



Probabilistic modeling of anatomical variability using a low dimensional parameterization of diffeomorphisms



Miaomiao Zhang^{a,*}, William M. Wells III^{a,b}, Polina Golland^a

^a Computer Science and Artificial Intelligence Laboratory, MIT, Massachusetts

^b Brigham and Women's Hospital, Harvard Medical School, Massachusetts

ARTICLE INFO

Article history:

Received 1 March 2017

Revised 19 June 2017

Accepted 28 June 2017

Available online 8 July 2017

Keywords:

Probabilistic modeling

Principal geodesic analysis

Diffeomorphic shape variability

Bandlimited space

ABSTRACT

We present an efficient probabilistic model of anatomical variability in a linear space of initial velocities of diffeomorphic transformations and demonstrate its benefits in clinical studies of brain anatomy. To overcome the computational challenges of the high dimensional deformation-based descriptors, we develop a latent variable model for principal geodesic analysis (PGA) based on a low dimensional shape descriptor that effectively captures the intrinsic variability in a population. We define a novel shape prior that explicitly represents principal modes as a multivariate complex Gaussian distribution on the initial velocities in a bandlimited space. We demonstrate the performance of our model on a set of 3D brain MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Our model yields a more compact representation of group variation at substantially lower computational cost than the state-of-the-art method such as tangent space PCA (TPCA) and probabilistic principal geodesic analysis (PPGA) that operate in the high dimensional image space.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The study of anatomical shape variability across populations and its relationship with disease processes plays an important role in medical image analysis. For example, identifying pathological brain shape changes caused by neurodegenerative disorders from brain MRI scans provides new insights into the nature of the disease and supports treatment (Gerig et al., 2001; Nemmi et al., 2015). Research in shape analysis mainly focuses on developing statistical models with well defined shape descriptors such as landmarks (Cootes et al., 1995; Bookstein, 1997), medial axes (Pizer et al., 1999), and deformation-based representations (Christensen et al., 1993). This paper focuses on a deformation-based shape descriptor with the underlying assumption that the geometric information in the deformations explicitly reflects the shape changes, i.e., shrinkage or expansion, of local structures. In many clinical applications, it is natural to require the deformation to be a diffeomorphism, which guarantees a differentiable bijective mapping with a differentiable inverse. A well developed framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) endowed with a distance metric in the space of diffeomorphisms was introduced by Beg et al. (2005) to estimate such deformations.

The deformable template approach, which is also known as atlas building, is commonly used for statistical shape analysis of diffeomorphic transformations (Joshi et al., 2004; Twining et al., 2005; Vialard et al., 2011). This class of methods employs image registration to match a template to each individual subject and then computes statistics of the resulting transformations. However, the high dimensional nature of the imaging data, for instance, a 128^3 or 256^3 image grid as a shape descriptor for a 3D brain MRI presents substantial challenges for model selection and uncertainty estimation if only a small number of image scans is available. Statistical inference in such a high dimensional space demands large computational resources and special programming techniques. Moreover, the optimization landscape contains numerous local minima. To address this problem, data dimensionality reduction methods that extract relevant latent structure from image transformations have been proposed in the diffeomorphic setting. Vaillant et al. (2004) performed principal component analysis (PCA) in the linearized tangent space of diffeomorphisms (TPCA) on the initial momenta, performing statistical modeling of transformations as a step that follows the estimation of deformations. Similar approaches based on the parameterization of stationary velocity fields (Sweet and Pennec, 2010) and free-form B-spline deformations (Onofrey et al., 2013) were also developed. Qiu et al. (2012) constructed an empirical shape distribution by using TPCA to estimate the intrinsic dimensionality of the diffeomorphic surface variation. A Bayesian model of shape

* Corresponding author.

E-mail addresses: miao86@mit.edu, miao86@csail.mit.edu (M. Zhang).

variability has been proposed to extract the principal modes after estimating a covariance matrix of transformations (Gori et al., 2013). A unified framework of principal geodesic analysis (PGA) was first developed by Fletcher et al. (2003) to infer the principal modes of variation simultaneously with the data fitting procedure. This method generalized PCA to finite-dimensional manifolds and estimated the geodesic subspaces by minimizing the sum-of-squared geodesic distances. Moreover, PGA enabled factor analysis of diffeomorphisms that treated data variability as a joint inference problem in a probabilistic principal geodesic analysis (PPGA) model (Zhang and Fletcher, 2014; 2015a). All prior models reviewed here were designed to find a compact low dimensional space to represent the data. However, their estimation still remains computationally expensive due to the fact that each operation has to be performed numerically on dense image grids in a high dimensional space.

