

# Geometric PCA of images

J r mie Bigot<sup>1</sup>, Ra l Gouet<sup>2</sup> & Alfredo Lopez<sup>2</sup>

Institut de Math matiques de Toulouse et CNRS (UMR 5219)<sup>1</sup>

Universit  de Toulouse

Center for Mathematical Modelling<sup>2</sup>

Universidad de Chile

February 2, 2012

## Abstract

We describe a method for analyzing the shape variability of images, called geometric PCA. Our approach is based on the use of deformation operators to model the geometric variability of images around a reference mean pattern. This leads to a new algorithm for estimating shape variability. Some numerical experiments on real images are proposed to highlight the benefits of this approach. The consistency of this procedure is also analyzed in statistical deformable models.

*Keywords:* Geometric variability; Principal component analysis; Mean pattern estimation; Fr chet mean; Image registration; Deformable models.

*AMS classifications:* Primary 62H25; secondary 62H35.

## Acknowledgments

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS, and the financial support from the French embassy in Chile. J.Bigot would like to thank the Center for Mathematical Modeling and the CNRS for financial support and excellent hospitality while visiting Santiago where part of this work was carried out.

## 1 Introduction

In many applications observations are in the form of a set of  $n$  gray-level images  $y_1, \dots, y_n$  (e.g. in geophysics, biomedical imaging or in signal processing for neurosciences), which can be considered as square-integrable functions on a domain  $\Omega$ , a convex subset of  $\mathbb{R}^d$ . Such data are

generally two or three dimensional images. In many situations the observed images share the same structure. This may lead to the assumption that these observations are random elements, which vary around the same mean pattern. Estimating such a mean pattern and characterizing the modes of individual variations around this common shape, is of fundamental interest. Principal components analysis (PCA) is a widely used method to estimate the variations in intensity of images around the usual Euclidean mean  $\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$ . However, such data typically exhibit not only a classical source of photometric variability (a pixel intensity changes from one image to another) but also a (less standard) geometric source of variability in shape, which cannot be recovered by standard PCA. The goal of this paper is to describe a method for analyzing the shape variability of images, called geometric PCA. We present some numerical experiments to highlight the benefits of this approach in applications, and we study the consistency of this procedure in statistical deformable models.

### 1.1 PCA in a Hilbert space

First, let us introduce some tools and notations to perform a standard PCA in a Hilbert space that will be used throughout the paper. Let  $\mathcal{H}$  a separable Hilbert space endowed with inner product  $\langle \cdot, \cdot \rangle$  and associated norm  $\| \cdot \|$ . Let  $Z$  be an  $\mathcal{H}$ -valued random variable. If  $\mathbb{E}\|Z\| < +\infty$ , then  $Z$  has expectation  $\mathbb{E}Z \in \mathcal{H}$  which happens to be the unique element satisfying  $\langle \mathbb{E}Z, h \rangle = \mathbb{E}\langle Z, h \rangle$ , for all  $h \in \mathcal{H}$ . If  $\mathbb{E}\|Z\|^2 < +\infty$ , then the (population) covariance operator  $K : \mathcal{H} \rightarrow \mathcal{H}$  corresponding to  $Z$  is given by

$$Kh = \mathbb{E}\langle Z - \mathbb{E}Z, h \rangle (Z - \mathbb{E}Z) \text{ for } h \in \mathcal{H}.$$

Moreover, the operator  $K$  is self-adjoint, positive semidefinite and trace-class. Hence,  $K$  is compact, with nonnegative (population) eigenvalues  $(\gamma_\lambda)_{\lambda \in \Lambda}$  and orthonormal (population) eigenvectors  $(\phi_\lambda)_{\lambda \in \Lambda}$  and such that

$$Kh = \sum_{\lambda \in \Lambda} \gamma_\lambda \langle h, u_\lambda \rangle u_\lambda$$

where  $\Lambda = \{1, \dots, \dim(\mathcal{H})\}$ , if  $\dim(\mathcal{H}) < \infty$  or  $\Lambda = \mathbb{N}$  otherwise. If we assume that the eigenvalues are arranged in decreasing order  $\gamma_1 \geq \gamma_2 \geq \dots \geq 0$ , the  $\lambda$ -th mode of variation of the random variable  $Z$ , is defined as the function  $\mathbb{E}Z + \rho\sqrt{\gamma_\lambda}u_\lambda$ , for  $\lambda \in \Lambda$ , where  $\rho$  is a real playing the role of a weight parameter (typically one takes  $\rho$  between  $-1$  and  $1$ ). Thus, the PCA of  $Z$  is obtained by diagonalizing the covariance operator  $K$ .

Now, let  $z_1, \dots, z_n \in \mathcal{H}$  and  $\hat{K}_n : \mathcal{H} \rightarrow \mathcal{H}$ , their (empirical) covariance operator defined as

$$\hat{K}_n h = \frac{1}{n} \sum_{i=1}^n \langle z_i - \bar{z}_n, h \rangle (z_i - \bar{z}_n), \text{ for } h \in \mathcal{H},$$

where,  $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i$ . As  $\hat{K}_n$  is also self-adjoint, positive semidefinite and compact, it admits the decomposition

$$\hat{K}_n h = \sum_{\lambda \in \Lambda} \hat{\gamma}_\lambda \langle h, \hat{u}_\lambda \rangle \hat{u}_\lambda,$$

where  $\hat{\gamma}_1 \geq \hat{\gamma}_2 \geq \dots \geq 0$  are the (empirical) eigenvalues, and  $(\hat{u}_\lambda)_{\lambda \in \Lambda}$  is the set of (empirical) orthonormal eigenvectors of  $\hat{K}_n$ . Now, the  $\lambda$ -th (empirical) mode of variation of the data  $z_1, \dots, z_n$  is defined as  $\bar{z}_n + \rho\sqrt{\hat{\gamma}_\lambda}\hat{u}_\lambda$ , for  $\lambda \in \Lambda$ . Thus, the empirical PCA of the data  $z_1, \dots, z_n$  is obtained by diagonalizing the operator  $\hat{K}_n$ .

If  $\mathcal{H}$  is finite-dimensional, diagonalizing  $\hat{K}_n$  corresponds to the empirical PCA for vectors in a finite dimensional Euclidean space. If  $\mathcal{H} = L^2(\Omega) = \{f : \Omega \rightarrow \mathbb{R}, \|f\|_2^2 := \int_\Omega |f(x)|^2 dx < \infty\}$  diagonalizing  $\hat{K}_n$  is usually referred to as the method of functional PCA in nonparametric statistics (see e.g. [RS05] for an introduction to functional data analysis). Various authors (see e.g. [DPR82], [Sil96] and references therein), under standard probabilistic assumptions on the data, have studied the consistency of empirical PCA in Hilbert spaces and proposed sufficient conditions to ensure that the empirical eigenvalues and eigenvectors converge to the population eigenvalues and eigenvectors as  $n \rightarrow +\infty$ .

Empirical PCA, with  $\mathcal{H} = L^2(\Omega)$ , applied to a set of  $n$  images  $y_1, \dots, y_n$ , is a method to compute the principal directions of photometric variability of the  $y_i$ 's around the usual Euclidean mean  $\bar{y}_n$ . However, in many situations, images also exhibit a large variability in shape, see the example of images of handwritten digits displayed in Figure 1 and Figure 2. In such cases, the standard Euclidean mean  $\bar{y}_n$  is not a satisfying estimator of the typical shape of each individual image, see Figure 1 (a) and Figure 2 (a), and standard PCA does not meaningfully reflect the modes of variability of the data, see Figure 1 (b),(c) and Figure 2 (b),(c). In particular, the second empirical modes of variation are no longer a single digit but rather the superposition of two digits in different orientations.

It is well known that the usual Euclidean mean  $\bar{y}_n$  is the minimizer of the sum-of-squares Euclidean distances to each of the data points, namely

$$\bar{y}_n = \arg \min_{f \in L^2(\Omega)} \frac{1}{n} \sum_{m=1}^n \|f - y_m\|_2^2.$$

The idea underlying PCA is that the Hilbert space  $L^2(\Omega)$ , equipped with the standard inner product, is well suited to model natural images. However, the set of such objects (as those in Figures 1-2) typically cannot be considered as a linear sub-space of  $L^2(\Omega)$ . Therefore, the Euclidean distance  $\|f_1 - f_2\|_2$  is generally not well suited, since it is not adapted to the geometry of the set to which the images  $f_1$  and  $f_2$  truly belong. Actually one can see that the images in Figures 1-2 have mainly a geometric variability in space, which is much more important than the photometric variability.

## 1.2 Organization of the paper

In Section 2, we define the geometric PCA of random images. We describe the numerical implementation of this method, and we illustrate its advantages over standard PCA for the analysis of handwritten digits. A review of PCA-like methods to analyze geometric variability of images is also given. In Section 3, we study the consistency of geometric PCA in statistical

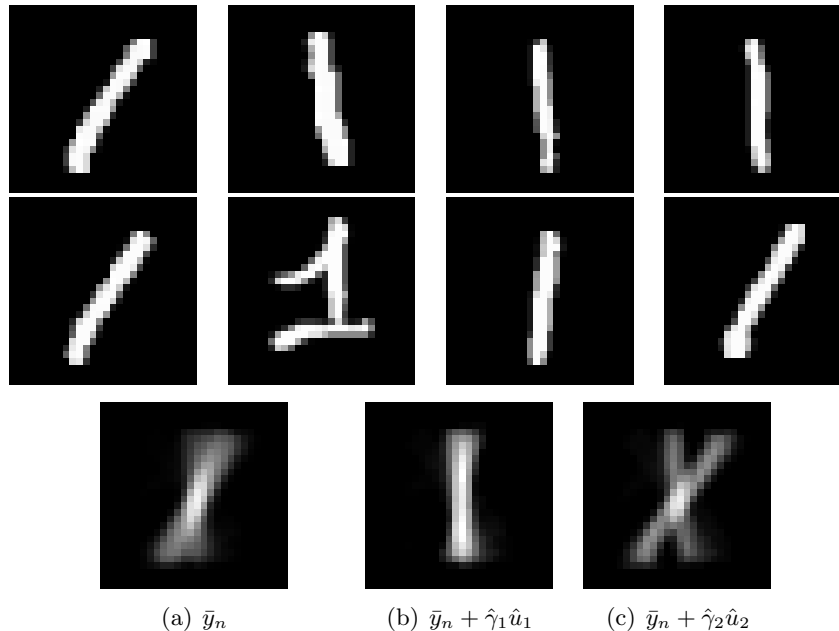


Figure 1: Digit 1. Two first rows: a sample of 8 images out of  $n = 100$  taken from the Mnist data base [LBBH98] - Last row: standard Euclidean mean, first and second empirical modes of variation.

deformable models. We conclude the paper in Section 4 by a short discussion. All proofs are gathered in a technical Appendix.

## 2 Geometric PCA

For convenience, we prefer to present the ideas of geometric PCA under the assumption that the images are observed on a continuous domain  $\Omega$ . In practice, such data are obviously observed on a discrete set of time points or pixels. However, assuming that the data are random elements of  $L^2(\Omega)$  is more convenient for dealing with statistical aspects of an inferential procedure, as it avoids the treatment of the bias introduced by any discretization of the domain  $\Omega$ .

### 2.1 Grenander's pattern theory of deformable templates

Following the ideas of Grenander's pattern theory (see [GM07] for a recent overview), one may consider that the data  $y_1, \dots, y_n$  are obtained through the deformation of the same reference image. In this setting, images are treated as points in  $L^2(\Omega)$  and the geometric variations of the images are modeled by the action of Lie groups on the domain  $\Omega$ . Recently, there has been a growing interest in Lie groups of transformations to model the geometric variability of images (see e.g. [BMTY05, TY05a, TY05b, You10] and references therein), and applications

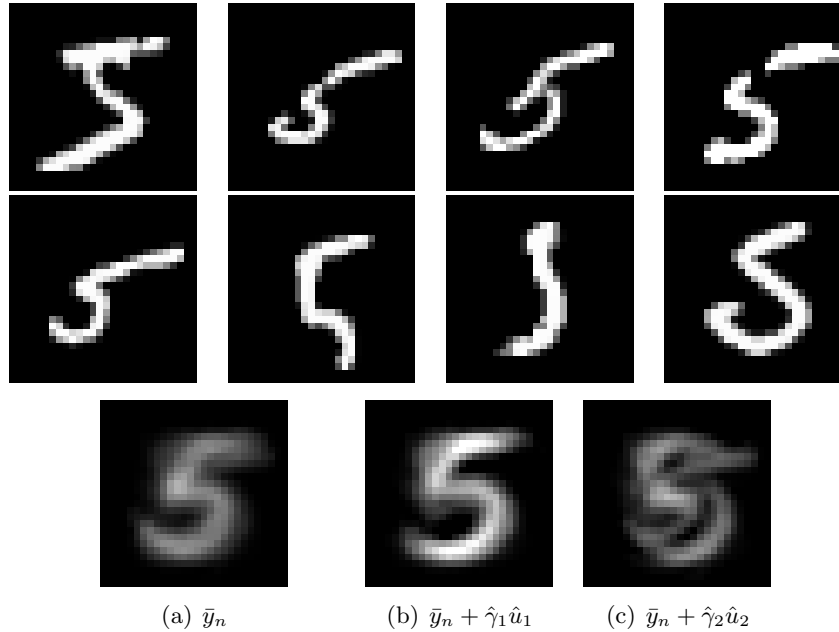


Figure 2: Digit 5. Two first rows: a sample of 8 images out of  $n = 100$  taken from the Mnist data base [LBBH98] - Last row: standard Euclidean mean, first and second empirical modes of variation.

are numerous in particular in biomedical imaging, see e.g. [FLPJ04, JDJG04].

Grenander's pattern theory leads to the construction of non-Euclidean distances between images. In this paper, we propose to model shape variability through the use of deformation operators (acting on  $\Omega$ ) that are parameterized by a separable Hilbert space  $\mathcal{V}$ , with inner product  $\langle \cdot, \cdot \rangle$ . We also assume that  $\Omega$  is equipped with a metric denoted by  $d_\Omega$ .

**Definition 2.1.** *Let  $\mathcal{V}$  be a Hilbert space. A deformation operator parameterized by  $\mathcal{V}$  is a mapping  $\varphi : \mathcal{V} \times \Omega \rightarrow \Omega$  such that, for any  $v \in \mathcal{V}$ , the function  $x \mapsto \varphi(v, x)$  is a homeomorphism on  $\Omega$ . Moreover,  $\varphi(0, \cdot)$  is the identity on  $\Omega$  and, for any  $v \in \mathcal{V}$ , there exists  $v^* \in \mathcal{V}$  such that  $\varphi^{-1}(v, \cdot) = \varphi(v^*, \cdot)$ .*

In this paper, we will study as illustrative examples of deformation operators the cases of translations, rigid deformations and non-rigid deformations generated by stationary vector fields.

