



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <http://oatao.univ-toulouse.fr/>
Eprints ID: 10949

To link to this article: DOI: 10.1137/120864556

URL: <http://dx.doi.org/10.1137/120864556>

To cite this version: Bigot, Jérémie and Gouet, Raúl and López, Alfredo *Geometric PCA of Images*. (2013) SIAM Journal on Imaging Sciences, vol. 6 (n° 4). pp. 1851-1879. ISSN 1936-4954

Geometric PCA of Images*

J. Bigot[†], R. Gouet[‡], and A. López[§]

Abstract. We describe a method for analyzing the principal modes of geometric variability of images. For this purpose, we propose a general framework based on the use of deformation operators for modeling the geometric variability of images around a reference mean pattern. In this setting, we describe a simple algorithm for estimating the geometric variability of a set of images. Some numerical experiments on real data are proposed for highlighting the benefits of this approach. The consistency of this procedure is also analyzed in statistical deformable models.

Key words. geometric variability, principal component analysis, mean pattern estimation, Fréchet mean, image registration, deformable models, consistent estimation

AMS subject classifications. Primary, 62H25; Secondary, 62H35

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, in biomedical imaging, or in signal processing for neurosciences), which can be considered as square-integrable functions on a domain Ω , 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 are of fundamental interest. Principal component analysis (PCA) is a widely used method for estimating 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 which cannot be recovered by standard PCA.

The goal of this paper is to provide a general framework for analyzing the geometric variability of images through the use of deformation operators that can be parametrized by elements in a Hilbert space. This setting leads to a simple algorithm for estimating the main

[†]Institut de Mathématiques de Toulouse et CNRS (UMR 5219), Université de Toulouse, Toulouse 31062, France (jeremie.bigot@isae.fr). This author's research was supported by the Center for Mathematical Modeling and the CNRS in Santiago.

[‡]Departamento de Ingeniería Matemática, Center for Mathematical Modeling (CNRS UMI 2807), Universidad de Chile, Santiago 8370459, Chile (rgouet@dim.uchile.cl). This author's research was supported by projects PFB-03-CMM and Fondecyt 1120408.

[§]Center for Mathematical Modeling (CNRS UMI 2807), Universidad de Chile, Santiago 8370459, Chile (lopezalfa@gmail.com).

modes of geometric variability of images, and we prove the consistency of this approach in statistical deformable models.

1.1. PCA in a Hilbert space. First, let us introduce some tools and notation for performing standard PCA in a Hilbert space that will be used throughout the paper. Let \mathcal{H} be 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

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

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

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

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 λ th mode of variation of the random variable Z is defined as the function $\mathbb{E}Z + \rho\sqrt{\gamma_\lambda}\phi_\lambda$, for $\lambda \in \Lambda$, where ρ is a real playing the role of a weight parameter (typically one takes ρ between -3 and 3 , since the interval $[-3, 3]$ corresponds to the practical range of a standard normal variable). 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

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

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 λ 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, 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 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., [28] for an introduction to functional data analysis), and various authors (see, e.g., [18, 31] and references therein) have studied the consistency of empirical PCA in Hilbert spaces.

Empirical PCA, with $\mathcal{H} = L^2(\Omega)$, applied to a set of n images y_1, \dots, y_n is a method for computing 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 geometric variability; see the example of images of handwritten digits displayed in Figure 1. 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 standard PCA does not meaningfully reflect the modes of variability of the data; see Figure 1(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.

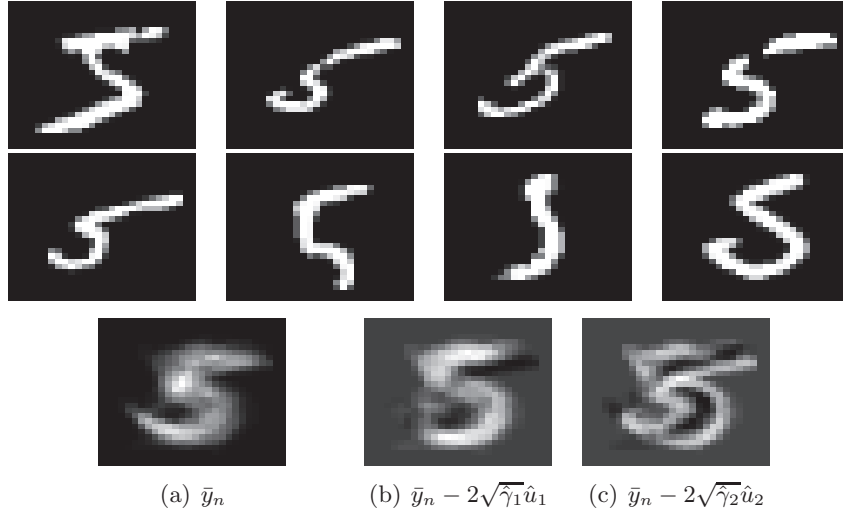


Figure 1. Digit 5. First two rows: a sample of 8 images out of $n = 30$ taken from the Mnist data base [26]. Last row: standard Euclidean mean, first and second empirical modes of variation.

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 (such as those in Figure 1) typically cannot be considered as a linear subspace 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 Figure 1 have mainly a geometric variability in space, which is much more important than the photometric variability.

1.2. A brief overview of PCA-like methods for analyzing geometric variability. A standard approach for analyzing geometric variability is to use registration. This well-known approach consists in computing geometric transforms of a set of images y_1, \dots, y_n , so that they can be compared. Then the main idea for estimating the geometric variability of such data is to apply classical PCA to the resulting transformation parameters after registration

and not to the images themselves. This approach is at the core of several methods for estimating the geometrical variability of images that we choose to call *geometric PCA methods* in what follows.

