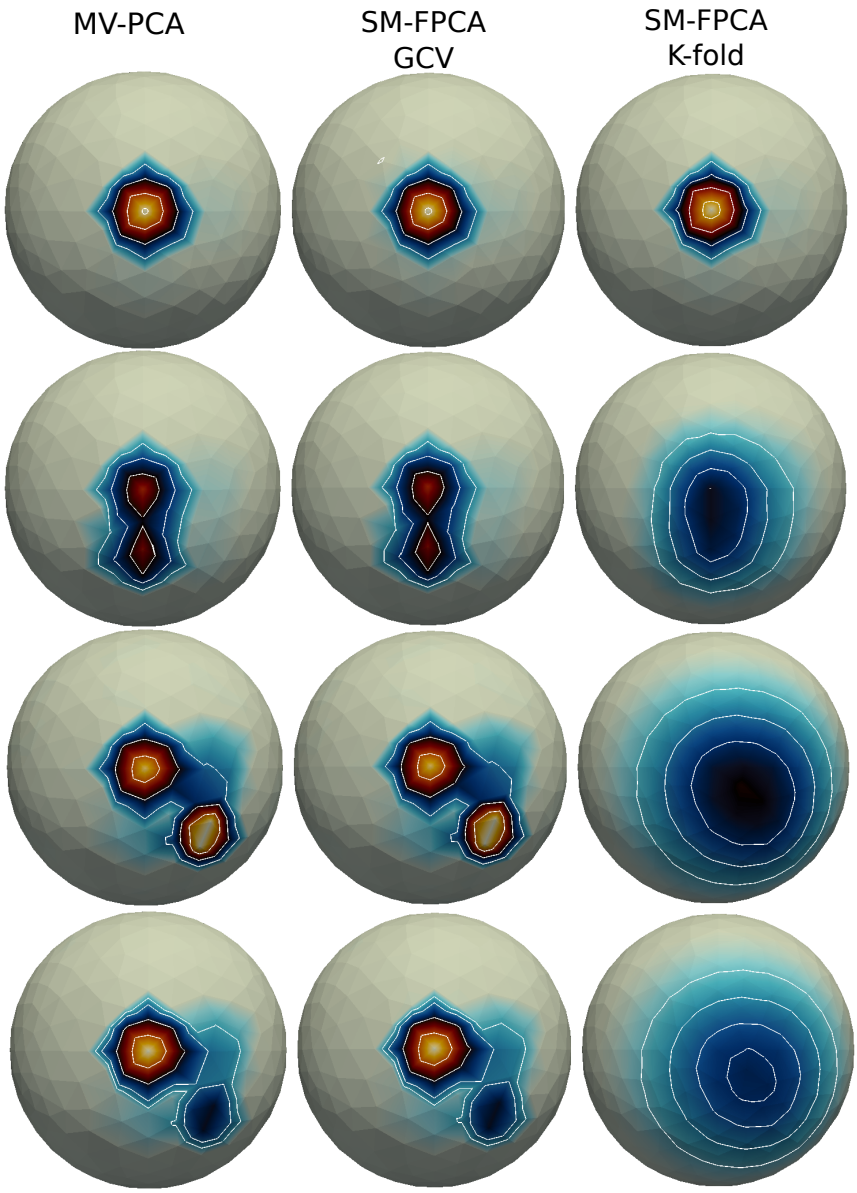

FIG 5. *From top to bottom, plot of the estimates computed on 4 different generated datasets. From left to right, plot of the estimate of the first PC function computed respectively with MV-PCA, SM-FPCA GCV and SM-FPCA K-fold.*

In Figure 5 we show the estimates computed with MV-PCA, SM-FPCA GCV and SM-FPCA K-fold ($K = 5$) for four different datasets generated as in (B.1). In the top row we show a situation where the PC function estimated with MV-PCA shows a satisfactory result. In this case also SM-FPCA GCV and SM-FPCA K-fold show a similar behavior. However, in the bottom three rows the estimates of the PC function computed with MV-PCA and SM-FPCA GCV show some artefacts introduced by the misalignment, while the estimate computed with SM-FPCA K-fold better preserves the shape of the PC function, renouncing however to spatial localization. The results obtained with SM-FPCA K-fold suggest to interpret the phenomena at a more macroscopical scale, due to the high local variability introduced by the spatial mismatching.

The different behavior of SM-FPCA, when the smoothing parameter is chosen by GCV with respect to K-fold cross-validation, can be explained by the fact that this first approach concerns with the choice of λ only in the regression step (3.7), where the choice of λ is only driven by the presence of noise on the vector $\mathbf{X}^T\mathbf{u}$. On the contrary, SM-FPCA K-fold is based on a direct comparison of the PC function estimated on the training and validation sets, obtained partitioning the dataset.

REFERENCES

Braess, D. (2007). *Theory, fast solvers, and applications in elasticity theory.* Cambridge University Press.

Dziuk, G. and Elliott, C. M. (2013). Finite element methods for surface PDEs. *Acta Numerica* **22** 289–396.

Wahba, G. (1981). Spline Interpolation and Smoothing on the Sphere. *SIAM Journal on Scientific and Statistical Computing* **2** 5-16.

Cambridge Centre for Analysis
University of Cambridge
E-mail: e.lila@maths.cam.ac.uk

Statistical Laboratory
University of Cambridge
E-mail: j.aston@statslab.cam.ac.uk

MOX - Dipartimento di Matematica
Politecnico di Milano
E-mail: laura.sangalli@polimi.it