*Translations:* Let  $\Omega = [0, 1)^d$ , for some integer  $d \geq 1$  and  $\mathcal{V} = \mathbb{R}^d$ . Let also

$$\varphi(v, x) = (\text{mod}(x_1 + v_1, 1), \dots, \text{mod}(x_d + v_d, 1)), \quad (2.1)$$

with

$$\varphi^{-1}(v, x) = (\text{mod}(x_1 - v_1, 1), \dots, \text{mod}(x_d - v_d, 1)),$$

for all  $v = (v_1, \dots, v_d) \in \mathbb{R}^d$  and  $x = (x_1, \dots, x_d) \in \Omega$ , where  $\text{mod}(a, 1)$  denotes the modulo operation between a real  $a$  and 1. Clearly,  $\varphi(0, \cdot)$  is the identity in  $\Omega$  and  $v^* = -v$ . Moreover, it can be shown (see Section 3.3) that  $\varphi(v, \cdot)$  is an homeomorphism.

*Rigid deformations of 2D images:* Let  $\Omega = \mathbb{R}^2$  and  $\mathcal{V} = \mathbb{R} \times \mathbb{R}^2$ . Let also

$$\varphi(v, x) = R_\alpha x + b,$$

with

$$\varphi^{-1}(v, x) = R_{-\alpha}(x - b),$$

for all  $v = (\alpha, b) \in \mathbb{R} \times \mathbb{R}^2$ , and  $x \in \mathbb{R}^2$  where  $R_\alpha$  is the rotation matrix of angle  $\alpha$  and  $b \in \mathbb{R}^2$  defines a translation. Observe that,  $\varphi((0, 0), \cdot)$  is the identity in  $\Omega$  and  $v^* = (-\alpha, -R_{-\alpha}b)$ . Clearly,  $\varphi(v, \cdot)$  is an homeomorphism.

*Diffeomorphic deformations generated by stationary vector fields:* Let  $\Omega = [0, 1]^d$  for some integer  $d \geq 1$ , and  $\mathcal{V}$  a separable Hilbert space of smooth vector fields such that  $\mathcal{V}$  is continuously embedded the set of functions  $v : \Omega \rightarrow \mathbb{R}^d$  which are continuously differentiable and such  $v$  and its derivatives vanish at the boundary of  $\Omega$ . For  $x \in \Omega$  and  $v \in \mathcal{V}$ , define  $\varphi(v, x)$  as the solution at time  $t = 1$  of the following ordinary differential equation (O.D.E.)

$$\frac{d\phi_t}{dt} = v(\phi_t), \quad (2.2)$$

with initial condition  $\phi_0 = x \in \Omega$ . It is well known (see e.g. [You10]) that, for any  $v \in \mathcal{V}$ , the function  $x \mapsto \varphi(v, x)$  is a  $C^1$  diffeomorphism on  $\Omega$ . The inverse of  $x \mapsto \varphi(v, x)$  is given by  $x \mapsto \varphi(-v, x)$ , and thus  $v^* = -v$ . Hence,  $\varphi(v, \cdot)$  satisfies the conditions in Definition 2.1.

## 2.2 Registration

Registration is a widely used method in image processing that consist in geometric transforms of a set of images  $y_1, \dots, y_n \in L^2(\Omega)$ , so that they be compared. This method can be described as an optimization problem which amounts to minimizing a dissimilarity functional between images.

**Definition 2.2** (Dissimilarity functional). *Let  $\varphi$  be a deformation operator, as described in Definition 2.1,  $\mathbf{v} = (v_1, \dots, v_n) \in \mathbf{V} := \mathcal{V}^n$  and  $\mathbf{y} = (y_1, \dots, y_n)$ , with  $y_i \in L^2(\Omega)$ ,  $i = 1, \dots, n$ .*

(a) *The template dissimilarity functional corresponding to  $\mathbf{v}$ ,  $\mathbf{y}$  and  $f \in L^2(\Omega)$  is defined as*

$$M^t(\mathbf{v}, \mathbf{y}, f) := \frac{1}{n} \sum_{i=1}^n \int_{\Omega} \left( y_i(\varphi(v_i, x)) - f(x) \right)^2 dx. \quad (2.3)$$

(b) *The groupwise dissimilarity functional corresponding to  $\mathbf{v}$  and  $\mathbf{y}$  is defined as*

$$M^g(\mathbf{v}, \mathbf{y}) := \frac{1}{n} \sum_{i=1}^n \int_{\Omega} \left( y_i(\varphi(v_i, x)) - \frac{1}{n} \sum_{j=1}^n y_j(\varphi(v_j, x)) \right)^2 dx. \quad (2.4)$$

Template registration of the images  $y_1, \dots, y_n$  onto some known template  $f \in L^2(\Omega)$ , is defined as the problem of minimizing the criterion given by the dissimilarity functional (2.3), with respect to  $v$  in

$$\mathcal{V}_\mu := \{v = (v_1, \dots, v_n), v_i \in \mathcal{V}_\mu\},$$

where  $\mathcal{V}_\mu := \{v \in \mathcal{V} : \|v\| \leq \mu\}$ , for some regularization parameter  $\mu \geq 0$  whose choice has to be discussed. Note that imposing the constraint  $\|v_i\| \leq \mu$  allows to explicitly control the norm of the vector  $v_i$  which is generally proportional to the distance between the deformation  $\varphi(v_i, \cdot)$  and the identity.

On the other hand, groupwise registration of  $y_1, \dots, y_n$  is defined as the problem of minimizing the functional (2.4) with respect to  $v$  in  $\mathcal{U} \subseteq \mathcal{V}_\mu$ . Two possible choices for  $\mathcal{U}$ , defined in terms of linear constraints on  $v$ , are

$$\mathcal{U}_0 := \left\{v \in \mathcal{V}_\mu, \sum_{m=1}^n v_m = 0\right\} \quad \text{and} \quad \mathcal{U}_1 := \left\{v \in \mathcal{V}_\mu, v_1 = 0\right\}. \quad (2.5)$$

Choosing  $\mathcal{U} = \mathcal{U}_0$  amounts to imposing that the deformation parameters  $(v_1, \dots, v_n)$  used to align the data have an empirical mean equal to zero, while taking  $\mathcal{U} = \mathcal{U}_1$  corresponds to choosing  $y_1$  as a reference template onto which  $y_2, \dots, y_n$  will be aligned.

Geometric PCA applied to a set of images  $\mathbf{y} = (y_1, \dots, y_n)$  is the following two step procedure. In the first step, one applies either a template or a groupwise registration, which leads to the computation of

$$\hat{v} \in \arg \min_{v \in \mathcal{V}_\mu} M^t(v, \mathbf{y}, f) \quad \text{or} \quad \hat{v} \in \arg \min_{v \in \mathcal{U}} M^g(v, \mathbf{y}). \quad (2.6)$$

In the second step, a standard PCA is carried out on  $\hat{v} = (\hat{v}_1, \dots, \hat{v}_n)$ , based on the following covariance operator

$$\hat{K}_n v = \frac{1}{n} \sum_{i=1}^n \langle \hat{v}_i - \bar{v}_n, v \rangle (\hat{v}_i - \bar{v}_n), \quad \text{for } v \in \mathcal{V}, \quad (2.7)$$

with  $\bar{v}_n = \frac{1}{n} \sum_{i=1}^n \hat{v}_i$ . This operator admits the decomposition

$$\hat{K}_n v = \sum_{\lambda \in \Lambda} \hat{\kappa}_\lambda \langle v, \hat{\phi}_\lambda \rangle \hat{\phi}_\lambda, \quad (2.8)$$

where  $\hat{\kappa}_1 \geq \hat{\kappa}_2 \geq \dots \geq 0$  and  $(\hat{\phi}_\lambda)_{\lambda \in \Lambda}$  are the eigenvalues and orthonormal eigenvectors of  $\hat{K}_n$ . We now state the definition of geometric PCA of a set of images.

**Definition 2.3** (Geometric PCA). *Let  $\varphi$  be a deformation operator parametrized by  $\mathcal{V}$ , as described in Definition 2.1. Let  $(\hat{\kappa}_\lambda, \hat{\phi}_\lambda)_{\lambda \in \Lambda}$  be the eigenvalues and orthonormal eigenvectors of the operator  $\hat{K}_n$  in (2.7). For  $\lambda \in \Lambda$ , the  $\lambda$ -th empirical mode of geometric variation of the data  $y_1, \dots, y_n$  is the homeomorphism  $\hat{\psi}_\lambda : \Omega \rightarrow \Omega$  defined by*

$$\hat{\psi}_\lambda(x) = \varphi^{-1}(\bar{v}_n + \sqrt{\hat{\kappa}_\lambda} \hat{\phi}_\lambda, x), \quad x \in \Omega. \quad (2.9)$$

We also denote  $\hat{\psi}_{\lambda, \rho}(x) = \varphi^{-1}(\bar{v}_n + \rho \sqrt{\hat{\kappa}_\lambda} \hat{\phi}_\lambda, x)$ , where  $\rho \in \mathbb{R}$  is a weighting value.

After the registration step, we obtain a set of deformed images  $y_1 \circ \varphi(\hat{v}_1, \cdot), \dots, y_n \circ \varphi(\hat{v}_n, \cdot)$ , each of them aligned either with respect to  $f$  in the case of template registration, or with respect to  $\hat{f} := \frac{1}{n} \sum_{j=1}^n y_j(\varphi(\hat{v}_j, x))$ , in the case of groupwise registration. Hence, in the case of template registration,  $f \circ \hat{\psi}_\lambda$  can be used to visualize the  $\lambda$ -th mode of geometric variation of the data. Similarly, in the case of groupwise registration, one uses  $\hat{f} \circ \hat{\psi}_\lambda$ . Note that  $\hat{f}$  can be interpreted as a mean pattern image. Moreover, the computation of  $\hat{f}$  is closely related to the notion of Fréchet mean of images, recently studied in [BC11], from a statistical point of view.

## 2.3 Numerical implementation and application of geometric PCA to handwritten digits data

In this section we explain in detail the implementation of geometric PCA, in the case of groupwise registration, using the class of diffeomorphic deformation described in Section 2.1. The method is applied to a set of  $n = 30$  images, defined on the domain  $\Omega = [0, 1]^2$ , taken from the Mnist data base of handwritten digits [LBBH98].

### 2.3.1 Specification of the Hilbert space of parameter $\mathcal{V}$

We choose  $\mathcal{V}$  as the vector space of functions from  $\Omega$  to  $\mathbb{R}^2$ , generated by a B-Splines basis of functions, because they have good properties for approximating continuous functions and implementing efficient computations [UAE93a, UAE93b]. Let  $\{b_k : \Omega \rightarrow \mathbb{R}, k = 1, \dots, p\}$  denote a set of bi-dimensional tensor product B-Splines, with knots defined on a regular grid of  $\Omega$ , and  $p$  some integer whose choice has to be discussed. We define  $\mathcal{V}$  as the space of vector fields of the form  $v = \sum_{k=1}^p \tilde{v}_k b_k$ , where  $\tilde{v}_k = (\tilde{v}_k^{(1)}, \tilde{v}_k^{(2)}) \in \mathbb{R}^2, k = 1, \dots, p$ . We denote by  $v^{(1)}, v^{(2)} : \Omega \rightarrow \mathbb{R}$  the coordinates of  $v \in \mathcal{V}$ , i.e.,  $v(x) = (v^{(1)}(x), v^{(2)}(x))$  for  $x \in \Omega$ . Note that the dimension of  $\mathcal{V}$  is  $2p$  and that a basis is given by

$$\{(b_1, 0), \dots, (b_p, 0), (0, b_1), \dots, (0, b_p)\}. \quad (2.10)$$

We endow  $\mathcal{V}$  with the inner product

$$\langle u, v \rangle := \langle u^{(1)}, v^{(1)} \rangle_L + \langle u^{(2)}, v^{(2)} \rangle_L, \quad u, v \in \mathcal{V},$$

where  $\langle u^{(1)}, v^{(1)} \rangle_L := \int_{\Omega} Lu^{(1)}(x)Lv^{(1)}(x)dx$ ,  $\langle u^{(2)}, v^{(2)} \rangle_L := \int_{\Omega} Lu^{(2)}(x)Lv^{(2)}(x)dx$  and  $L$  is a differential operator. As suggested in [BMTY05] we take  $L = \gamma I + \alpha \Delta$ , where  $I$  is the identity operator,  $\Delta$  is the laplacian operator and  $\gamma, \alpha$  are positive scalars. By using the basic properties of differentiation and integration of B-Splines [UAE93a], we derive an explicit formula for computing the inner product in  $\mathcal{V}$ , that can be implemented using convolution filters. By an adequate design of the B-Spline grid, we ensure that the values of  $v$  and its derivatives are zero at the boundary of  $\Omega$ .

### 2.3.2 Minimization of the dissimilarity functional $D^g$

In the case of groupwise registration, we minimize the dissimilarity functional (2.4) over the set  $\mathcal{U}_0$  defined in (2.2). Thanks to the above choice for  $\mathcal{V}$ , the minimization of the criterion

(2.4) has to be performed over a subset of  $\mathbb{R}^{2p}$ . In order to take into account the constraint  $\|v_i\| \leq \mu$ ,  $1 \leq i \leq n$ , we use a logarithmic barrier approach to obtain an approximate solution. Then, for the minimization, we use a gradient descent algorithm, with an adaptive step. Such an algorithm requires the computation of the deformation operator  $\varphi : \mathcal{V} \times \Omega \rightarrow \Omega$  and its gradient, with respect to the coefficients  $\tilde{v}_k = (\tilde{v}_k^{(1)}, \tilde{v}_k^{(2)})$  that parameterize the vector field  $v$ . In the case of diffeomorphic deformation operators,  $\varphi(v, x)$  corresponds to the solution at time  $t = 1$  of the ODE (2.2). We solve the O.D.E. using a forward Euler integration scheme. For a comparison of different methods for solving such O.D.E., we refer to [BZO08]. It can be shown (see Lemma 2.1 in [BMTY05]), that the gradient of  $\varphi$  with respect to the  $\tilde{v}_k = (\tilde{v}_k^{(1)}, \tilde{v}_k^{(2)})$ 's has a closed-form expression.

### 2.3.3 Spectral decomposition of the empirical covariance operator

Let  $\hat{v}_1, \dots, \hat{v}_n$  be the vector fields in  $\mathcal{V}$  obtained after the registration step described above. Recall that the empirical covariance operator of the  $\hat{v}_i$ 's is defined as  $\hat{K}_n v = \frac{1}{n} \sum_{i=1}^n \langle \hat{v}_i - \bar{v}_n, v \rangle (\hat{v}_i - \bar{v}_n)$ ,  $v \in \mathcal{V}$ . In what follows, we describe how to perform the spectral decomposition of  $\hat{K}_n$ .