In [29], a linear and finite-dimensional space of nonrigid transformations is considered as the admissible space of deformations onto which standard PCA is carried out. This is the so-called statistical deformation model that is inspired by Cootes active shape models. An important limitation of this approach is the lack of invertibility of the deformations in such models. In several cases, the invertibility is a desirable property from the point of view of physical (for instance, when analyzing geometric variability of a determined organ) and mathematical modeling. Within the context of linear space of deformations, the noninvertibility issue had been addressed by enforcing the positiveness of the Jacobian determinant [22, 30]. However, such methods are unsuited for further statistical analysis, as statistical procedures on resulting transformation parameters (such as the empirical Euclidean mean) do 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 geometric transformations between images in the context of Grenander’s pattern theory [5, 24, 4]. In this framework, the set of admissible diffeomorphic transformations is considered as a Riemannian manifold, and thus first- and second-order statistical analysis on manifolds [20, 27] can be applied to perform statistical analysis of diffeomorphic deformations [3, 15, 23, 36]. Such diffeomorphisms are constructed as solution of an ordinary differential equation (ODE) governed by a time-dependent vector field belonging to a linear space; see, e.g., [5]. In this way, it is possible to build sets of diffeomorphisms that have mathematical properties very similar to Lie groups. This approach, which leads to the representation of the geometric variability of images through the use of standard PCA on the elements in the “Lie algebra” of vector fields, is discussed in detail in [32, 33, 37]. In particular, 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 key property, called momentum conservation, allows one to carry out PCA on the Hilbert space of initial momentums; see, e.g., [36].

A particular subclass of diffeomorphic deformations is the set of diffeomorphisms generated by an ODE governed by stationary vector fields. In this way, diffeomorphisms are directly characterized by vector fields belonging to a Hilbert space, and thus a standard statistical analysis such as PCA can be carried out on the vector fields computed after image registration. Compared to the case of diffeomorphisms generated by nonstationary vector fields, the resulting deformations do not have the same desirable properties in terms of group structure. Nevertheless, the natural parametrization of these deformations by a linear space makes them well suited for the purpose of geometric PCA. Moreover, the computational cost of the registration step when using stationary vector fields is considerably smaller, while keeping comparable accuracy according to the experimental results reported in [4, 24]. The properties of diffeomorphisms generated by stationary vector fields also allow simple and fast computations of the diffeomorphism associated with a vector field and vice versa. Hence, PCA methods for manifolds can be implemented to analyze geometric variability of diffeomorphic deformations generated by stationary vector fields; see [3, 15, 23].

1.3. Main contributions and organization of this paper. In section 2, we propose analyzing geometric PCA methods using a general framework where the spatial deformation operators (to represent geometric variability) are invertible and can be parametrized by elements of a Hilbert space. For estimating geometric variability, as is done in geometric PCA methods, we use a preliminary registration step. Then we apply classical PCA on Hilbert spaces to the resulting parameters representing the deformations after registration. The main contributions of this paper are then the following ones. First, for the application considered in this paper and for algorithmic purposes, we use diffeomorphic deformations parametrized by stationary vector fields that are expanded into a finite-dimensional basis of a linear space of functions. We show that such deformations are well suited for the analysis of handwritten digits. In this setting, an important and new contribution is to provide an automatic method for choosing the regularization parameter that represents the usual balance between the regularity of the spatial deformations and the quality of image alignment. Second, in section 3, we consider the problem of building geometric PCA methods that are consistent. To the best of our knowledge, this issue has not been studied very much in the literature on geometric PCA. We discuss the appropriate asymptotic setting for such an analysis, and we prove the consistency of our approach in statistical deformable models. We conclude the paper in section 4 by a short discussion. All proofs are gathered in two appendices at the end of the paper.

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 Ω . 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 the statistical aspects of an inferential procedure, as it avoids the treatment of the bias introduced by any discretization of the domain Ω . We refer the reader to section 3 for a detailed discussion on this point.

2.1. Grenander’s pattern theory of deformable templates. Following the ideas of Grenander’s pattern theory (see [21] 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 Ω . Recently, there has been a growing interest in Lie groups of transformations for modeling the geometric variability of images (see, e.g., [5, 32, 33, 37] and references therein), and applications are numerous in particular in biomedical imaging; see, e.g., [20, 25].

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

Definition 2.1. *Let \mathcal{V} be a Hilbert space. A deformation operator parametrized 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 Ω . Moreover, $\varphi(0, \cdot)$ is the identity on Ω 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 nonrigid deformations generated by stationary vector fields.

Translations. Let $\Omega = [0, 1]^d$ for some integer $d \geq 1$ and $\mathcal{V} = \mathbb{R}^d$. Also, let

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

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 Ω 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 $\mathcal{V} = \mathbb{R} \times \mathbb{R}^2$. Also, let

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

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_α is the rotation matrix of angle α and $b \in \mathbb{R}^2$ defines a translation. Observe that $\varphi((0, 0), \cdot)$ is the identity in Ω 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 let \mathcal{V} be a separable Hilbert space of smooth vector fields such that \mathcal{V} is continuously embedded in the set of functions $v : \Omega \rightarrow \mathbb{R}^d$ which are continuously differentiable and such that v and its derivatives vanish at the boundary of Ω . For $x \in \Omega$ and $v \in \mathcal{V}$, define $\varphi(v, x)$ as the solution at time $t = 1$ of the ODE

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

with initial condition $\phi_0 = x \in \Omega$. It is well known (see, e.g., [37]) that, for any $v \in \mathcal{V}$, the function $x \mapsto \varphi(v, x)$ is a C^1 diffeomorphism on Ω . 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 consists of geometric transforms of a set of images $y_1, \dots, y_n \in L^2(\Omega)$, so that they can 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 φ be a deformation operator, as described in Definition 2.1, and let $\mathbf{v} = (v_1, \dots, v_n) \in \mathcal{V} := \mathcal{V}^n$ and $\mathbf{y} = (y_1, \dots, y_n)$, with $y_i \in L^2(\Omega)$, $i = 1, \dots, n$.*

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

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

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

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

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

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

where $\mathcal{V}_\mu := \{v \in \mathcal{V} : \|v\| \leq \mu\}$ for some regularization parameter $\mu \geq 0$. Note that imposing the constraint $\|v_i\| \leq \mu$ allows one 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. The choice of μ is obviously of primary importance. A data-based procedure for its calibration is thus discussed in detail in section 2.3.4.

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 \mathbf{v} , are

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

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

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

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

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

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

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

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 φ 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 λ 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

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

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 λ 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 [7], 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 deformations 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 [26].

2.3.1. Specification of the Hilbert space of parameter \mathcal{V} . We choose \mathcal{V} as the vector space of functions from Ω to \mathbb{R}^2 , generated by a B-spline basis of functions, because they have good properties for approximating continuous functions and implementing efficient computations [34, 35]. Let $\{b_k : \Omega \rightarrow \mathbb{R}, k = 1, \dots, p\}$ denote a set of bidimensional tensor product B-splines, with knots defined on a regular grid of Ω , 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

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

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 [5] we take $L = \gamma I + \alpha \Delta$, where I is the identity operator, Δ is the Laplacian operator, and γ, α are positive scalars. By using the basic properties of differentiation and integration of B-splines [34], 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 Ω .