Let  $\tilde{v}_i = (\tilde{v}_i^{(1)}, \tilde{v}_i^{(2)})$  with  $\tilde{v}_i^{(1)} = (\tilde{v}_{i,1}^{(1)}, \dots, \tilde{v}_{i,p}^{(1)})$  and  $\tilde{v}_i^{(2)} = (\tilde{v}_{i,1}^{(2)}, \dots, \tilde{v}_{i,p}^{(2)})$  being the coefficients of  $\hat{v}_i$  with respect to the base (2.10), i.e.  $\hat{v}_i = \sum_{k=1}^p (\tilde{v}_{i,k}^{(1)}, \tilde{v}_{i,k}^{(2)}) b_k$ . We identify the Hilbert space  $\mathcal{V}$  with  $\mathbb{R}^{2p}$  endowed by the inner product

$$\langle \tilde{u}, \tilde{v} \rangle := \tilde{u} \Sigma \tilde{v}^t, \quad \tilde{u}, \tilde{v} \in \mathbb{R}^{2p}, \quad (2.11)$$

where  $\Sigma$  is a  $2p \times 2p$  matrix with entries  $\Sigma_{j,k} = \Sigma_{j+p,k+p} := \langle b_j, b_k \rangle_L$  for  $j, k = 1, \dots, p$  and  $\Sigma_{j,k} := 0$  in the other cases. Hence, the operator  $\hat{K}_n$  can be identified with a  $2p \times 2p$  matrix  $\tilde{K}_n$ , given by

$$\tilde{K}_n := \frac{1}{n} \tilde{v} \tilde{v}^t \Sigma,$$

where  $\tilde{v}$  is the  $2p \times n$  matrix with  $i$ -th column equals to  $\tilde{v}_i^t - \frac{1}{n} \sum_{j=1}^n \tilde{v}_j^t$ . The matrix  $\Sigma$  is symmetric, hence admits a diagonalization  $\Sigma = P \Lambda P$  with  $\Lambda$  diagonal matrix and  $P^t P = P P^t = I$ . The idea now is to reduce the problem to a standard diagonalization of the symmetric matrix  $M := \frac{1}{n} \Lambda^{\frac{1}{2}} P^t \tilde{v} \tilde{v}^t P \Lambda^{\frac{1}{2}}$ , namely we find the decomposition  $M = W D W^t$  with  $D$  diagonal matrix and  $W^t W = W W^t = I$ . We obtain the following spectral decomposition of  $\tilde{K}_n$  respect to the inner product (2.11)

$$\tilde{K}_n = U D U^t \Sigma,$$

where  $U := P^t W$ . Remark that the columns of  $U$  are orthonormal vectors in  $\mathbb{R}^{2p}$  with respect to the inner product (2.11). Indeed, it holds that  $U^t \Sigma U = I$ . Finally, we define  $\hat{\kappa}_\lambda$  as the  $\lambda$ -th elements of the diagonal matrix  $D$  and we let  $\hat{\phi}_\lambda := \sum_{k=1}^p (U_{k,\lambda}, U_{k+p,\lambda}) b_k$ , for  $\lambda = \{1, \dots, 2p\}$ . It can be checked that  $(\hat{\phi}_\lambda)_{\lambda=1}^{2p}$  are orthonormal vectors of  $\mathcal{V}$ , and we thus obtain that  $\hat{K}_n v = \sum_{\lambda=1}^{2p} \hat{\kappa}_\lambda \langle v, \hat{\phi}_\lambda \rangle_L \hat{\phi}_\lambda$ . If we assume that  $\hat{\kappa}_1 \geq \dots \geq \hat{\kappa}_{2p}$ , then

$$\hat{\psi}_{\lambda,\rho} = \varphi^{-1}(\bar{v}_n + \rho \sqrt{\hat{\kappa}_\lambda} \hat{\phi}_\lambda, \cdot)$$

is the  $\lambda$ -th empirical mode of geometric variation according to Definition 2.3.

### 2.3.4 Application of geometric PCA to handwritten digits data and comparison with standard PCA

We now describe the application of geometric PCA to handwritten digits, taken from the Mnist data base [LBBH98], based on the numerical framework we had described so far. Also, we illustrate the benefits of geometric PCA over standard PCA.

For the B-Spline base of  $\mathcal{V}$ , we choose a B-Spline degree equals to 3, as it provides a good trade-off among smoothness and the size of the support. The number of B-Spline knots is  $p = 81$  arranged in a  $9 \times 9$  regular grid. Such value of  $p$  provides a fine B-Spline grid with respect to an image size of  $28 \times 28$ . For defining the differential operator  $L$ , we take  $\gamma = 100$  and  $\alpha = 1$ . Note that  $\gamma \gg \alpha$  in order to compensate for scaling factor associated to the inter knot spacing. Finally, we choose a regularization parameter  $\mu = 10$ . This value was determined experimentally, by trying to obtain a good compromise among regularity of the vector fields and matching of the images, at the registration step.

For each available digit (from 0 to 9), we took  $n = 30$  images and we carried out a geometric PCA on each of these image sets. In this database, for each digit, one observes a large source of geometrical variability that can be modeled by diffeomorphic deformations. We proceed by groupwise registration, as there are no reference images available. To illustrate the advantages of our procedure, we have also carried out a standard PCA of each digit, which amounts to analyzing the photometric variability of the data. Thus we compute

$$\bar{y}_n + \rho \sqrt{\hat{\gamma}_\lambda} \hat{u}_\lambda,$$

the  $\lambda$ -th standard empirical mode of photometric variation of the data as described in Section 1.1. Figures 3, 4 and 5 show the geometric modes of variations by displaying the images

$$\hat{f} \circ \hat{\psi}_{\lambda, \rho} \text{ where } \hat{f}(x) = \frac{1}{n} \sum_{j=1}^n y_j(\varphi(\hat{v}_j, x)),$$

with  $\lambda = 1, 2$  and  $\rho = 2, -2$ . Results using the standard PCA are also displayed in the same Figures. We observe that geometric PCA better reflects the main modes of variability of the digits. In contrast, standard PCA fails in several cases in representing the geometric variability of some digits, and it results sometimes in a blurring of the images. Also, it can be seen that  $\hat{f}$  is a much better mean pattern of the data than the Euclidean mean  $\bar{y}_n$ .

## 2.4 PCA-like methods for analyzing geometric variability

The idea of applying a PCA analysis to the resulting transformation parameters after registration of a set of images is at the core of several methods to estimate the geometrical variability of images.

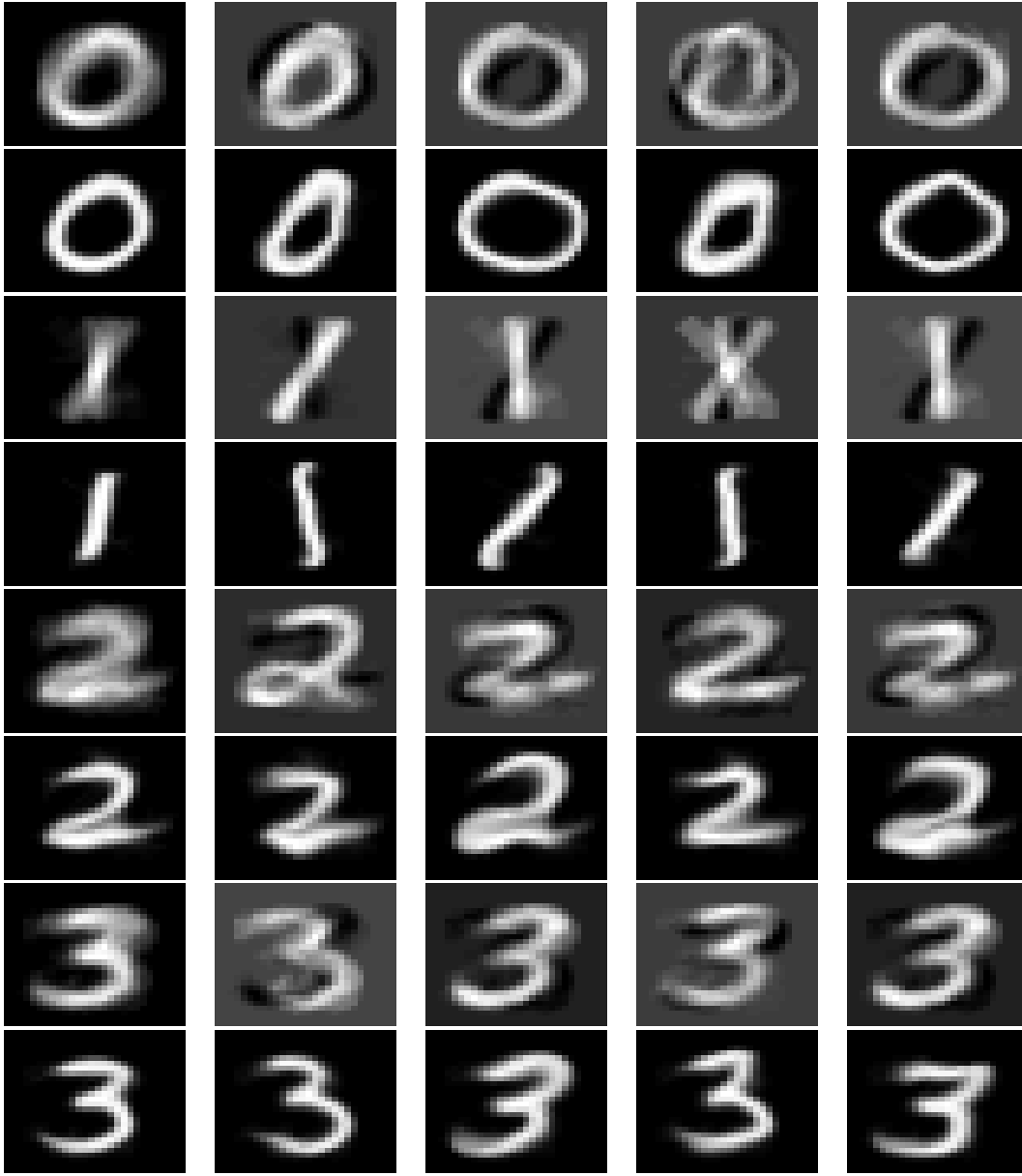


Figure 3: Visualization of the geometric PCA and standard PCA results for digits 1 to 3. For each digit, the superior row from left to right corresponds to  $\bar{y}$ ,  $\bar{y}_n - 2\sqrt{\hat{\gamma}_1}\hat{u}_1$ ,  $\bar{y}_n + 2\sqrt{\hat{\gamma}_1}\hat{u}_1$ ,  $\bar{y}_n - 2\sqrt{\hat{\gamma}_2}\hat{u}_2$  and  $\bar{y}_n + 2\sqrt{\hat{\gamma}_2}\hat{u}_2$  respectively. The inferior row from left to right corresponds to  $\hat{f}$ ,  $\hat{f} \circ \hat{\psi}_{1,-2}$ ,  $\hat{f} \circ \hat{\psi}_{1,2}$ ,  $\hat{f} \circ \hat{\psi}_{2,-2}$  and  $\hat{f} \circ \hat{\psi}_{2,2}$ . We observe, in several cases, that standard PCA results (first row of each digit) do not recover well shapes of the digits and produce a blurring of the images. In contrast, geometric PCA results (second row of each digit) recover the geometric features of the digits.



Figure 4: Visualization of the geometric PCA and standard PCA results for digits 4 to 6. For each digit, the superior row from left to right corresponds to  $\bar{y}$ ,  $\bar{y}_n - 2\sqrt{\hat{\gamma}_1}\hat{u}_1$ ,  $\bar{y}_n + 2\sqrt{\hat{\gamma}_1}\hat{u}_1$ ,  $\bar{y}_n - 2\sqrt{\hat{\gamma}_2}\hat{u}_2$  and  $\bar{y}_n + 2\sqrt{\hat{\gamma}_2}\hat{u}_2$  respectively. The inferior row from left to right corresponds to  $\hat{f}$ ,  $\hat{f} \circ \hat{\psi}_{1,-2}$ ,  $\hat{f} \circ \hat{\psi}_{1,2}$ ,  $\hat{f} \circ \hat{\psi}_{2,-2}$  and  $\hat{f} \circ \hat{\psi}_{2,2}$ . We observe, in several cases, that standard PCA results (first row of each digit) do not recover well shapes of the digits and produce a blurring of the images. In contrast, geometric PCA results (second row of each digit) recover the geometric features of the digits.

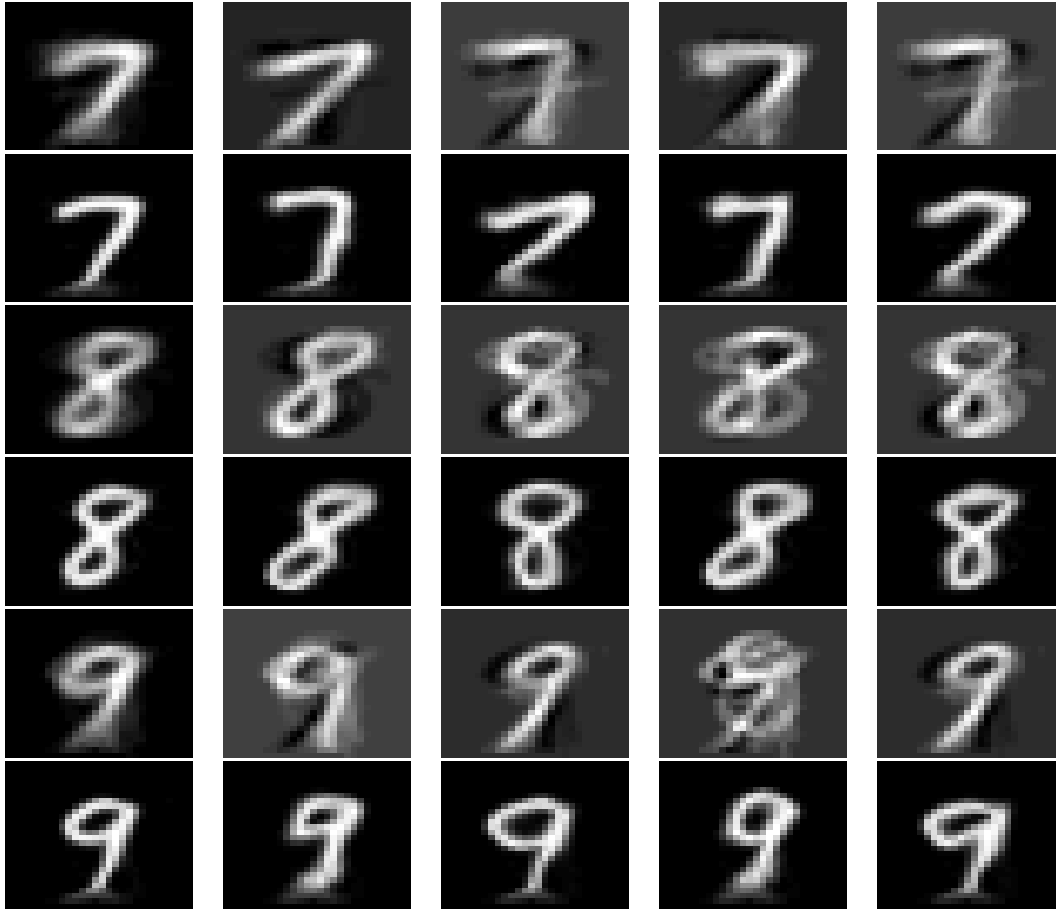


Figure 5: Visualization of the geometric PCA and standard PCA results for digits 7 to 9. For each digit, the superior row from left to right corresponds to  $\bar{y}$ ,  $\bar{y}_n - 2\sqrt{\hat{\gamma}_1}\hat{u}_1$ ,  $\bar{y}_n + 2\sqrt{\hat{\gamma}_1}\hat{u}_1$ ,  $\bar{y}_n - 2\sqrt{\hat{\gamma}_2}\hat{u}_2$  and  $\bar{y}_n + 2\sqrt{\hat{\gamma}_2}\hat{u}_2$  respectively. The inferior row from left to right corresponds to  $\hat{f}$ ,  $\hat{f} \circ \hat{\psi}_{1,-2}$ ,  $\hat{f} \circ \hat{\psi}_{1,2}$ ,  $\hat{f} \circ \hat{\psi}_{2,-2}$  and  $\hat{f} \circ \hat{\psi}_{2,2}$ . We observe, in several cases, that standard PCA results (first row of each digit) do not recover well shapes of the digits and produce a blurring of the images. In contrast, geometric PCA results (second row of each digit) recover the geometric features of the digits.

In [RFS03], a linear and finite dimensional space of non rigid transformations is considered as the admissible space of deformations on which PCA is carried out. This is the so-called *statistical deformation model* that succeed Coote’s active shape models. An important limitation of this approach is the lack of invertibility of the deformations. In several cases, the invertibility is a desirable property from the point of view of physical (for instance when analyzing shape variability of a determined organ) and mathematical modelling. Within the context of linear space of deformations, the non-invertibility issue had been addressed by enforcing the positiveness of the Jacobian determinant [HM06, Sdi08]. However, such methods are unsuited for further statistical analysis, as statistical procedures on resulting transformation parameters (such the empirical Euclidean mean), does not lead to invertible transforms. Moreover, the inverse transforms do not belong to the initial space of transformations.

More recently, diffeomorphisms have been used to model deformations between images in the context of Grenander’s pattern theory [BMTY05, HBO09, Ash07]. In this framework, the set of admissible diffeomorphic deformations is considered as a Riemmaniann manifold, and thus first and second order statistics analysis on manifolds [FLPJ04, Pen06] can be applied to perform statistical analysis of diffeomorphic deformations [WBR<sup>+</sup>07, ACPA06, BHO07, HBO07]. Such diffeomorphisms are constructed as solution of a differential equation governed by a *time dependent vector field* [BMTY05]. The optimal diffeomorphism after the registration of two images can be fully characterized by the initial point in time (or equivalently by the *initial momentum*) of the associated time dependent vector field. This important *momentum conservation* property allows to carry out PCA on the Hilbert space of initial momentums [WBR<sup>+</sup>07].

A particular sub-class of diffeomorphic deformations is of the set of diffeomorphisms generated by an O.D.E. governed by stationary vector fields. In this way, diffeomorphisms are directly characterized by vector fields in a Hilbert space and statistical analysis like PCA can be carried on these vector fields. Compared to the case of diffeomorphisms generated by nonstationary vector fields, the computational cost of the registration is considerably smaller, while keeping comparable accuracy according to experimental results reported in [HBO09, Ash07]. Moreover, diffeomorphisms generated by stationary vector fields have some desirable properties that allow simple and fast computations of *exponential* and *logarithms*, that is, computing the diffeomorphism associated to a vector field and vice versa. Hence, PCA methods for manifolds can be implemented to analyze shape variability of diffeomorphic deformations generated by stationary vector fields, see [ACPA06, BHO07, HBO07].

### 3 Consistency of geometric PCA in statistical deformable models

In the last decades, in the framework of Grenander’s pattern theory, there has been a growing interest in the use of first order statistics for the computation of mean pattern from a set of images [AAT07, AKT10, BC11, BGV09, BGL09] and in the construction of consistent procedures. However, there is not so much work in the statistical literature on the consistency of second order statistics for the analysis of geometric variability of images. In particular, the convergence

of such procedures in simple statistical models has generally not been established.

We study the consistency of geometric PCA, in the context of the following statistical deformable model

$$Y_i(x) = f^*(\varphi^{-1}(V_i, x)) + \epsilon W_i(x), \quad x \in \Omega, \quad i = 1, \dots, n, \quad (3.1)$$

where

- $f^*$  is an unknown mean pattern belonging to  $L^2(\Omega)$ ,
- $\varphi$  is a deformation operator associated to a Hilbert space  $\mathcal{V}$ , (in the sense of Definition 2.1), equipped with inner product  $\langle \cdot, \cdot \rangle$  and induced norm  $\| \cdot \|$ ,
- $V_1, \dots, V_n$  are independent copies of  $V$ , a zero-mean, square-integrable  $\mathcal{V}$ -valued random variable (i.e.  $\mathbb{E}V = 0$  and  $\mathbb{E}\|V\|^2 < \infty$ ),
- There exists  $\mu > 0$  (regularization parameter) such that  $\mathbb{P}(V \in \mathcal{V}_\mu) = 1$ ,
- $\epsilon > 0$  is a noise level parameter,
- $W_1, \dots, W_n$  are independent copies of a zero mean Gaussian process  $W \in L^2(\Omega)$ , such that  $\mathbb{E}\|W\|_2^2 = 1$ ,
- $(V_1, \dots, V_n)$  and  $(W_1, \dots, W_n)$  are mutually independent.

Additionally, we assume that the eigenvalues  $\kappa_1 \geq \kappa_2 \geq \dots \geq 0$  of the population covariance operator  $K$ , defined as

$$Kv = \mathbb{E}\langle V, v \rangle V, \quad v \in \mathcal{V}, \quad (3.2)$$

have algebraic multiplicity 1, i.e.  $\kappa_1 > \kappa_2 > \dots \geq 0$ . This implies that the  $\lambda$ -th eigen-gap, defined as  $\delta_\lambda := \min_{\lambda' \in \Lambda \setminus \{\lambda\}} |\kappa_\lambda - \kappa_{\lambda'}|$ , is strictly positive, for any  $\lambda \in \Lambda$ .

Observe that the function  $f^*$  in (3.1) models the common shape of the  $Y_i$ 's. The  $W_i$ 's represent the individual variations in intensity of the data around the mean pattern  $f^*$ , and thus correspond to a classical source of variability that could be analyzed by standard PCA. To the contrary, the random elements  $\varphi^{-1}(V_i, \cdot)$  model deformations of the domain  $\Omega$ , and thus correspond to a source of shape variability in the data.

**Definition 3.1** (Population geometric modes of variations). *Let  $K$  be the population covariance operator, defined in (3.2), with (population) eigenvalues  $\kappa_1 > \kappa_2 > \dots \geq 0$  and (population) orthonormal eigenvectors  $\phi_1, \phi_2, \dots$ . For  $\lambda \in \Lambda$ , the  $\lambda$ -th population mode of geometric variation of the random variable  $V$  is the homeomorphism  $\psi_\lambda : \Omega \rightarrow \Omega$  defined by*

$$\psi_\lambda(x) = \varphi^{-1}(\sqrt{\kappa_\lambda} \phi_\lambda, x), \quad x \in \Omega.$$

In this paper we say that geometric PCA is a consistent procedure if, for data  $\mathbf{Y} = (Y_1, \dots, Y_n)$  following model (3.1), and for all  $\lambda \in \Lambda$ , the  $\lambda$ -th empirical mode of geometric variation  $\hat{\psi}_\lambda$  (see equation (2.9)) tends to the  $\lambda$ -th population mode of geometric variation  $\psi_\lambda$ , as  $n \rightarrow +\infty$  and  $\epsilon \rightarrow 0$ , in a sense to be made precise later on. In this context, the empirical

modes of geometric variation are obtained from the eigenvalues  $\hat{\kappa}_\lambda$  and the eigenvectors  $\hat{\phi}_\lambda$  of the empirical covariance operator

$$\hat{K}_n v = \frac{1}{n} \sum_{i=1}^n \langle \hat{V}_i - \bar{V}_n, v \rangle (\hat{V}_i - \bar{V}_n), \text{ for } v \in \mathcal{V}, \quad (3.3)$$

where  $(\hat{V}_1, \dots, \hat{V}_n)$  belongs to  $\arg \min_{\mathbf{v} \in \mathcal{V}_\mu} M^t(\mathbf{v}, \mathbf{Y}, f^*)$  or  $\arg \min_{\mathbf{v} \in \mathcal{U}} M^g(\mathbf{v}, \mathbf{Y})$ . Consequently, from now on, empirical eigenvalues, eigenvectors and modes of geometric variations will be considered as random elements.

**Remark 3.1.** *Observe that, for template registration, where  $\hat{\mathbf{v}} \in \arg \min_{\mathbf{v} \in \mathcal{V}_\mu} M^t(\mathbf{v}, \mathbf{y}, f^*)$ , each coordinate  $\hat{v}_i$  depends only on  $v_i$ . This fact implies that  $\hat{V}_1, \dots, \hat{V}_n$  are i.i.d. However, this is not the case for groupwise registration, where  $\hat{v}_i$  may depend on all the  $v_i$ 's.*

The asymptotic  $n \rightarrow +\infty$  is rather natural and correspond to the setting of a growing number of images. On the other hand, the setting  $\epsilon \rightarrow 0$  corresponds to the analysis of the influence of the additive term  $\epsilon W_i$  in model (3.1). The main result of this section is that geometric PCA is consistent only in the double asymptotic setting  $n \rightarrow +\infty$  and  $\epsilon \rightarrow 0$ . This result illustrates the fact that the photometric perturbations  $\epsilon W_i, i = 1, \dots, n$  in model (3.1) have to be sufficiently small in order to recover the geometric modes of variation.

**Definition 3.2.** *A deformation operator  $\varphi$  (see Definition 2.1) is said to be  $\mu$ -regular if there exists  $\mu > 0$  such that*

(i)

$$\int_{\Omega} f^2(\varphi^{-1}(v, x)) dx \leq A_\mu \int_{\Omega} f^2(x) dx, \quad (3.4)$$

for all  $f \in L^2(\Omega)$ ,  $v \in \mathcal{V}_\mu$  and some constant  $A_\mu > 0$ ; and

(ii) *the mapping  $v \rightarrow \varphi(v, \cdot)$  from  $\mathcal{V}_\mu$  to  $C(\Omega, \Omega)$  is continuous,*

where  $C(\Omega, \Omega)$  is the space of continuous function from  $\Omega$  to  $\Omega$ , endowed with the metric  $d_C(\psi, \phi) := \sup_{x \in \Omega} d_\Omega(\psi(x), \phi(x))$ .

Note that if  $\varphi(v, \cdot)$  is sufficiently smooth such that the determinant of its Jacobian matrix is bounded, that is  $|\det(J(\varphi(v, x)))| \leq A_\mu$  for all  $v \in \mathcal{V}_\mu$  and  $x \in \Omega$ , then (3.4) follows from a change of variable.

Finally, before stating our consistency results, we define convergence in probability in the double asymptotic setting  $n \rightarrow \infty, \epsilon \rightarrow 0$ . Let  $X_{n,\epsilon}, X_n, X_\epsilon, X, n = 1, 2, \dots, \epsilon > 0$  random variables with values on a metric space  $(S, d)$ . The notation  $\text{plim } X_{n,\epsilon} = X_n$  stands for  $d(X_{n,\epsilon}, X_n) \rightarrow 0$  in probability as  $\epsilon \rightarrow 0$ ;  $\text{plim } X_{n,\epsilon} = X_\epsilon$  denotes  $d(X_{n,\epsilon}, X_\epsilon) \rightarrow 0$  in probability as  $n \rightarrow \infty$ . Finally,  $\text{plim}_{n,\epsilon} X_{n,\epsilon} = X$  is equivalent to  $\text{plim}_n \text{plim}_\epsilon X_{n,\epsilon} = \text{plim}_\epsilon \text{plim}_n X_{n,\epsilon} = X$ .

In this paper, all equalities and inequalities involving random variables are understood in the almost sure sense. We require the following definition, in our main results: for  $u, v \in \mathcal{V}$ ,  $\sin(u, v) := \sqrt{1 - \langle u/\|u\|, v/\|v\| \rangle^2}$ .

### 3.1 Case of template registration

**Theorem 3.1.** *Let  $\mathbf{Y} = (Y_1, \dots, Y_n)$  i.i.d. observations of model (3.1), with deformation operator  $\varphi$  and regularization parameter  $\mu$ . Let  $\hat{\kappa}_\lambda$  and  $\hat{\phi}_\lambda, \lambda \in \Lambda$ , be the empirical eigenvalues and eigenvectors corresponding to template registration of  $\mathbf{Y}$ , with  $f = f^*$ . Suppose that  $\varphi$  is  $\mu$ -regular and the mapping  $\varphi^* : \mathcal{V}_\mu \rightarrow L^2(\Omega)$ , defined by  $\varphi^*(v) := f^* \circ \varphi^{-1}(v, \cdot)$  for  $v \in \mathcal{V}_\mu$ , is injective and its inverse  $\varphi^{*-1} : \varphi^*(\mathcal{V}_\mu) \rightarrow \mathcal{V}_\mu$  is continuous. Then  $\text{plim}_{n, \epsilon} \hat{\kappa}_\lambda = \kappa_\lambda$  and*

$$\text{plim}_{n, \epsilon} \sin^2(\hat{\phi}_\lambda, \phi_\lambda) = 0, \text{ for all } \lambda \in \Lambda.$$

**Remark 3.2.** *Observe that, under the hypothesis of Theorem 3.1, the  $\lambda$ -th empirical mode of geometric variation  $\hat{\psi}_\lambda$  converges the population mode of geometric variation  $\psi_\lambda$  in probability, when  $n \rightarrow +\infty$  and  $\epsilon \rightarrow 0$ , as elements of  $(C(\Omega, \Omega), d_C)$ . Indeed, this result follows from the continuity of the mapping  $v \rightarrow \varphi(v, \cdot)$ , which is guaranteed by the  $\mu$ -regularity of  $\varphi$ .*

We show below how a stronger regularity assumption on  $\varphi$ , allows one to obtain rates of convergence for  $\hat{\kappa}_\lambda$  and  $\hat{\phi}_\lambda$ , via a concentration inequality that depends explicitly on  $n$  and  $\epsilon$ .