2.3.2. Minimization of the dissimilarity functional M^g . In the case of groupwise registration, we minimize the dissimilarity functional (2.4) over the set \mathcal{U}_0 defined in (2.5). 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 parametrize 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 ODE using a forward Euler integration scheme. For a comparison of different methods for solving such an ODE, we refer the reader to [16]. It can be shown (see Lemma 2.1 in [5]) that the gradient of φ with respect to the $\tilde{v}_k = (\tilde{v}_k^{(1)}, \tilde{v}_k^{(2)})$'s has a closed-form expression, and thus it can be explicitly computed to derive a gradient descent algorithm.

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

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

where Σ 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 whose i th column equals $\tilde{v}_i^t - \frac{1}{n} \sum_{j=1}^n \tilde{v}_j^t$. The matrix Σ is symmetric and hence admits a diagonalization $\Sigma = P \Lambda P$ with a Λ 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 a D diagonal matrix and $W^t W = W W^t = I$. We obtain the following spectral decomposition of \tilde{K}_n with 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 λ 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 λ th empirical mode of geometric variation according to Definition 2.3.

2.3.4. Choice of the regularization parameter μ and application of geometric PCA to handwritten digit data. We now describe the application of geometric PCA to handwritten digits, taken from the Mnist data base [26], based on the numerical framework we have described so far. We also discuss the problem of automatically selecting the regularization parameter μ , and we finally illustrate the benefits of geometric PCA over standard PCA.

For the B-spline base of \mathcal{V} , we choose a B-spline degree equal to 3, as it provides a good trade-off between smoothness and the size of the support. The number of B-spline knots is $p = 81$ arranged in a 9×9 regular grid. Such a value of p provides a fine B-spline grid with respect to an image size of 28×28 . For defining the differential operator L , we take $\gamma = 100$ and $\alpha = 1$. Note that $\gamma \gg \alpha$ in order to compensate for the scaling factor associated with the interknot spacing.

For each available digit (from 0 to 9) we determine the regularization parameter μ experimentally, by trying to obtain a good balance between the regularity of the vector fields and the matching of the images during the preliminary registration step. Our approach is inspired by the classical L-curve method in inverse problems. More precisely, for each digit, we took $n = 30$ images and we carried out registration 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. For each digit and for a given $\mu > 0$, we define $r(\mu)$ as the relative percentage value between

- the dissimilarity $M^g(\hat{\mathbf{v}}_\mu, \mathbf{y})$ of the images after registration with regularization parameter μ (see (2.6))

and

- the dissimilarity $M^g(\hat{\mathbf{v}}_0, \mathbf{y})$ of the images before registration (i.e., with regularization parameter $\mu = 0$),

that is,

$$r(\mu) = 100 * \frac{M^g(\hat{\mathbf{v}}_\mu, \mathbf{y})}{M^g(\hat{\mathbf{v}}_0, \mathbf{y})}.$$

We define also the finite difference derivative $\Delta_h r(\mu) = -(r(\mu + h) - r(\mu))/h$ for $h > 0$. For $h = 2$ and for all digits, we observed that the curves $\mu \rightarrow r(\mu)$ (with $\mu = 0, 0 + h, \dots, 30$) have an approximate convex shape and that the curves $\mu \rightarrow \Delta_h r(\mu)$, with $\mu = 0, 0 + h, \dots, 28$, have a decreasing trend to 0. We display the curves $\mu \rightarrow r(\mu)$ in Figure 2 for digits 0 and 1. It is reasonable to say that taking μ such that $\Delta_h r(\mu)$ is large corresponds to a situation of underfitting, whereas the case such that $\Delta_h r(\mu - h)$ is small corresponds to a situation of overfitting.

Hence, an automatic choice for the regularization parameter is to take

$$\mu^* = \max\{0, 0 + h, \dots, 28 : \Delta r(\mu) > t\},$$

where $0 \leq t \leq 100/h$ is a threshold value. One thus has the following interpretation: for μ larger than the selected value μ^* the rate of decrease of $r(\mu)$ is less than $t\%$. By setting $t = 2$, we have obtained $\mu^* = 12, 16, 14, 14, 18, 18, 12, 12, 10, 10$ for digits 0, 1, \dots , 9, respectively. We observed that $r(\mu^*)$ ranges from 0.1 to 0.3 among all digits; that is, the dissimilarity between the images after registration using the regularization parameter μ^* corresponds to 10%–30% of the dissimilarity between the images before registration.

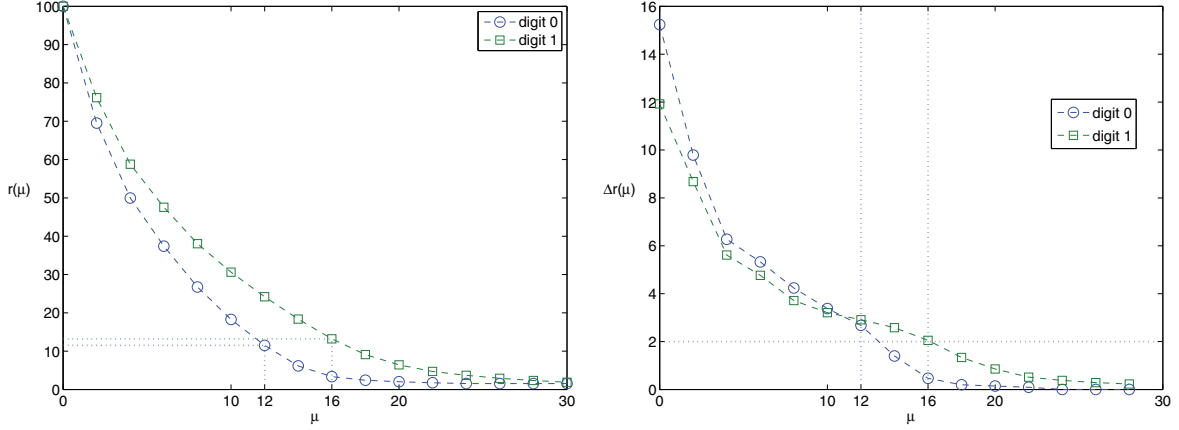


Figure 2. Choice of the regularization parameter μ^* for digits 0 and 1 through the analysis of the curves $\mu \rightarrow r(\mu)$ (figure on left-hand side) and $\mu \rightarrow \Delta r(\mu)$ (figure on right-hand side) with threshold $t = 2\%$.