**Theorem 3.2.** *Under the hypothesis of Theorem 3.1 and if  $\varphi^{*-1}$  is uniformly Lipschitz (in the sense that  $\|u - v\|^2 \leq L(f^*, \mu) \|\varphi^*(u) - \varphi^*(v)\|_2^2$ , for every  $u, v \in \mathcal{V}_\mu$  and some constant  $L(f^*, \mu) > 0$  depending only on  $f^*$  and  $\mu$ ), then*

$$\mathbb{P} \left( |\hat{\kappa}_\lambda - \kappa_\lambda|^2 > C(f^*, \mu) \max(h(u, n, \epsilon) + \sqrt{h(u, n, \epsilon)}; g(u, n)) \right) \leq \exp(-u),$$

for any  $u > 0$ , where  $C(f^*, \mu) > 0$  is a constant depending only on  $f^*$  and  $\mu$ ;  $h(u, n, \epsilon) = \epsilon^2 (1 + 2\frac{u}{n} + 2\sqrt{\frac{u}{n}})$  and  $g(u, n) = \left( \frac{u}{n} + \sqrt{\frac{u^2}{n^2} + \frac{u}{n}} \right)^2$ .

Take now  $u^* > 0$  such that

$$C(f^*, \mu) \max(h(u^*, n, \epsilon) + \sqrt{h(u^*, n, \epsilon)}; g(u^*, n)) < (\delta_\lambda/2)^2,$$

then for any  $0 < u \leq u^*$

$$\mathbb{P} \left( \sin^2(\hat{\phi}_\lambda, \phi_\lambda) > (2/\delta_\lambda)^2 C(f^*, \mu) \max(h(u, n, \epsilon) + \sqrt{h(u, n, \epsilon)}; g(u, n)) \right) \leq 2 \exp(-u).$$

### 3.2 Case of groupwise registration

In order to prove consistency for groupwise registration, we require model (3.1) to satisfy the set of identifiability assumptions, shown below. For  $\mathbf{u}, \mathbf{v} \in \mathcal{V} = \mathcal{V}^n$ , let

$$D^g(\mathbf{u}, \mathbf{v}) := M^g(\mathbf{u}, (f_1^*, \dots, f_n^*)) = \frac{1}{n} \sum_{i=1}^n \int_{\Omega} \left( f_i^*(\varphi(u_i, x)) - \frac{1}{n} \sum_{j=1}^n f_j^*(\varphi(u_j, x)) \right)^2 dx, \quad (3.5)$$

where  $f_i^*(x) := f^*(\varphi^{-1}(v_i, x))$ ,  $x \in \Omega$ ,  $i = 1, \dots, n$ . Observe that,  $D^g(\mathbf{u}, \mathbf{V}) = M^g(\mathbf{u}, \mathbf{Y})$  when  $\mathbf{Y} = (Y_1, \dots, Y_n)$  follows model (3.1) with  $\epsilon = 0$ .

**Definition 3.3** (*g-identifiability*). *Model (3.1) is said to be g-identifiable if*

- (i) *there exists a measurable function  $\mathbf{u}^* : \mathcal{V}_\mu \rightarrow \mathcal{U}$  such that for every  $\eta > 0$  there exists a constant  $C > 0$  not depending on  $n$ , with  $D^g(\mathbf{u}, \mathbf{v}) - D^g(\mathbf{u}^*, \mathbf{v}) > C$ , for every  $\mathbf{u} \in \mathcal{U}$  satisfying  $\bar{d}^2(\mathbf{u}^*, \mathbf{u}) > \eta$ , and*
- (ii)  $\text{plim}_n \bar{d}^2(\mathbf{u}^*(\mathbf{V}), \mathbf{V}) = 0$ ,

where  $\bar{d}^2(\mathbf{u}, \mathbf{v}) := \frac{1}{n} \sum_{i=1}^n \|u_i - v_i\|^2$ , for  $\mathbf{u}, \mathbf{v} \in \mathcal{V}$ .

Observe that condition (i) above implies that, for every  $\mathbf{v} \in \mathcal{V}_\mu$ ,  $D^g(\mathbf{u}, \mathbf{v})$  has a unique measurable minimizer  $\mathbf{u}^*(\mathbf{v})$  on  $\mathcal{U}$ .

**Theorem 3.3.** *Let  $\mathbf{Y} = (Y_1, \dots, Y_n)$  i.i.d. observations of model (3.1), with deformation operator  $\varphi$  and regularization parameter  $\mu$ . Let  $\hat{\kappa}_\lambda$  and  $\hat{\phi}_\lambda$ ,  $\lambda \in \Lambda$ , be the empirical eigenvalues and eigenvectors corresponding to groupwise registration of  $\mathbf{Y}$ . Suppose that  $\varphi$  is  $\mu$ -regular and that (3.1) is g-identifiable. Then  $\text{plim}_{n,\epsilon} \hat{\kappa}_\lambda = \kappa_\lambda$  and  $\text{plim}_{n,\epsilon} \sin^2(\hat{\phi}_\lambda, \phi_\lambda) = 0$ , for all  $\lambda \in \Lambda$ .*

**Remark 3.3.** *Observe that, as in the case of template registration, it can be shown that under the hypothesis of Theorem 3.3,  $\hat{\psi}_\lambda$  converges to  $\psi_\lambda$  in probability, as  $n \rightarrow +\infty$  and  $\epsilon \rightarrow 0$ .*

### 3.3 Translation operators

We study the applicability of Theorems 3.1, 3.2 and 3.3 to translation operator  $\varphi$  given by (2.1). In this case,  $\Omega = [0, 1]^d$ , for some integer  $d \geq 1$ , is equipped with the distance  $d_\Omega(x, y) := \sum_{k=1}^d \min\{|x_k - y_k|, 1 - |x_k - y_k|\}$ , for  $x = (x_1, \dots, x_d) \in \Omega$ . Let also  $\mathcal{V} = \mathbb{R}^d$ , be equipped with the usual Euclidean inner product.

Now, let us show that  $\varphi$  is a deformation operator in the sense of Definition 2.1. It holds that  $\varphi(0, \cdot)$  is the identity in  $\Omega$  and  $\varphi^{-1}(v, \cdot) = \varphi(-v, \cdot)$ . Last, from (i) in Lemma C.1, it follows that the mapping  $\varphi(v, \cdot)$  is continuous for all  $v \in \mathcal{V}$ . Moreover, we prove that  $\varphi$  is  $\mu$ -regular, for all  $\mu > 0$ : for (i) in Definition 3.2, take  $f \in L^2(\Omega)$  and consider its periodic extension  $f_{per}$  to  $\mathbb{R}^d$ . Then (3.4) is a consequence of  $f(\varphi(v, x)) = f_{per}(x + v)$  which holds for all  $v \in \mathbb{R}^d, x \in \Omega$ . Finally, condition (ii) in Definition 3.2, follows from (ii) in Lemma C.1.

We impose further conditions on model (3.1) implying that  $\varphi^{*-1}$ , defined in Theorem 3.1 is Lipschitz. Let  $\theta_k$ ,  $k = 1, \dots, d$ , be the low frequency Fourier coefficients of the template  $f^*$ , that is

$$\theta_k := \int_{\Omega} f^*(x) e^{-i2\pi x_k} dx \neq 0 \text{ for all } 1 \leq k \leq d. \quad (3.6)$$

**Lemma 3.1.** *Suppose that  $f^*$  is such that  $\theta_k \neq 0$  for all  $1 \leq k \leq d$  and let  $\mu < 1/2$ , then  $\varphi^*$  is injective and  $\varphi^{*-1}$  is uniformly Lipschitz.*

Hence, if  $\theta_k \neq 0$  for all  $1 \leq k \leq d$  and  $\mu < 1/2$ , the hypotheses of Theorems 3.1 and 3.2 are verified. Thus the geometric PCA is consistent, in the case of template registration, with translation operator. Observe that hypotheses of Lemma 3.1 imply that translation invariant templates  $f^*$  are excluded.

We now turn our attention to groupwise registration. We have to impose further conditions on model (3.1) ensuring  $g$ -identifiability, so that Theorem 3.3 applies. The set of deformations parameters  $\mathcal{U} \subset \mathcal{V}_\mu$  over which  $M^g(\mathbf{v}, \mathbf{y})$  will be minimized, is  $\mathcal{U} = \mathcal{U}_0$ , given in (2.2). We have the following.

**Proposition 3.1.** *Suppose  $\theta_k \neq 0$  for all  $1 \leq k \leq d$ , and that  $\mathbb{P}(V \in [-\rho, \rho]^d) = 1$  with  $\rho = \min(\frac{\mu}{2}, \frac{\mu}{\sqrt{d}})$  and  $0 < \mu < \frac{1}{12}$ . Then*

$$D^g(\mathbf{u}, \mathbf{v}) - D^g(\mathbf{u}^*(\mathbf{v}), \mathbf{v}) \geq C(f^*, \mu) \bar{d}^2(\mathbf{u}, \mathbf{u}^*(\mathbf{v})), \text{ for all } \mathbf{u} \in \mathcal{U}, \quad (3.7)$$

where  $\mathbf{u}^*(\mathbf{v}) := (v_1 - \frac{1}{n} \sum_{i=1}^n v_i, \dots, v_n - \frac{1}{n} \sum_{i=1}^n v_n)$  and  $C(f^*, \mu) > 0$  is a constant depending only on  $f^*$  and  $\mu$ .

Remark that in Proposition 3.1,  $D^g(\mathbf{u}^*(\mathbf{v}), \mathbf{v}) = 0$ . This shows that  $D^g(\mathbf{u}, \mathbf{v})$  is bounded below by a quadratic functional.

We are now ready to prove  $g$ -identifiability under the hypotheses of the previous proposition. Observe that (i) in Definition 3.3 follows at once from (3.7). For (ii) note that  $\bar{d}^2(\mathbf{u}^*(\mathbf{V}), \mathbf{V}) = \|\frac{1}{n} \sum_{i=1}^n V_i\|^2$ . Hence, given that  $\mathbb{E}V = 0$ , from Bernstein's inequality for bounded random variables in a Hilbert space (see e.g. [Bos00], Theorem 2.6) we conclude that, for any  $\eta > 0$ ,

$$\mathbb{P}(d(\mathbf{u}^*(\mathbf{V}), \mathbf{V}) > \eta) \leq 2 \exp\left(-\frac{n\eta^2}{2\mathbb{E}\|V\|^2 + \frac{\mu}{3}\eta}\right).$$

Therefore  $\bar{d}^2(\mathbf{u}^*(\mathbf{V}), \mathbf{V})$  converges in probability to 0 as  $n \rightarrow +\infty$ .

Finally, having checked the  $g$ -identifiability of the model, we conclude that the geometric PCA is consistent, in the case of groupwise registration, with translation operator.

## 4 Conclusion and discussion

The contribution of this paper is two folds. First, the use of deformation operators (as introduced in this paper) provides a general framework for modelling and analyzing the shape variability of a

set of images. As a particular case, it allows the use of diffeomorphic deformations parametrized by stationary vector fields. In the case of diffeomorphisms computed with nonstationary vector fields as in [BMTY05], the link with our framework is not straightforward. Indeed, in this setting, there are two possibilities for defining the deformation operators. One can parameterizing them either by the Hilbert space of time-dependent vector fields, or by the Hilbert space of initial velocities. Both cases are rather complex from the analytical and the computational point of views and treating them is beyond the scope of this paper. In contrast, due to its analytical and numerical tractability, we have preferred to focus on diffeomorphic deformation operators parametrized by stationary vector fields.

The second contribution of this paper is the study of the consistency of geometric PCA in statistical deformable models which, to the best of our knowledge, has not been investigated before. We also hope that the methods presented in this paper will stimulate further investigation into the development of consistent statistical procedures for the analysis of shape variability.

## A Preliminary technical results

In this sub-section, we give a deviation in probability of  $\sup_{\mathbf{v} \in \mathbf{V}_\mu} |D^g(\mathbf{v}) - M^g(\mathbf{v}, \mathbf{Y})|$  under appropriate assumptions on the deformation operators and the additive noise in model (3.1), where  $M^g(\mathbf{v}, \mathbf{Y})$  and  $D^g(\mathbf{v})$  are defined in (2.4) and (3.5) respectively.

**Lemma A.1.** *Consider model (3.1), with  $\mu$ -regular deformation operator  $\varphi$ . Let*

$$Q(\mathbf{v}) = \frac{\epsilon^2}{n} \sum_{i=1}^n \int_{\Omega} W_i^2(\varphi(v_i, x)) dx, \quad \mathbf{v} \in \mathbf{V}_\mu. \quad (\text{A.1})$$

Then, for any  $s > 0$ ,

$$\mathbb{P} \left( \sup_{\mathbf{v} \in \mathbf{V}_\mu} Q(\mathbf{v}) \geq A_\mu h(s, n, \epsilon) \right) \leq \exp(-s),$$

where  $A_\mu$  is given in Definition 3.2 (i) and  $h(s, n, \epsilon) = \epsilon^2 (1 + 2\frac{s}{n} + 2\sqrt{\frac{s}{n}})$ .