For each digit, we carried out geometric PCA with a preliminary registration step as described in the previous paragraph and with regularization parameter μ^* . To illustrate the advantages of our procedure, we have also carried out standard PCA of each digit, which amounts to analyzing the photometric variability of the data. Thus, we have computed

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

the λ th standard empirical mode of photometric variation of the data as described in section 1.1. In Figure 3, we show the geometric modes of variations by displaying the images

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

with $\lambda = 1, 2$ and $\rho = 2, -2$. Results using standard PCA are also displayed in Figure 3. We observe that geometric PCA better reflects the main modes of variability of the digits. To the contrary, standard PCA fails in several cases in representing geometric variability of some digits, and it sometimes results 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 .

We also use the learned Fréchet mean \hat{f} and the learned empirical eigenvalues $(\hat{\kappa}_\lambda)_{\lambda \in \Lambda}$ and eigenvectors $(\hat{\phi}_\lambda)_{\lambda \in \Lambda}$ to produce simulated images consisting of a random warp of the Fréchet mean. More precisely, we generate a new image $Y \in L^2(\Omega)$ from the model

$$(2.12) \quad Y(x) = \hat{f}(\varphi^{-1}(V, x)), \quad x \in \Omega,$$

where $V = \bar{v}_n + \sum_{\lambda=1}^q \rho_\lambda \sqrt{\hat{\kappa}_\lambda} \hat{\phi}_\lambda$, $q \leq 2p$, and ρ_λ , $\lambda = 1, \dots, q$, are independent standard normal random variables. Observe that $\mathbb{E}(V) = \bar{v}_n$ and that the covariance operator \tilde{K}_n of V is the projection of \hat{K}_n onto the space generated by ϕ_1, \dots, ϕ_q . In Figure 4, we display, for each digit, five independent random images obtained from (2.12) with $q = 8$. For such a choice of q we have that the ratio between the trace of \tilde{K}_n and the trace of \hat{K}_n , that is, $\sum_{\lambda=1}^{2p} \hat{\kappa}_\lambda / \sum_{\lambda=1}^q \hat{\kappa}_\lambda$, ranges from 0.75 to 0.9 among digits 0, \dots , 9.

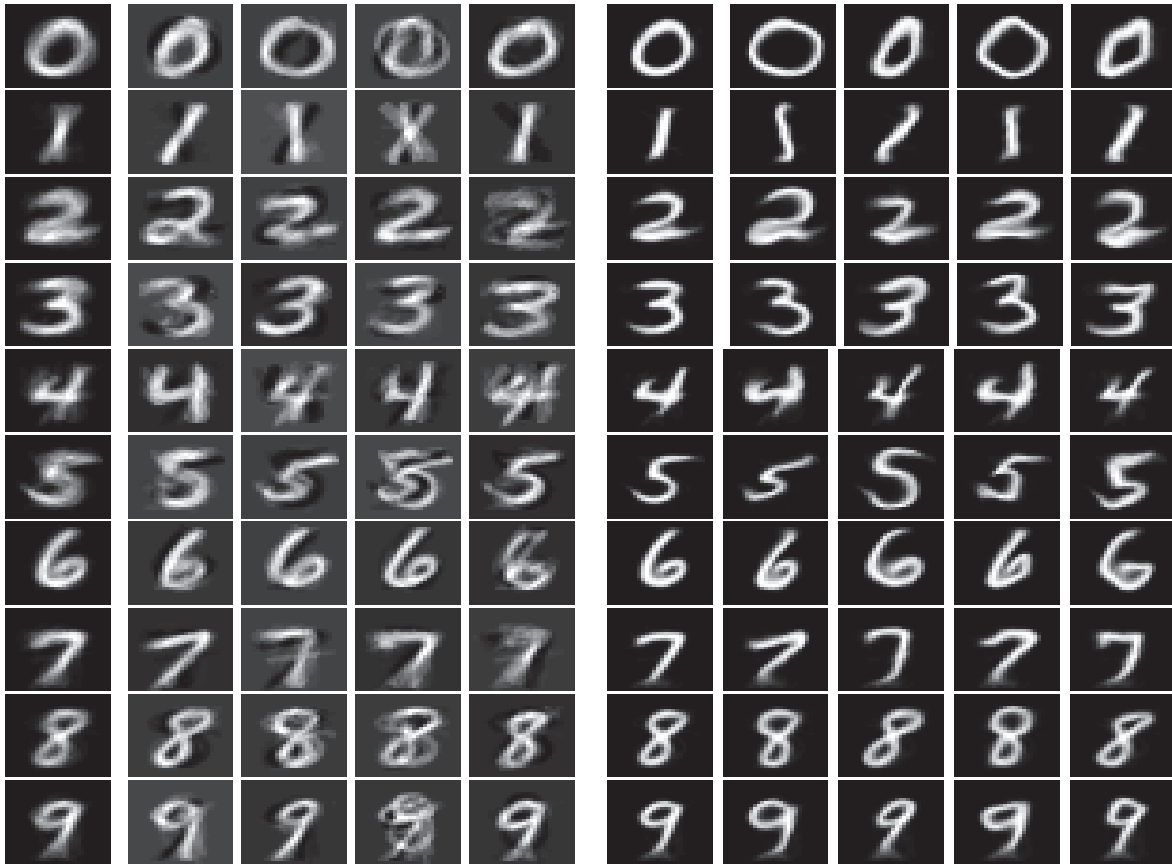


Figure 3. Visualization of standard PCA (block of five images in the left-hand side) and geometric PCA (block of five images in the right-hand side) for digits 0 to 9. For each digit, from left to right, the images correspond to \bar{y} , $\bar{y}_n - 2\sqrt{\hat{\gamma}_1}\hat{u}_1$, $\bar{y}_n + 2\sqrt{\hat{\gamma}_1}\hat{u}_1$, $\bar{y}_n - 2\sqrt{\hat{\gamma}_2}\hat{u}_2$, and $\bar{y}_n + 2\sqrt{\hat{\gamma}_2}\hat{u}_2$ and then 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 do not recover the shape of the digits well and that they produce a blurring of the images. In contrast, geometric PCA results recover the geometric features of the digits.

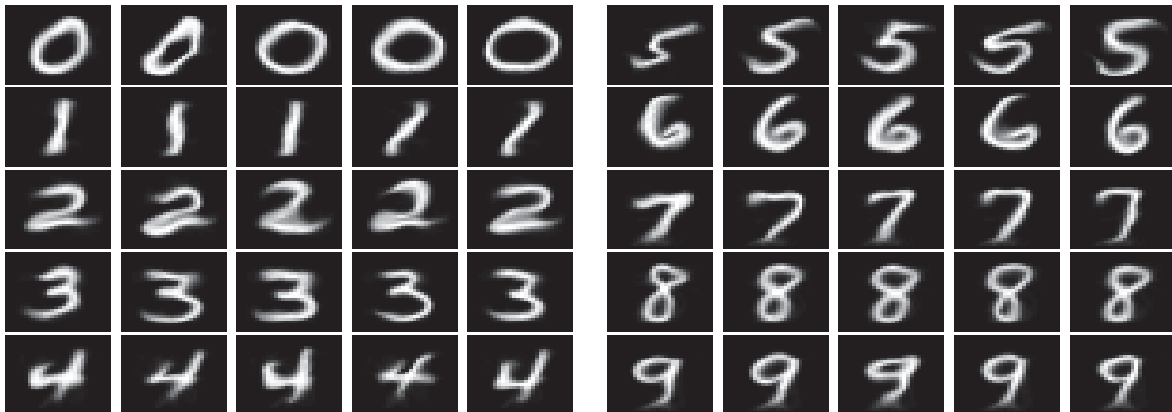


Figure 4. Simulated images for each digit $0, \dots, 9$ from model (2.12) based on the learning of the Fréchet mean \hat{f} and the eigenvalues/eigenvectors of the empirical covariance operator \hat{K}_n .

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 a mean pattern from a set of images [1, 2, 7, 10, 9] 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:

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

where

- f^* is an unknown mean pattern belonging to $L^2(\Omega)$,
- φ is a deformation operator associated with 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 $\mathbb{E}\|W\|_2^2 = 1$, and
- (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

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

have algebraic multiplicity 1, i.e., $\kappa_1 > \kappa_2 > \dots \geq 0$. This implies that the λ th eigengap, 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 Ω and thus correspond to a source of geometric variability in the data.

Model (3.1) is somewhat ideal, since images are never observed in a continuous domain but rather on a discrete set of pixels. A detailed discussion on this issue can be found in [1, 2], where it is proposed to deform a template model and not the observed discrete images themselves for the purpose of template estimation. However, to study the asymptotic properties of a statistical procedure, it is simpler to assume that the data are random elements of $L^2(\Omega)$ to avoid the treatment of the bias introduced by any discretization scheme.

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 ϕ_1, ϕ_2, \dots . For $\lambda \in \Lambda$, the λ 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 λ th empirical mode of geometric variation $\hat{\psi}_\lambda$ (see (2.9)) tends to the λ th population mode of geometric variation ψ_λ , 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

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

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 independent and identically distributed (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 it corresponds 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 ϵW_i in model (3.1). In the statistical literature (see, e.g., [17]), it has been shown that the setting $\epsilon \rightarrow 0$ in a white noise model such as (3.1) corresponds to the setting where a number $N \sim \epsilon^{-2}$ of pixels would tend to infinity in a related model of images sampled on a discrete grid of size N . Therefore, one may interpret $\epsilon \rightarrow 0$ as the asymptotic setting where one observes images with a growing number $N \sim \epsilon^{-2}$ of pixels.

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 ϵW_i , $i = 1, \dots, n$, in model (3.1) have to be sufficiently small in order to recover the geometric modes of variation. One may argue that the main interest for practical purposes is the asymptotic setting where $n \rightarrow +\infty$ and ϵ is fixed. However, recent results show that, in the setting where ϵ is fixed, it is not possible to recover the random variables V_i encoding the deformations in model (3.1) by any statistical procedure; see, e.g., [7, 8]. In the double asymptotic setting $n \rightarrow +\infty$ and $\epsilon \rightarrow 0$, a detailed analysis of the problem of recovering the template f^* in model (3.1) has been carried out in [11]. In particular, some answers are given in [11] on the relative rate between n and ϵ that is needed to guarantee a consistent estimation of f^* via the use of the Fréchet mean. However, it is out of the scope of this paper to discuss such issues for the problem of consistent estimation of the main modes of geometric variations.

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

(i)

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

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}_μ to $C(\Omega, \Omega)$ is continuous,

where $C(\Omega, \Omega)$ is the space of continuous functions from Ω to Ω , 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 variables.

Finally, before stating our consistency results, we define convergence in probability in the double asymptotic setting $n \rightarrow \infty$, $\epsilon \rightarrow 0$. Let $X_{n,\epsilon}, X_n, X_\epsilon, X$, $n = 1, 2, \dots, \epsilon > 0$, be random variables with values on a metric space (S, d) . The notation $\text{plim}_\epsilon X_{n,\epsilon} = X_n$ stands for $d(X_{n,\epsilon}, X_n) \rightarrow 0$ in probability as $\epsilon \rightarrow 0$; $\text{plim}_n X_{n,\epsilon} = X_\epsilon$ 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 3.3. *Let $\mathbf{Y} = (Y_1, \dots, Y_n)$ be i.i.d. observations of model (3.1) with deformation operator φ and regularization parameter μ . 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 φ is μ -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 $\text{plim}_{n,\epsilon} \sin^2(\hat{\phi}_\lambda, \phi_\lambda) = 0$ for all $\lambda \in \Lambda$.*

The injectivity condition of φ^* in the previous theorem implies that if there were no additive noise in model (3.1), then the registration of the observations \mathbf{Y} onto the template f^* would lead exactly to the nonobserved deformation parameters V_1, \dots, V_n . In other words, if $\epsilon = 0$ in (3.1), then the template dissimilarity functional $M^t(\mathbf{v}, \mathbf{Y}, f^*)$ (see (2.3)) has a unique minimizer over \mathcal{V}_μ given by $\mathbf{v} = (V_1, \dots, V_n)$. The condition of continuity of φ^{*-1} ensures that the registration problem with noise level ϵ will converge to the registration problem with no noise as $\epsilon \rightarrow 0$. In section 3.3, we analyze the case where φ is the translator operator defined in section 2.1 and we provide some conditions on f^* and φ ensuring that the hypotheses of Theorem 3.3 are satisfied. In the case where φ is the diffeomorphic deformation operator defined in section 2.1, it is necessary to impose much stronger assumptions on the template f^* and the space of vector fields \mathcal{V} to ensure that $\varphi^* : \mathcal{V}_\mu \rightarrow \varphi^*(\mathcal{V}_\mu)$ has a continuous inverse. In the concluding section, section 4, we further discuss this issue.