*Proof.* From Definition 3.2 (i), we have, for  $\mathbf{v} \in \mathbf{V}_\mu$ ,

$$Q(\mathbf{v}) \leq A_\mu \frac{\epsilon^2}{n} \sum_{i=1}^n \int_{\Omega} W_i^2(x) dx. \quad (\text{A.2})$$

Let  $g \in L^2(\Omega)$  and  $K^W g(x) = \int_{\Omega} k(x, y) g(y) dy$  be the covariance operator of the random process  $W$ , where  $k(x, y) = \mathbb{E}W(x)W(y)$  for  $x, y \in \Omega$ . Then, there exist orthonormal eigenfunctions  $(\phi_\lambda)_{\lambda \in \Lambda}$  in  $L^2(\Omega)$  and positive eigenvalues  $(w_\lambda)_{\lambda \in \Lambda}$ , such that  $K^W \phi_\lambda = w_\lambda \phi_\lambda$ , with  $w_1 \geq w_2 \geq \dots \geq 0$  and  $\Lambda = \{1, 2, \dots\}$ . For any  $1 \leq i \leq n$ , the Gaussian process  $W_i$  can thus be decomposed as

$$W_i = \sum_{\lambda \in \Lambda} w_\lambda^{1/2} \xi_{i,\lambda} \phi_\lambda,$$

where  $\xi_{i,\lambda} = w_\lambda^{-1/2} \langle W_i, \phi_\lambda \rangle_2$  is a Gaussian variable with zero mean and variance 1, such that  $\mathbb{E} \xi_{i,\lambda} \xi_{i,\lambda'} = 0$  for  $\lambda \neq \lambda'$ . Therefore,  $\|W_i\|_2^2 = \sum_{\lambda=1}^{+\infty} w_\lambda \xi_{i,\lambda}^2$  where  $\xi_{i,k}, i = 1 \dots, n, k \geq 1$  are i.i.d. standard Gaussian random variables. We have, from the assumptions on  $W$ ,  $\mathbb{E} \|W_i\|_2^2 = \sum_{\lambda=1}^{+\infty} w_\lambda = 1 < +\infty$ , and one can thus consider the following centered random variable

$$Z = \sum_{i=1}^n \sum_{\lambda=1}^{+\infty} w_\lambda (\xi_{i,\lambda}^2 - 1).$$

Let  $0 < t < (2w_1)^{-1}$ . Since the generating function of a  $\chi^2$  random variable, with one degree of freedom, is  $\mathbb{E} \left( e^{s\xi_{i,\lambda}^2} \right) = (1 - 2s)^{-1/2}$ , for  $s > 0$ , it follows that

$$\log (\mathbb{E} (e^{tZ})) = -n \sum_{\lambda=1}^{+\infty} \left( tw_\lambda + \frac{1}{2} \log(1 - 2tw_\lambda) \right). \quad (\text{A.3})$$

Then, using the inequality  $-s - \frac{1}{2} \log(1 - 2s) \leq \frac{s^2}{1-2s}$ , which holds for all  $0 < s < \frac{1}{2}$ , from (A.3) we obtain

$$\log (\mathbb{E} (e^{tZ})) \leq n \sum_{\lambda=1}^{+\infty} \frac{t^2 w_\lambda^2}{1 - 2tw_\lambda} \leq \frac{t^2 n}{1 - 2tw_1} \left( \sum_{\lambda=1}^{+\infty} w_\lambda \right)^2 = \frac{t^2 n}{1 - 2tw_1} < \infty.$$

Arguing e.g. as in [BM98], the above inequality implies that, for any  $s > 0$ ,

$$\mathbb{P} (Z > 2w_1 s + 2\sqrt{ns}) \leq \exp(-s). \quad (\text{A.4})$$

By (A.2), it follows that

$$\sup_{\mathbf{v} \in \mathcal{V}_\mu} Q(\mathbf{v}) \leq A_\mu \frac{\epsilon^2}{n} \sum_{i=1}^n \int_{\Omega} W_i^2(x) dx = A_\mu \frac{\epsilon^2}{n} \sum_{i=1}^n \sum_{\lambda=1}^{+\infty} w_\lambda \xi_{i,\lambda}^2 = A_\mu \frac{\epsilon^2}{n} (Z + n).$$

Hence, it follows from (A.4) that

$$\mathbb{P} \left( \sup_{\mathbf{v} \in \mathcal{V}_\mu} Q(\mathbf{v}) > A_\mu \frac{\epsilon^2}{n} (n + 2w_1 s + 2\sqrt{ns}) \right) \leq \exp(-s),$$

for any  $s > 0$ . The conclusion follows noting that  $w_1 \leq \mathbb{E} \|W\|_2^2 = 1$ .  $\square$

**Lemma A.2.** Consider model (3.1), with  $\mu$ -regular deformation operator  $\varphi$ . Let

$$D^t(\mathbf{v}) := \frac{1}{n} \sum_{i=1}^n \int_{\Omega} (f_i^*(\varphi(v_i, x)) - f^*(x))^2 dx, \quad \mathbf{v} \in \mathcal{V}, \quad (\text{A.5})$$

with  $f_i^*(x) := f^*(\varphi^{-1}(V_i, x))$ ,  $x \in \Omega$ ,  $i = 1, \dots, n$ . Then

$$\mathbb{P} \left( \sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^t(\mathbf{v}) - M^t(\mathbf{v}, \mathbf{Y}, f^*)| > C \left( h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)} \right) \right) \leq \exp(-s), \quad s > 0,$$

where  $M^t$  is defined in (2.3),  $C > 0$  is a constant depending only on  $f^*$  and  $\mu$ , and  $h(s, n, \epsilon)$  is defined in Lemma A.1.

*Proof.* For  $\mathbf{v} \in \mathcal{V} = \mathcal{V}^n$ , let

$$R(\mathbf{v}) = 2\epsilon \frac{1}{n} \sum_{i=1}^n \int_{\Omega} (f_i^*(\varphi(v_i, x)) - f^*(x)) W_i^*(\varphi(v_i, x)) dx.$$

For any  $\mathbf{v} \in \mathcal{V}_\mu$ , we have the decomposition

$$M^t(\mathbf{v}, \mathbf{Y}, f^*) = D^t(\mathbf{v}) + Q(\mathbf{v}) + R(\mathbf{v}), \quad (\text{A.6})$$

where  $Q$  is defined in (A.1).

By applying the Cauchy-Schwartz inequality in  $L^2(\Omega)$  and in  $\mathbb{R}^n$  we obtain  $R(\mathbf{v}) \leq 2\sqrt{D^t(\mathbf{v})}\sqrt{Q(\mathbf{v})}$ . Also, the  $\mu$ -regularity of  $\varphi$  implies  $D^t(\mathbf{v}) \leq 4A_\mu^2 \|f^*\|_2^2$ , and therefore  $R(\mathbf{v}) \leq 4A_\mu \|f^*\|_2 \sqrt{Q(\mathbf{v})}$ . Now, using the decomposition (A.6), one obtains

$$\sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^t(\mathbf{v}) - M^t(\mathbf{v}, \mathbf{Y}, f^*)| \leq \max(1, 4A_\mu \|f^*\|_2) \left( \sup_{\mathbf{v} \in \mathcal{V}_\mu} Q(\mathbf{v}) + \sup_{\mathbf{v} \in \mathcal{V}_\mu} \sqrt{Q(\mathbf{v})} \right)$$

and the result follows from Lemma A.1.  $\square$

**Lemma A.3.** Consider model (3.1), with  $\mu$ -regular deformation operator  $\varphi$ . Then

$$\mathbb{P} \left( \sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^g(\mathbf{v}, \mathbf{V}) - M^g(\mathbf{v}, \mathbf{Y})| > C \left( h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)} \right) \right) \leq \exp(-s), \quad s > 0,$$

where  $M^g$  and  $D^g$  are defined in (2.4) and (3.5) respectively;  $C > 0$  is a constant, depending only on  $f^*$  and  $\mu$ , and  $h(s, n, \epsilon)$  is defined in Lemma A.1.

*Proof.* Let

$$Q^g(\mathbf{v}) = \frac{\epsilon^2}{n} \sum_{i=1}^n \int_{\Omega} \left( W_i(\varphi(v_i, x)) - \frac{1}{n} \sum_{j=1}^n W_j(\varphi(v_j, x)) \right)^2 dx, \quad \mathbf{v} \in \mathcal{V}$$

and

$$R(\mathbf{v}) = 2\epsilon \frac{1}{n} \sum_{i=1}^n \int_{\Omega} \left( \frac{1}{n} \sum_{j=1}^n f_j^*(\varphi(v_j, x)) - f_i^*(\varphi(v_i, x)) \right) \times \left( \frac{1}{n} \sum_{j=1}^n W_j^*(\varphi(v_j, x)) - W_i^*(\varphi(v_i, x)) \right) dx, \quad \mathbf{v} \in \mathcal{V}.$$

Then, for any  $\mathbf{v} \in \mathcal{V}_\mu$ , we have the decomposition

$$M^g(\mathbf{v}, \mathbf{Y}) = D^g(\mathbf{v}) + Q^g(\mathbf{v}) + R(\mathbf{v}). \quad (\text{A.7})$$

From the Cauchy-Schwartz inequality in  $L^2(\Omega)$  and in  $\mathbb{R}^n$ , we have  $R(\mathbf{v}) \leq 2\sqrt{D^g(\mathbf{v})}\sqrt{Q^g(\mathbf{v})}$ . Also, from the  $\mu$ -regularity of  $\varphi$ , we obtain  $D^g(\mathbf{v}) \leq \frac{1}{n} \sum_{i=1}^n \int_{\Omega} (f_i^*(\varphi^{-1}(v_i, x)))^2 dx \leq A_\mu^2 \|f^*\|_2^2$ . So  $R(\mathbf{v}) \leq 2A_\mu \|f^*\|_2 \sqrt{Q^g(\mathbf{v})}$ . Now, using the decomposition (A.7), one obtains

$$\sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^g(\mathbf{v}) - M^g(\mathbf{v}, \mathbf{Y})| \leq \max(1, 2A_\mu \|f^*\|_2) \left( \sup_{\mathbf{v} \in \mathcal{V}_\mu} Q^g(\mathbf{v}) + \sup_{\mathbf{v} \in \mathcal{V}_\mu} \sqrt{Q^g(\mathbf{v})} \right)$$

and the result follows from the fact that  $Q^g \leq Q$  (see (A.1)) and Lemma A.1.  $\square$

**Remark A.1.** Observe that Lemma A.2 and Lemma A.3 imply

$$\text{plim}_{\epsilon} \sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^t(\mathbf{v}) - M^t(\mathbf{v}, \mathbf{Y}, f^*)| = \text{plim}_{\epsilon} \sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^g(\mathbf{v}, \mathbf{V}) - M^g(\mathbf{v}, \mathbf{Y})| = 0, \quad (\text{A.8})$$

and

$$\text{plim}_{n, \epsilon} \sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^t(\mathbf{v}) - M^t(\mathbf{v}, \mathbf{Y}, f^*)| = \text{plim}_{n, \epsilon} \sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^g(\mathbf{v}, \mathbf{V}) - M^g(\mathbf{v}, \mathbf{Y})| = 0. \quad (\text{A.9})$$

The proofs of Theorems 3.1, 3.2 and 3.3 rely on the following two propositions that establish the consistency of the registration procedures described in Section 2.2.

**Proposition A.1.** Let  $\hat{\mathbf{V}} \in \arg \min_{\mathbf{v} \in \mathcal{V}_\mu} M^t(\mathbf{v}, \mathbf{Y}, f^*)$  the parameters obtained from template registration of  $\mathbf{Y}$  on  $f^*$ . Then,

(i) under the hypotheses of Theorem 3.1,  $\text{plim}_{n, \epsilon} \bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) = 0$ , and

(ii) under the hypotheses of Theorem 3.2,

$$\mathbb{P} \left( \bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) > C \left( h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)} \right) \right) \leq \exp(-s), \quad s > 0,$$

where  $C > 0$  is a constant, depending only on  $f^*$ ,  $\mu$  and  $h(s, n, \epsilon)$  is defined in Lemma A.1.

*Proof.* Observe that  $D^t(\mathbf{V}) = 0$  and so

$$D^t(\mathbf{v}) = D^t(\mathbf{v}) - D^t(\mathbf{V}) \leq 2 \sup_{\mathbf{v} \in \mathcal{V}_\mu} |M^t(\mathbf{v}, \mathbf{Y}, f^*) - D^t(\mathbf{v})|, \quad \mathbf{v} \in \mathcal{V}_\mu, \quad (\text{A.10})$$

where  $D^t$  and  $M^t$  are defined in (A.5) and (2.3) respectively. On the other hand, from Definition 3.2 (i),

$$\frac{1}{n} \sum_{i=1}^n \|\varphi^*(V_i) - \varphi^*(v_i)\|_2^2 \leq A_\mu D^t(\mathbf{v}), \quad \mathbf{v} \in \mathcal{V}_\mu.$$

Hence,

$$\frac{1}{n} \sum_{i=1}^n \|\varphi^*(V_i) - \varphi^*(\hat{V}_i)\|_2^2 \leq 2A_\mu \sup_{\mathbf{v} \in \mathcal{V}_\mu} |M^t(\mathbf{v}, \mathbf{Y}, f^*) - D^t(\mathbf{v})|. \quad (\text{A.11})$$

We proceed now to prove part (i). From (A.11) and (A.8) we have  $\text{plim}_\epsilon \|\varphi^*(V_i) - \varphi^*(\hat{V}_i)\|_2^2 = 0$ , that is,  $\text{plim}_\epsilon \varphi^*(\hat{V}_i) = \varphi^*(V_i)$  for  $i = 1, \dots, n$ . From the continuity of  $\varphi^{*-1}$  we have  $\text{plim}_\epsilon \hat{V}_i = V_i$ , for  $i = 1, \dots, n$ , therefore

$$\text{plim}_n \text{plim}_\epsilon \frac{1}{n} \sum_{i=1}^n \|V_i - \hat{V}_i\|^2 = 0.$$

Now, the fact that  $\|V_1 - \hat{V}_1\|^2$  is bounded by  $2\mu$  and tends to 0 in probability as  $\epsilon \rightarrow 0$ , implies that  $\mathbb{E}\|V_1 - \hat{V}_1\|^2 \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Noting that  $(V_i - \hat{V}_i)_{i \geq 1}$  are i.i.d. (see Remark 3.1), we conclude from the weak law of large number that

$$\text{plim}_\epsilon \text{plim}_n \frac{1}{n} \sum_{i=1}^n \|V_i - \hat{V}_i\|^2 = 0$$

thus proving part (i).