Remark 3.2. Observe that, under the hypotheses of Theorem 3.3, the λ th empirical mode of geometric variation $\hat{\psi}_\lambda$ converges the population mode of geometric variation ψ_λ 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 μ -regularity of φ .

We show below how a stronger regularity assumption on φ 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 ϵ .

Theorem 3.4. *Under the hypotheses of Theorem 3.3 and if φ^{*-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 μ), 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 μ , $h(u, n, \epsilon) = \epsilon^2 \left(1 + 2\frac{u}{n} + 2\sqrt{\frac{u}{n}}\right)$, and $g(u, n) = \left(\frac{u}{n} + \sqrt{\frac{u^2}{n^2} + \frac{u}{n}}\right)^2$.

Now, take $u^* > 0$ 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

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

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 3.5 (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 3.6. Let $\mathbf{Y} = (Y_1, \dots, Y_n)$ be i.i.d. observations of model (3.1) with deformation operator φ and regularization parameter μ . 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 φ is μ -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 hypotheses of Theorem 3.6, $\hat{\psi}_\lambda$ converges to ψ_λ in probability as $n \rightarrow +\infty$ and $\epsilon \rightarrow 0$.

3.3. Translation operators. We study the applicability of Theorems 3.3, 3.4, and 3.6 to translation operator φ 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 $x = (x_1, \dots, x_d) \in \Omega$. Also, let $\mathcal{V} = \mathbb{R}^d$ be equipped with the usual Euclidean inner product.

Now, let us show that φ is a deformation operator in the sense of Definition 2.1. It holds that $\varphi(0, \cdot)$ is the identity in Ω 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 φ is μ -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 φ^{*-1} , defined in Theorem 3.3, is Lipschitz. Let θ_k , $k = 1, \dots, d$, be the low frequency Fourier coefficients of the template f^* , that is,

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

Lemma 3.7. *Suppose that f^* is such that $\theta_k \neq 0$ for all $1 \leq k \leq d$, and let $\mu < 1/2$; then φ^* is injective and φ^{*-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 3.3 and 3.4 are verified. Thus geometric PCA is consistent, in the case of template registration, with translation operators. Observe that the hypotheses of Lemma 3.7 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 3.6 applies. The set of deformation 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 (2.5). We have the following.

Proposition 3.8. *Suppose that $\theta_k \neq 0$ 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*

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

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 μ .

Remark that in Proposition 3.8, $D^g(\mathbf{u}^*(\mathbf{v}), \mathbf{v}) = 0$. 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 3.5 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., [14, Theorem 2.6]) we conclude that, for any $\eta > 0$,

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

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 operators.

4. Conclusion and discussion. The contribution of this paper is twofold. First, the use of deformation operators (as introduced in this paper) provides a general framework for modeling and analyzing geometric variability 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 [5], the link with our framework is not straightforward. Indeed, in this setting, there are two possibilities for defining the deformation operators. One can parametrize 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 points of view, 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 belonging to a finite-dimensional Hilbert space.

The second contribution of this paper is the study of the consistency of geometric PCA methods in statistical deformable models, which, to the best of our knowledge, has not been investigated so far. One can remark that our consistency results rely on strong assumptions on the template f^* and the deformation operator φ . For the case of translations, we have provided (see section 3.3) verifiable conditions to satisfy such assumptions. A similar analysis, in the case of diffeomorphic deformations, is much more complex. For the case of template registration, one of our main assumptions is that 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. Such a condition together with some regularity conditions on φ ensures that if $\mathbf{Y} = (Y_1, \dots, Y_n)$ is sampled from model (3.1) with noise level $\epsilon = 0$, then the registration problem $\min_{\mathbf{v} \in \mathcal{V}_\mu} M^t(\mathbf{v}, \mathbf{Y}, f^*)$ has a unique solution given by the nonobserved deformation parameters $\mathbf{V} = (V_1, \dots, V_n)$, where M^t is the template dissimilarity functional defined in (2.3). In particular, in the case where φ is a diffeomorphic deformation operator, this injectivity condition necessarily requires the template f^* to not be constant in any open region of Ω , which is a quite restrictive assumption that does not hold in applications. One possibility for removing this injectivity condition could be to try to estimate the deformation parameters of minimal norm among those that left the template f^* unchanged, that is,

$$(4.1) \quad \|V_i\| = \min\{\|v\| : v \in \mathcal{V}, \varphi^*(v) = \varphi^*(V_i)\} \text{ a.s.}$$

for all $i = 1, \dots, n$. However, such an approach makes the analysis on the consistency of our procedure much more difficult. Nevertheless, we hope that the methods presented in this paper will stimulate further investigation into the development of consistent statistical procedures for the analysis of geometric variability.

Appendix A. Preliminary technical results. In this appendix, we give a deviation in probability of $\sup_{\mathbf{v} \in \mathcal{V}_\mu} |D^g(\mathbf{v}) - M^g(\mathbf{v}, \mathbf{Y})|$ 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 μ -regular deformation operator φ . Let*

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

Then, for any $s > 0$,

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

where A_μ 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 $\mathbf{v} \in \mathcal{V}_\mu$,

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

Let $g \in L^2(\Omega)$, and let $K^W g(x) = \int_{\Omega} k(x, y)g(y)dy$ 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 , that $\mathbb{E}\|W_i\|_2^2 = \sum_{\lambda=1}^{+\infty} w_\lambda = 1 < +\infty$, 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 χ^2 random variable, with one degree of freedom, is $\mathbb{E}(e^{s\xi_{i,\lambda}^2}) = (1 - 2s)^{-1/2}$ for $s > 0$, it follows that

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

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 [12], the above inequality implies that, for any $s > 0$,

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

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 $w_1 \leq \mathbb{E}\|W\|_2^2 = 1$. \blacksquare

Lemma A.2. Consider model (3.1) with μ -regular deformation operator φ . Let

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

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 μ , 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

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

where Q is defined in (A.1).

By applying the Cauchy–Schwarz 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 μ -regularity of φ 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. \blacksquare

Lemma A.3. Consider model (3.1) with μ -regular deformation operator φ . 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 μ , 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

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

From the Cauchy–Schwarz 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 μ -regularity of φ , 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. \blacksquare

Remark A.1. Observe that Lemmas A.2 and A.3 imply

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

and

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

The proofs of Theorems 3.3, 3.4, and 3.6 rely on the following two propositions that establish the consistency of the registration procedures described in section 2.2.

Proposition A.4. *Let $\hat{\mathbf{V}} \in \arg \min_{\mathbf{v} \in \mathcal{V}_\mu} M^t(\mathbf{v}, \mathbf{Y}, f^*)$ be the parameters obtained from template registration of \mathbf{Y} on f^* . Then,*

- (i) *under the hypotheses of Theorem 3.3, $\text{plim}_{n, \epsilon} \bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) = 0$, and*
- (ii) *under the hypotheses of Theorem 3.4,*

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

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

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

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

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,

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

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 φ^{*-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 2μ 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 the fact that φ^{*-1} is uniformly Lipschitz, with constant $L(f^*, \mu) > 0$, imply that

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

and the result follows from Lemma A.2. \blacksquare

Proposition A.5. *Let $\hat{\mathbf{V}} \in \arg \min_{\mathbf{v} \in \mathcal{U}} M^g(\mathbf{v}, \mathbf{Y})$ be the parameters obtained from groupwise registration of \mathbf{Y} . Then, under the hypotheses of Theorem 3.6, $\text{plim}_{n,\epsilon} \bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) = 0$.*

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 $\bar{d}^2(\hat{\mathbf{V}}, \mathbf{V}) \leq 2\bar{d}^2(\hat{\mathbf{V}}, \mathbf{u}^*) + 2\bar{d}^2(\mathbf{u}^*, \mathbf{V})$. \blacksquare

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.6. *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 us define $\epsilon_i = v_i - u_i$, so that 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 $\|\epsilon_i\| \leq \|u_i\| + \|v_i\| \leq 2r$,

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

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. \blacksquare

The following theorem follows from the theory developed in [6, 19].

Theorem A.7. *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 eigenvalue/eigenvector pairs $(\kappa_\lambda, \phi_\lambda)_{\lambda \geq 1}$ and $(\hat{\kappa}_\lambda, \hat{\phi}_\lambda)_{\lambda \geq 1}$, respectively. Then*

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

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

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

Appendix B. Proofs of main results.

B.1. Proof of Theorem 3.3.

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

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

The first term in the right-hand side of inequality (B.1) can be controlled by using Lemma A.6 and noting that $\|V_i\|, \|\hat{V}_i\| \leq \mu, i = 1, \dots, n$, that is,

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

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 inequality for Hilbert–Schmidt operators (see, e.g., [13, Chapter 3]), it follows that

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

for some constant $\tilde{C}(\mu) > 0$ depending only on μ . Hence, we can combine (B.1), (B.2), and (B.3) with Proposition A.4(i) to obtain that for any $\eta > 0$, $\lim_{n,\epsilon} \mathbb{P}(\|\hat{K}_n - K\|_{HS}^2 > \eta) = 0$, that is,

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

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 λ th empirical eigengap 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

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

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

B.2. Proof of Theorem 3.4.

Proof. Combining (B.1), (B.2), and (B.3) with Proposition A.4(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 μ 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

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

Now, take $s^* > 0$ 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 $0 < s \leq s^*$,

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

B.3. Proof of Theorem 3.6.

Proof. We proceed similarly as in the proof of Theorem 3.3. In the case of groupwise registration, inequalities (B.1), (B.2), and (B.3) are still valid and can be combined with Proposition A.5 to obtain $\text{plim}_{n, \epsilon} \hat{K}_n = K$. The rest of proof is identical to that of Theorem 3.3. \blacksquare

Appendix C. Technical results for translation operators.

Lemma C.1. *Let φ 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 $v \in \mathcal{V}$, and
- (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 $a \geq b$,

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

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 $d_\Omega \leq \frac{1}{2}$, and hence $d_\Omega(\varphi(u, x), \varphi(v, x)) \leq |u - v|$ if $|u - v| \geq 1$. On the other hand, 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 $d_C(\varphi(u, \cdot), \varphi(v, \cdot)) \leq |u - v|$. ■

In order to prove Lemma 3.7 and Proposition 3.8, 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 θ_k is defined in (3.6).

C.1. Proof of Lemma 3.7.

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

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

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,

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

C.2. Proof of Proposition 3.8.

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

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

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

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

Further, remark that

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

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 hypotheses of the proposition, one has that $|(u_m^{(k)} - V_m^{(k)}) - (u_{m'}^{(k)} - V_{m'}^{(k)})| \leq 2(\mu + \rho) < 1/4$. Therefore, for $\alpha = 2(\mu + \rho)$, it follows that

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

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

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

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

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

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

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

with $C_0(f^*, \mu) = 2C(\alpha) \min_{1 \leq k \leq d} \{|\theta_{\ell^{(k)}}|^2\}$ 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. \blacksquare

Acknowledgments. We are very much indebted to the referees and the Associate Editor for their constructive comments and remarks that resulted in a significant improvement of the original manuscript.