For (ii), note that inequality A.11 and that  $\varphi^{*-1}$  is uniformly Lipschitz, with constant  $L(f^*, \mu) > 0$ , imply

$$\bar{d}^2(\mathbf{u}, \mathbf{v}) \leq 2A_\mu L(f^*, \mu) \sup_{\mathbf{v} \in \mathcal{V}} |M^t(\mathbf{v}, \mathbf{Y}, f^*) - D^t(\mathbf{v})|$$

and the result follows from Lemma A.2.  $\square$

**Proposition A.2.** *Let  $\hat{\mathbf{V}} \in \arg \min_{\mathbf{v} \in \mathcal{U}} M^g(\mathbf{v}, \mathbf{Y})$  the parameters obtained from groupwise registration of  $\mathbf{Y}$ . Then, under the hypotheses of Theorem 3.3,  $\text{plim}_{n, \epsilon} \bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) = 0$ .*

*Proof.* Let  $\mathbf{u}^*(\mathbf{V})$  be the unique minimizer of  $D^g(u, \mathbf{V})$  on  $\mathcal{U}$ , which exists because the model is  $g$ -identifiable. Since (by definition)  $\hat{\mathbf{V}} \in \arg \min_{\mathbf{v} \in \mathcal{U}} M^g(\mathbf{v}, \mathbf{Y})$ , one obtains that

$$\begin{aligned} D^g(\hat{\mathbf{V}}, \mathbf{V}) - D^g(\mathbf{u}^*, \mathbf{V}) &\leq 2 \sup_{\mathbf{u} \in \mathcal{U}} |M^g(\mathbf{u}, \mathbf{Y}) - D^g(\mathbf{u}, \mathbf{V})| \\ &\leq 2 \sup_{\mathbf{u} \in \mathcal{V}_\mu} |M^g(\mathbf{u}, \mathbf{Y}) - D^g(\mathbf{u}, \mathbf{V})|. \end{aligned}$$

Therefore, from (A.9) and the  $g$ -identifiability of the model, we have  $\text{plim}_{n, \epsilon} \bar{d}^2(\hat{\mathbf{V}}, \mathbf{u}^*) = 0$ . Also, the  $g$ -identifiability implies that  $\text{plim}_{n, \epsilon} \bar{d}^2(\mathbf{u}^*, \mathbf{V}) = 0$ . Finally, the conclusion follows from the inequality  $\bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) \leq 2\bar{d}^2(\hat{\mathbf{V}}, \mathbf{u}^*) + 2\bar{d}^2(\mathbf{u}^*, \mathbf{V})$ .  $\square$

In what follows,  $\| \cdot \|_{HS}$  denotes the Hilbert-Schmidt norm of operators on a Hilbert space  $\mathcal{H}$ . Recall that, given an orthonormal basis  $\{e_j\}_{j \geq 1}$  of  $\mathcal{H}$ , the Hilbert-Schmidt norm of an operator  $K$  is defined as  $\|K\|_{HS}^2 = \sum_{j,k} \langle K(e_j), e_k \rangle^2$ .

**Lemma A.4.** *Let  $\mathcal{H}$  be a separable Hilbert space with inner product  $\langle \cdot, \cdot \rangle$  and induced norm  $\| \cdot \|$ . Let  $\{u_i\}_{i=1}^n, \{v_i\}_{i=1}^n$  in  $B_r = \{h \in \mathcal{H} : \|h\| \leq r\}$  for some  $r > 0$ . Define the covariance operators  $K_u, K_v : \mathcal{H} \rightarrow \mathcal{H}$  by  $K_u(h) = \frac{1}{n} \sum_{i=1}^n \langle u_i - \bar{u}, h \rangle (u_i - \bar{u})$  and  $K_v(h) = \frac{1}{n} \sum_{i=1}^n \langle v_i - \bar{v}, h \rangle (v_i - \bar{v})$ , where  $\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i$  and  $\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$ . Then*

$$\|K_v - K_u\|_{HS}^2 \leq (6r)^2 \frac{1}{n} \sum_{i=1}^n \|v_i - u_i\|^2.$$

*Proof.* Let define  $\epsilon_i = v_i - u_i$ , so we can write  $v_i = u_i + \epsilon_i$ . Let  $h \in H$  and write

$$K_v(h) = \frac{1}{n} \sum_{i=1}^n \langle u_i - \bar{u} + \epsilon_i - \bar{\epsilon}, h \rangle (u_i - \bar{u} + \epsilon_i - \bar{\epsilon}) = K_u + L + L^* + S,$$

where  $L^*$  denotes the adjoint the operator of  $L = \frac{1}{n} \sum_{i=1}^n \langle u_i - \bar{u}, h \rangle (\epsilon_i - \bar{\epsilon})$  and  $S = \frac{1}{n} \sum_{i=1}^n \langle \epsilon_i - \bar{\epsilon}, h \rangle (\epsilon_i - \bar{\epsilon})$ . Then, after some simple calculations, we get

$$\|L\|_{HS}^2 = \sum_{j \geq 1} \sum_{k \geq 1} \left( \frac{1}{n} \sum_{i=1}^n \langle u_i - \bar{u}, e_j \rangle \langle \epsilon_i - \bar{\epsilon}, e_k \rangle \right)^2 \leq \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n \|u_i - \bar{u}\| \|u_{i'} - \bar{u}\| \|\epsilon_i - \bar{\epsilon}\| \|\epsilon_{i'} - \bar{\epsilon}\|.$$

Hence, since  $u_i \in B_r, i = 1, \dots, n$ , one has

$$\|L\|_{HS}^2 \leq \left( \frac{1}{n} \sum_{i=1}^n \|u_i - \bar{u}\| \|\epsilon_i - \bar{\epsilon}\| \right)^2 \leq (2r)^2 \left( \frac{1}{n} \sum_{i=1}^n \|\epsilon_i - \bar{\epsilon}\| \right)^2 \leq (2r)^2 \frac{1}{n} \sum_{i=1}^n \|\epsilon_i\|^2.$$

Similarly, since  $\|\epsilon_i\| \leq \|u_i\| + \|v_i\| \leq 2r$ ,

$$\|S\|_{HS}^2 \leq \left( \frac{1}{n} \sum_{i=1}^n \|\epsilon_i - \bar{\epsilon}\|^2 \right) \leq \left( \frac{1}{n} \sum_{i=1}^n \|\epsilon_i\|^2 \right) \leq (2r)^2 \frac{1}{n} \sum_{i=1}^n \|\epsilon_i\|^2.$$

Finally,  $\|K_v - K_u\|_{HS} \leq 2\|L\|_{HS} + \|S\|_{HS} \leq 6r \left( \sum_{i=1}^n \|\epsilon_i\|^2 \right)^{\frac{1}{2}}$ , which completes the proof.  $\square$

The following theorem follows from the theory developed in [BDM83, DK70].

**Theorem A.1.** *Let  $\mathcal{H}$  be a separable Hilbert space endowed with inner product  $\langle \cdot, \cdot \rangle_H$ . Let  $A, \hat{A} : H \rightarrow H$  be self-adjoint Hilbert-Schmidt operators on  $\mathcal{H}$ , with eigenvalues/eigenvectors pairs  $(\kappa_\lambda, \phi_\lambda)_{\lambda \geq 1}$  and  $(\hat{\kappa}_\lambda, \hat{\phi}_\lambda)_{\lambda \geq 1}$  respectively. Then,*

$$\sup_{\lambda \geq 1} |\kappa_\lambda - \hat{\kappa}_\lambda| \leq \|A - \hat{A}\|_{HS}. \quad (\text{A.12})$$

Moreover, if  $\hat{\delta}_\lambda = \min_{\lambda' \in \Lambda \setminus \{\lambda\}} |\kappa_\lambda - \hat{\kappa}_{\lambda'}| > 0$ , then

$$\sin(\phi_\lambda, \hat{\phi}_\lambda) \leq \hat{\delta}_\lambda^{-1} \|A - \hat{A}\|_{HS}. \quad (\text{A.13})$$

## B Proofs of main results

### B.1 Proof of Theorem 3.1

*Proof.* Let  $\tilde{K}_n$  be the sample covariance operator  $V_1, \dots, V_n$ , that is  $\tilde{K}_n v = \frac{1}{n} \sum_{i=1}^n \langle V_i - \bar{V}_n, v \rangle (V_i - \bar{V}_n)$ , with  $\bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_i$ . Note that

$$\|\hat{K}_n - K\|_{HS}^2 \leq 2\|\hat{K}_n - \tilde{K}_n\|_{HS}^2 + 2\|\tilde{K}_n - K\|_{HS}^2, \quad (\text{B.1})$$

The first term in the right hand side of inequality (B.1) can be controlled by using Lemma A.4 and noting that  $\|V_i\|, \|\hat{V}_i\| \leq \mu, i = 1, \dots, n$ , that is,

$$\|\hat{K}_n - \tilde{K}_n\|_{HS}^2 \leq (6\mu)^2 \frac{1}{n} \sum_{i=1}^n \|\hat{V}_i - V_i\|^2. \quad (\text{B.2})$$

Let us now bound the second term in the right hand side of (B.1). To do so, remark that  $\|V\|, \|V_i\| \leq \mu, i = 1, \dots, n$ , and, thanks to a Bernstein's inequality for Hilbert-Schmidt operators (see e.g. [Bos98], Chapter 3), it follows that

$$\mathbb{P}\left(\|\tilde{K}_n - K\|_{HS} > \eta\right) \leq 2 \exp\left(-\frac{n\eta^2}{\tilde{C}(\mu)(1+\eta)}\right), \quad \eta > 0, \quad (\text{B.3})$$

for some constant  $\tilde{C}(\mu) > 0$  depending only on  $\mu$ . Hence, we can combine (B.1), (B.2) and (B.3) with Proposition A.1 (i) to obtain that for any  $\eta > 0$ ,  $\lim_{n,\epsilon} \mathbb{P}\left(\|\hat{K}_n - K\|_{HS}^2 > \eta\right) = 0$ , that is,

$$\text{plim}_{n,\epsilon} \hat{K}_n = K. \quad (\text{B.4})$$

Now, from (A.12) and (B.4) we obtain  $\text{plim}_{n,\epsilon} \hat{\kappa}_\lambda = \kappa_\lambda, \lambda \in \Lambda$ . For  $\lambda \in \Lambda$  define the  $\lambda$ -th empirical eigen-gap as

$$\hat{\delta}_\lambda = \min_{\lambda' \in \Lambda \setminus \{\lambda\}} |\kappa_\lambda - \hat{\kappa}_{\lambda'}|.$$

From (A.12), it holds that  $\delta_\lambda \leq \hat{\delta}_\lambda + \max_{\lambda'} |\kappa_{\lambda'} - \hat{\kappa}_{\lambda'}| \leq \hat{\delta}_\lambda + \|\hat{K}_n - K\|_{HS}$ . By (B.4), it follows that

$$\lim_{n,\epsilon} \mathbb{P}\left(\hat{\delta}_\lambda > \frac{\delta_\lambda}{2}\right) = 1. \quad (\text{B.5})$$

Recalling that, from the specification of model (3.1), we have  $\delta_\lambda > 0$ , hence inequality (A.13) implies

$$\begin{aligned} \mathbb{P}\left(\sin(\hat{\phi}_\lambda, \phi_\lambda) > \eta\right) &\leq \mathbb{P}\left(\|\hat{K}_n - K\|_{HS} / \hat{\delta}_\lambda > \eta\right) \\ &= \mathbb{P}\left(\|\hat{K}_n - K\|_{HS} / \hat{\delta}_\lambda > \eta, \hat{\delta}_\lambda > \delta/2\right) + \mathbb{P}\left(\|\hat{K}_n - K\|_{HS} / \hat{\delta}_\lambda > \eta, \hat{\delta}_\lambda \leq \delta/2\right) \\ &\leq \mathbb{P}\left(\|\hat{K}_n - K\|_{HS} > (\delta_\lambda \eta) / 2\right) + \mathbb{P}\left(\hat{\delta}_\lambda \leq \delta/2\right). \end{aligned}$$

From the above inequality, combined with B.4 and B.5, we obtain  $\text{plim}_{n,\epsilon} \sin^2(\hat{\phi}_\lambda, \phi_\lambda) = 0$ .  $\square$

## B.2 Proof of Theorem 3.2

*Proof.* Combining (B.1), (B.2) and (B.3) with Proposition A.1 (ii), we obtain

$$\mathbb{P} \left( \|\hat{K}_n - K\|_{HS}^2 > C \max(h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)}; g(s, n)) \right) \leq 2 \exp(-s), \quad s > 0,$$

where  $C > 0$  is a constant depending only on  $f^*$  and  $\mu$  and  $g(s, n) = \left( \frac{s}{n} + \sqrt{\frac{s^2}{n^2} + \frac{s}{n}} \right)^2$ . Hence, from (A.12) we obtain

$$\mathbb{P} \left( |\hat{\kappa}_\lambda - \kappa_\lambda|^2 > C \max(h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)}; g(s, n)) \right) \leq 2 \exp(-s), \quad s > 0.$$

Take now  $s^* > 0$  such that  $C \max(h(s^*, n, \epsilon) + \sqrt{h(s^*, n, \epsilon)}; g(s^*, n)) < (\delta_\lambda/2)^2$ . Then, thanks to (B.5) and (A.13) we obtain, for any  $0 < s \leq s^*$ ,

$$\begin{aligned} 1 - 2 \exp(-s) &\leq \mathbb{P} \left( \|\hat{K}_n - K\|_{HS}^2 < C \max(h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)}; g(s, n)) \right) \\ &= \mathbb{P} \left( \|\hat{K}_n - K\|_{HS}^2 < C \max(h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)}; g(s, n)), \hat{\delta}_\lambda > \delta_\lambda/2 \right) \\ &\leq \mathbb{P} \left( (1/\hat{\delta}_\lambda)^2 \|\hat{K}_n - K\|_{HS}^2 < (2/\delta_\lambda)^2 C \max(h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)}; g(s, n)) \right) \\ &\leq \mathbb{P} \left( \sin^2(\hat{\phi}_\lambda, \phi_\lambda) < (2/\delta_\lambda)^2 C \max(h(s, n, \epsilon) + \sqrt{h(s, n, \epsilon)}; g(s, n)) \right). \end{aligned}$$

$\square$

## B.3 Proof of Theorem 3.3

*Proof.* We proceed similarly as in the proof of Theorem 3.1. In the case of groupwise registration, inequalities (B.1), (B.2) and (B.3) are still valid, and can be combined with Proposition A.2 to obtain  $\text{plim}_{n,\epsilon} \hat{K}_n = K$ . The rest of proof is identical to that of Theorem 3.1.  $\square$

## C Technical results for translation operators

**Lemma C.1.** *Let  $\varphi$  be defined by (2.1), then*

(i)  $d_\Omega(\varphi(v, x), \varphi(v, y)) = d_\Omega(x, y)$  for all  $x, y \in \Omega$  and  $v \in \mathcal{V}$ .

(ii)  $d_C(\varphi(u, \cdot), \varphi(v, \cdot)) \leq \sum_{k=1}^d |u_k - v_k|$  for all  $x \in \Omega$  and  $u, v \in \mathcal{V}$ .

*Proof.* Remark that, for any  $a \in \mathbb{R}$ , there exists a unique  $k(a) \in \mathbb{Z}$  such that  $\text{mod}(a, 1) = a + k(a)$ . Then

$$\text{mod}(a, 1) - \text{mod}(b, 1) = a - b + k(a) - k(b).$$

Take  $a, b \in \mathbb{R}$  such that  $|a - b| < 1$  and assume that  $a \geq b$ . Since  $a - b \in [0, 1)$  and  $\text{mod}(a, 1) - \text{mod}(b, 1) \in [-1, 1]$ , we obtain that

$$k(a) - k(b) = \begin{cases} 0 & \text{if } \text{mod}(a, 1) \geq \text{mod}(b, 1) \\ -1 & \text{if } \text{mod}(a, 1) < \text{mod}(b, 1). \end{cases}$$

Then

$$|\text{mod}(a, 1) - \text{mod}(b, 1)| = \begin{cases} a - b & \text{if } \text{mod}(a, 1) \geq \text{mod}(b, 1) \\ 1 - (a - b) & \text{if } \text{mod}(a, 1) < \text{mod}(b, 1) \end{cases}$$

We conclude that for  $a \geq b$

$$\min\{|\text{mod}(a, 1) - \text{mod}(b, 1)|, 1 - |\text{mod}(a, 1) - \text{mod}(b, 1)|\} = \min\{|b - a|, 1 - |b - a|\}. \quad (\text{C.1})$$

Because of the symmetry in the expression above, we conclude that (C.1) is valid for any  $a, b \in \mathbb{R}$ , such that  $|a - b| < 1$ .