REFERENCES

- [1] S. ALLASSONIÈRE, Y. AMIT, AND A. TROUVÉ, *Toward a coherent statistical framework for dense deformable template estimation*, J. R. Stat. Soc. Ser. B Stat. Methodol., 69 (2007), pp. 3–29.
- [2] S. ALLASSONIÈRE, E. KUHN, AND A. TROUVÉ, *Bayesian deformable models building via stochastic approximation algorithm: A convergence study*, Bernoulli, 16 (2010), pp. 641–678.
- [3] V. ARSIGNY, O. COMMOWICK, X. PENNEC, AND N. AYACHE, *A log-Euclidean framework for statistics on diffeomorphisms*, in Proceedings of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Part I, Lecture Notes in Comput. Sci. 4190, Springer-Verlag, Berlin, 2006, pp. 924–931.
- [4] J. ASHBURNER, *A fast diffeomorphic image registration algorithm*, NeuroImage, 38 (2007), pp. 95–113.
- [5] 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 (2005), pp. 139–157.
- [6] R. BHATIA, C. DAVIS, AND A. MCINTOSH, *Perturbation of spectral subspaces and solution of linear operator equations*, Linear Algebra Appl., 52/53 (1983), pp. 45–67.
- [7] J. BIGOT AND B. CHARLIER, *On the consistency of Fréchet means in deformable models for curve and image analysis*, Electron. J. Stat., 5 (2011), pp. 1054–1089.
- [8] J. BIGOT AND S. GADAT, *A deconvolution approach to estimation of a common shape in a shifted curves model*, Ann. Statist., 38 (2010), pp. 2422–2464.
- [9] J. BIGOT, S. GADAT, AND J. M. LOUBES, *Statistical M-estimation and consistency in large deformable models for image warping*, J. Math. Imaging Vision, 34 (2009), pp. 270–290.
- [10] J. BIGOT, F. GAMBOA, AND M. VIMOND, *Estimation of translation, rotation, and scaling between noisy images using the Fourier–Mellin transform*, SIAM J. Imaging Sci., 2 (2009), pp. 614–645.
- [11] J. BIGOT AND X. GENDRE, *Minimax properties of Fréchet means of discretely sampled curves*, Ann. Statist., 41 (2013), pp. 923–956.
- [12] L. BIRGÉ AND P. MASSART, *Minimum contrast estimators on sieves: Exponential bounds and rates of convergence*, Bernoulli, 4 (1998), pp. 329–375.
- [13] D. BOSQ, *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd ed., Lecture Notes in Statist. 110, Springer-Verlag, New York, 1998.
- [14] D. BOSQ, *Linear Processes in Function Spaces: Theory and Applications*, Lecture Notes in Statist. 149, Springer-Verlag, New York, 2000.
- [15] M. BOSSA, M. HERNANDEZ, AND S. OLMOS, *Contributions to 3d diffeomorphic atlas estimation: Application to brain images*, in Proceedings of the 10th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '07), Lecture Notes in Comput. Sci. 4791, N. Ayache, S. Ourselin, and A. Maeder, eds., Springer-Verlag, Berlin, 2007, pp. 667–674.
- [16] M. BOSSA, E. ZACUR, AND S. OLMOS, *Algorithms for computing the group exponential of diffeomorphisms: Performance evaluation*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '08), 2008, pp. 1–8.
- [17] L. D. BROWN AND M. G. LOW, *Asymptotic equivalence of nonparametric regression and white noise*, Ann. Statist., 24 (1996), pp. 2384–2398.
- [18] 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 (1982), pp. 136–154.
- [19] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation. III*, SIAM J. Numer. Anal., 7 (1970), pp. 1–46.
- [20] P. T. FLETCHER, C. LU, S. M. PIZER, AND S. JOSHI, *Principal geodesic analysis for the study of nonlinear statistics of shape*, IEEE Trans. Med. Imag., 23 (2004), pp. 995–1005.
- [21] U. GRENANDER AND M. MILLER, *Pattern Theory: From Representation to Inference*, Oxford University Press, Oxford, UK, 2007.

- [22] E. HABER AND J. MODERSITZKI, *Image registration with guaranteed displacement regularity*, Int. J. Comput. Vision, 71 (2007), pp. 361–372.
- [23] M. HERNANDEZ, M. BOSSA, AND S. OLMOS, *Estimation of Statistical Atlases Using Groups of Diffeomorphisms*, Technical report, Aragon Institute of Engineering Research (I3A), 2007; available online from <http://diec.unizar.es/intranet/articulos/uploads/I3ATR.pdf>.
- [24] M. HERNANDEZ, M. BOSSA, AND S. OLMOS, *Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows*, Int. J. Comput. Vision, 85 (2009), pp. 291–306.
- [25] S. JOSHI, B. DAVIS, M. JOMIER, AND G. GERIG, *Unbiased diffeomorphic atlas construction for computational anatomy*, NeuroImage, 23 (2004), pp. 151–160.
- [26] Y. LECUN, L. BOTTOU, Y. BENGIO, AND P. HAFFNER, *Gradient-based learning applied to document recognition*, Proc. IEEE, 86 (1998), pp. 2278–2324.
- [27] X. PENNEC, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, J. Math. Imaging Vision, 25 (2006), pp. 127–154.
- [28] J. O. RAMSAY AND B. W. SILVERMAN, *Functional Data Analysis*, 2nd ed., Springer Ser. Statist., Springer-Verlag, New York, 2005.
- [29] D. RUECKERT, A. F. FRANGI, AND J. A. SCHNABEL, *Automatic construction of 3d statistical deformation models using non-rigid registration*, IEEE Trans. Med. Imag., 22 (2003), pp. 77–84.
- [30] M. SDIKA, *A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization*, IEEE Trans. Med. Imag., 27 (2008), pp. 271–281.
- [31] B. W. SILVERMAN, *Smoothed functional principal components analysis by choice of norm*, Ann. Statist., 24 (1996), pp. 1–24.
- [32] A. TROUVÉ AND L. YOUNES, *Local geometry of deformable templates*, SIAM J. Math. Anal., 37 (2005), pp. 17–59.
- [33] A. TROUVÉ AND L. YOUNES, *Metamorphoses through Lie group action*, Found. Comput. Math., 5 (2005), pp. 173–198.
- [34] M. UNSER, A. ALDROUBI, AND M. EDEN, *B-Spline signal processing: Part I—Theory*, IEEE Trans. Signal Process., 41 (1993), pp. 821–833.
- [35] M. UNSER, A. ALDROUBI, AND M. EDEN, *B-Spline signal processing: Part II—Efficient design and applications*, IEEE Trans. Signal Process., 41 (1993), pp. 834–848.
- [36] L. WANG, F. BEG, T. RATNANATHER, C. CERITOGU, 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*, IEEE Trans. Med. Imag., 26 (2007), pp. 462–470.
- [37] L. YOUNES, *Shapes and Diffeomorphisms*, Appl. Math. Sci. 171, Springer-Verlag, Berlin, 2010.