For the sake of simplicity, let us prove the lemma in the one-dimensional case (i.e.  $d = 1$ ), where  $d_\Omega(x, y) := \min\{|x - y|, 1 - |x - y|\}$ . Take  $x, y \in \Omega$  and  $u, v \in \mathcal{V}$ . Part (i) is directly implied by (C.1), taking  $a := x + v$  and  $b := y + v$ . For part (ii), note that  $d_\Omega \leq \frac{1}{2}$ , hence  $d_\Omega(\varphi(u, x), \varphi(v, x)) \leq |u - v|$  if  $|u - v| \geq 1$ . In the other had, if  $|u - v| < 1$  we can use (C.1) with  $a := x + v$  and  $b := x + u$  to obtain  $d_\Omega(\varphi(u, x), \varphi(v, x)) \leq d_\Omega(u, v) \leq |u - v|$ . Finally, we obtain  $d_C(\varphi(u, \cdot), \varphi(v, \cdot)) \leq |u - v|$ .  $\square$

In order to prove Lemma 3.1 and Proposition 3.1, denote by  $e_\ell(x) = e^{i2\pi \sum_{k=1}^d \ell_k x_k}$  for  $x = (x_1, \dots, x_d) \in \Omega = [0, 1]^d$  and  $\ell = (\ell_1, \dots, \ell_d) \in \mathbb{Z}^d$  the Fourier basis of  $L^2([0, 1]^d)$ . Let  $\theta_\ell = \int_\Omega f(x) e_\ell(x) dx$ ,  $\ell \in \mathbb{Z}^d$  be the Fourier coefficients of  $f^*$ . For  $1 \leq k \leq d$ , denote by  $\ell^{(k)} = (\ell_1^{(k)}, \dots, \ell_d^{(k)})$  the vector of  $\mathbb{Z}^d$  such that  $\ell_{k'}^{(k)} = 0$  for  $k' \neq k$  and  $\ell_k^{(k)} = 1$ . Remark that, with this notation,  $\theta_k = \theta_{\ell^{(k)}}$ , where  $\theta_k$  is defined in (3.6).

### C.1 Proof of Lemma 3.1

*Proof.* Recall that  $\varphi^*(v) := f^* \circ \varphi^{-1}(v, \cdot)$ ,  $v \in \mathcal{V}_\mu$ . For  $u, v \in [-\rho, \rho]^d$  with  $0 < \rho < 1/2$ , we have

$$\begin{aligned} \|f^*(\varphi^{-1}(u, x)) - f^*(\varphi^{-1}(v, x))\|_2^2 &\geq \sum_{k=1}^d |\theta_{\ell^{(k)}} e^{-i2\pi u_k} - \theta_{\ell^{(k)}} e^{-i2\pi v_k}|^2 \\ &= \sum_{k=1}^d |\theta_{\ell^{(k)}}|^2 |e^{-i2\pi u_k} - e^{-i2\pi v_k}|^2. \end{aligned} \quad (\text{C.2})$$

Then, by the mean value theorem, we have  $|e^{-i2\pi u_k} - e^{-i2\pi v_k}|^2 = |\cos(2\pi u_k) - \cos(2\pi v_k)|^2 + |\sin(2\pi u_k) - \sin(2\pi v_k)|^2 \geq (2\pi)^2 \cos^2(\rho) |u_k - v_k|^2$ , for any  $0 \leq u_k, v_k \leq \rho$ . Hence,

$$\|f^*(\varphi^{-1}(u, x)) - f^*(\varphi^{-1}(v, x))\|_2^2 \geq (2\pi)^2 \cos^2(\rho) \min_{0 \leq k \leq d} |\theta_{\ell^{(k)}}|^2 \sum_{k=1}^d |u_k - v_k|^2.$$

□

## C.2 Proof of Proposition 3.1

*Proof.* Remark that  $D^g$ , defined in (3.5), has the following expression in the Fourier domain:

$$D^g(\mathbf{u}, \mathbf{V}) = \frac{1}{n} \sum_{m=1}^n \left( \sum_{\ell \in \mathbb{Z}^d} \left| \frac{1}{n} \sum_{j=1}^n \theta_{\ell} e^{-i2\pi \langle \ell, V_j - u_j \rangle} - \theta_{\ell} e^{-i2\pi \langle \ell, V_m - u_m \rangle} \right|^2 \right), \quad \mathbf{u} \in \mathcal{V}. \quad (\text{C.3})$$

For  $\mathbf{u} \in \mathcal{U}_0$  we have

$$\begin{aligned} D^g(\mathbf{u}, \mathbf{V}) &\geq \frac{1}{n} \sum_{m=1}^n \left( \sum_{k=1}^d |\theta_{\ell^{(k)}}|^2 \left| \frac{1}{n} \sum_{j=1}^n e^{-i2\pi (V_j^{(k)} - u_j^{(k)})} - e^{-i2\pi (V_m^{(k)} - u_m^{(k)})} \right|^2 \right) \\ &\geq \sum_{k=1}^d |\theta_{\ell^{(k)}}|^2 \left( 1 - \left| \frac{1}{n} \sum_{m=1}^n e^{i2\pi (u_m^{(k)} - V_m^{(k)})} \right|^2 \right). \end{aligned} \quad (\text{C.4})$$

Further, remark that

$$\left| \frac{1}{n} \sum_{m=1}^n e^{i2\pi (u_m^{(k)} - V_m^{(k)})} \right|^2 = \frac{1}{n} + \frac{2}{n^2} \sum_{m=1}^{n-1} \sum_{m'=m+1}^n \cos \left( 2\pi \left( (u_m^{(k)} - V_m^{(k)}) - (u_{m'}^{(k)} - V_{m'}^{(k)}) \right) \right).$$

Let  $0 \leq \alpha < 1/4$ . Using a second order Taylor expansion and the mean value theorem, one has that  $\cos(2\pi u) \leq 1 - C(\alpha)|u|^2$  for any real  $u$  such that  $|u| \leq \alpha$ , with  $C(\alpha) = 2\pi^2 \cos(2\pi\alpha)$ . Under the hypothesis of the proposition, one has that  $\left| (u_m^{(k)} - V_m^{(k)}) - (u_{m'}^{(k)} - V_{m'}^{(k)}) \right| \leq 2(\mu + \rho) < 1/4$ .

Therefore, for  $\alpha = 2(\mu + \rho)$ , it follows that

$$\begin{aligned} \left| \frac{1}{n} \sum_{m=1}^n e^{i2\pi (u_m^{(k)} - V_m^{(k)})} \right|^2 &\leq \frac{1}{n} + \frac{2}{n^2} \sum_{m=1}^{n-1} \sum_{m'=m+1}^n 1 - C(\alpha) \left| (u_m^{(k)} - V_m^{(k)}) - (u_{m'}^{(k)} - V_{m'}^{(k)}) \right|^2 \\ &\leq 1 - \frac{2}{n^2} \sum_{m=1}^{n-1} \sum_{m'=m+1}^n C(\alpha) \left| (u_m^{(k)} - V_m^{(k)}) - (u_{m'}^{(k)} - V_{m'}^{(k)}) \right|^2. \end{aligned}$$

Hence, using the lower bound (C.4), it follows that, for  $\mathbf{u} \in \mathcal{U}_0$ ,

$$D^g(\mathbf{u}, \mathbf{V}) \geq 2C(\alpha) \frac{1}{n^2} \sum_{m=1}^{n-1} \sum_{m'=m+1}^n \left( \sum_{k=1}^d |\theta_{\ell(k)}|^2 \left| \left( u_m^{(k)} - V_m^{(k)} \right) - \left( u_{m'}^{(k)} - V_{m'}^{(k)} \right) \right|^2 \right). \quad (\text{C.5})$$

The following identity is obtained from elementary algebraic manipulations and the fact that  $\mathbf{u} \in \mathcal{U}_0$  ( $\sum_{m=1}^n u_m^{(k)} = 0$ ).

$$\frac{1}{n} \sum_{m=1}^{n-1} \sum_{m'=m+1}^n \left| \left( u_m^{(k)} - V_m^{(k)} \right) - \left( u_{m'}^{(k)} - V_{m'}^{(k)} \right) \right|^2 = \sum_{m=1}^n \left| u_m^{(k)} - \left( V_m^{(k)} - \bar{V}_n^{(k)} \right) \right|^2,$$

where  $\bar{V}_n^{(k)} = \frac{1}{n} \sum_{m=1}^n V_m^{(k)}$ . Inserting the above equality in (C.5), we finally obtain

$$D^g(\mathbf{u}, \mathbf{V}) \geq C_0(f^*, \mu) \frac{1}{n} \sum_{m=1}^n \sum_{k=1}^d \left| u_m^{(k)} - \tilde{V}_m^{(k)} \right|^2, \quad (\text{C.6})$$

with  $C_0(f^*, \mu) = 2C(\alpha) \min_{1 \leq k \leq d} \left\{ |\theta_{\ell(k)}|^2 \right\}$  and  $\tilde{V}_m^{(k)} = V_m^{(k)} - \bar{V}_n^{(k)}$ . Thanks to the assumption  $\theta_{\ell(k)} \neq 0$  for all  $1 \leq k \leq d$ , it follows that  $C_0(f^*, \mu) > 0$ . The inequality  $\mu \geq 2\rho$ , implies that  $|\tilde{V}_m^{(k)}| = |V_m^{(k)} - \bar{V}_n^{(k)}| \leq 2\rho \leq \mu$  for any  $1 \leq k \leq d$  and all  $1 \leq m \leq n$ , therefore,  $\mathbf{u} \in \mathcal{U}_0$ . Then, using inequality (C.6) and  $D^g(\mathbf{u}^*, \mathbf{V}) = 0$ , the proof is completed.  $\square$

## References

- [AAT07] S. Allasonière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Statistical Royal Society (B)*, 69:3–29, 2007.
- [ACPA06] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In *Proc. of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Part I*, number 4190 in LNCS, pages 924–931, 2–4 October 2006.
- [AKT10] S. Allasonière, E. Kuhn, and A. Trouvé. Bayesian deformable models building via stochastic approximation algorithm: a convergence study. *Bernoulli*, 16:641–678, 2010.
- [Ash07] John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95 – 113, 2007.
- [BC11] J. Bigot and B. Charlier. On the consistency of frchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, 5:1054–1089, 2011.

- [BDM83] R. Bhatia, C. Davis, and A. McIntosh. Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl.*, 52/53:45–67, 1983.
- [BGL09] J. Bigot, S. Gadat, and J.M. Loubes. Statistical M-estimation and consistency in large deformable models for image warping. *Journal of Mathematical Imaging and Vision*, 34(3):270–290, 2009.
- [BGV09] J. Bigot, F. Gamboa, and M. Vimond. Estimation of translation, rotation and scaling between noisy images using the fourier mellin transform. *SIAM Journal on Imaging Sciences*, 2(2):614–645, 2009.
- [BHO07] Matias Bossa, Monica Hernandez, and Salvador Olmos. Contributions to 3d diffeomorphic atlas estimation: Application to brain images. In Nicholas Ayache, Sbastien Ourselin, and Anthony Maeder, editors, *MICCAI (1)*, volume 4791 of *Lecture Notes in Computer Science*, pages 667–674. Springer, 2007.
- [BM98] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.
- [BMTY05] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision*, 61(2):139–157, 2005.
- [Bos98] D. Bosq. *Nonparametric statistics for stochastic processes*, volume 110 of *Lecture Notes in Statistics*. Springer-Verlag, New York, second edition, 1998. Estimation and prediction.
- [Bos00] D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000. Theory and applications.
- [BZO08] M. Bossa, E. Zacur, and S. Olmos. Algorithms for computing the group exponential of diffeomorphisms: Performance evaluation. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–8, june 2008.
- [DK70] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.
- [DPR82] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.*, 12(1):136–154, 1982.
- [FLPJ04] P. T. Fletcher, C. Lu, Stephen M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, August 2004.

- [GM07] U. Grenander and M. Miller. *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, Oxford, 2007.
- [HBO07] Monica Hernandez, Matias Bossa, and Salvador Olmos. Estimation of statistical atlases using groups of diffeomorphisms. *Technical Report*, 2007.
- [HBO09] M. Hernandez, M. Bossa, and S. Olmos. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision*, 85(3):291–306, 2009.
- [HM06] Eldad Haber and Jan Modersitzki. Image registration with guaranteed displacement regularity. *Int. J. on Comp. Vision*, 2006.
- [JDJG04] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage*, 23:151–160, 2004.
- [LBBH98] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [Pen06] X. Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision*, 25(1):127–154, 2006.
- [RFS03] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3d statistical deformation models using non-rigid registration. *IEEE Transactions on Medical Imaging*, 22:77–84, 2003.
- [RS05] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.
- [Sdi08] M. Sdika. A fast nonrigid image registration with constraints on the jacobian using large scale constrained optimization. *Medical Imaging, IEEE Transactions on*, 27(2):271–281, feb. 2008.
- [Sil96] B. W. Silverman. Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, 24(1):1–24, 1996.
- [TY05a] A. Trouvé and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59, 2005.
- [TY05b] A. Trouvé and L. Younes. Metamorphoses through Lie group action. *Found. Comput. Math.*, 5(2):173–198, 2005.
- [UAE93a] M. Unser, A. Aldroubi, and M. Eden. B-Spline signal processing: Part I—Theory. *IEEE Transactions on Signal Processing*, 41(2):821–833, February 1993. IEEE Signal Processing Society’s 1995 best paper award.

- [UAE93b] M. Unser, A. Aldroubi, and M. Eden. B-Spline signal processing: Part II—Efficient design and applications. *IEEE Transactions on Signal Processing*, 41(2):834–848, February 1993.
- [WBR<sup>+</sup>07] L. Wang, F. Beg, T. Ratnanather, C. Ceritoglu, L. Younes, J.C. Morris, J.G. Csernansky, and M.I. Miller. Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type. *Medical Imaging, IEEE Transactions on*, 26(4):462–470, 2007.
- [You10] L. Younes. *Shapes and diffeomorphisms*, volume 171 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 2010.