In contrast, we propose to detect the latent subspaces of anatomical shape variability by using a low dimensional shape descriptor of diffeomorphisms via bandlimited initial velocity fields (Zhang and Fletcher, 2015b), in a model we call *low dimensional probabilistic principal geodesic analysis (LPPGA)*. More specifically, our contributions are as follows:

1. We define a low dimensional probabilistic framework of factor analysis in the context of diffeomorphic atlas building.
2. We dramatically reduce the computational cost of detecting principal geodesics of diffeomorphisms by employing a bandlimited parametrization in the Fourier space.
3. We enforce the orthogonality constraints on the principal modes, which is computationally intractable in high dimensional models like PPGA (Zhang and Fletcher, 2014).

This paper is an extension of a recently published conference paper (Zhang et al., 2016), with several additional developments. First, we provide in-depth derivations of the statistical model and inference procedure. Second, we include comprehensive experimental results that validate the method. Moreover, we demonstrate Markov Chain Monte Carlo sampling in the proposed shape space, which is computationally intractable on dense image grids. We report estimated principal modes in the ADNI brain MRI dataset (Jack et al., 2008) and compare them with the results of TPCA and PPGA of diffeomorphisms estimated on the full image grid. The experimental results show that the low dimensional statistics encode important features of the data, better capture the group variation and improve data interpretability. Moreover, our model requires substantially lower computational resources.

2. Background

In this section, we first briefly review the mathematical background of diffeomorphic atlas building in the LDDMM setting (Beg et al., 2005) with geodesic shooting (Younes et al., 2009; Vialard et al., 2012). We then provide a short summary of low dimensional Fourier representation that forms the basis of our method.

Let J_1, \dots, J_N be the N input images that are assumed to be square integrable functions defined on a d -dimensional torus domain $\Omega = \mathbb{R}^d / \mathbb{Z}^d$ ($J_n \in L^2(\Omega, \mathbb{R})$, $n \in \{1, \dots, N\}$) and $\text{Diff}(\Omega)$ be the space of diffeomorphisms. The problem of diffeomorphic atlas building is to find the template $I \in L^2(\Omega, \mathbb{R})$ and the deformation $\phi_n \in \text{Diff}(\Omega)$ from template I to each input image J_n that minimize the energy function

$$E(\{\phi_n\}, I) = \sum_{n=1}^N \text{Dist}(J_n, I \circ \phi_n^{-1}) + \text{Reg}(\phi_n), \quad (1)$$

where \circ is a composition operator that resamples I by the inverse of the smooth mapping ϕ_n , $\text{Dist}(\cdot, \cdot)$ denotes a distance func-

tion between images such as sum-of-squared difference (SSD), normalized cross correlation (NCC), or mutual information (MI), and $\text{Reg}(\cdot)$ is a regularization term that enforces smoothness of the transformations.

2.1. Flows of diffeomorphisms and geodesics

The optimization of the energy function (1) over the transformations $\{\phi_n\}$ is challenging due to the nonlinearity of the space of diffeomorphisms. Mathematically, we consider the time-varying deformation $\phi_n(t, x)$: $t \in [0, 1]$, $x \in \Omega$ to be generated by the integral flow of time-varying velocity field $v_n(t, x) \in V$ in the tangent space of diffeomorphisms at the identity Id ($V = T_{\text{Id}}\text{Diff}(\Omega)$):

$$\frac{d\phi_n(t, x)}{dt} = v_n \circ \phi_n(t, x), \quad \phi_n(0, x) = \text{Id}.$$

The geodesic path between the identity element and transformation ϕ_n is uniquely determined by a right-invariant Riemannian metric $\|\cdot\|_V$ on the time-dependent velocity fields as

$$\int_0^1 \|v_n(t, x)\|_V dt. \quad (2)$$

The geodesic is obtained at the minimum of (2) by integrating the Euler-Poincaré differential equation (EPDiff) (Arnol'd, 1966; Miller et al., 2006) with the initial condition of $v_n(t, x)$ at $t = 0$:

$$\begin{aligned} \frac{\partial v_n}{\partial t} &= -\text{ad}_v^\dagger v \\ &= -\mathcal{K}[(Dv_n)^T m_n + Dm_n v_n + m_n \text{div}(v_n)], \end{aligned} \quad (3)$$

where ad^\dagger is an adjoint operator, D is the Jacobian matrix and div is the divergence operator. The operator \mathcal{K} is the inverse of a symmetric, positive-definite differential operator $\mathcal{L}: V \rightarrow V^*$ that maps a velocity field $v_n \in V$ to a momentum vector $m_n \in V^*$ such that $m_n = \mathcal{L}v_n$ and $v_n = \mathcal{K}m_n$. Evaluation of Eq. (3) is known as *geodesic shooting* (Younes et al., 2009; Vialard et al., 2012). It has been shown that the geodesic shooting algorithm substantially reduces the computational complexity and improves the optimization landscape by only manipulating the initial velocity with the geodesic evolution Eq. (3). Therefore, in this paper we choose to optimize over initial velocities rather than the entire time-dependent velocity fields. With a slight abuse of notation, we set $v_n \triangleq v_n(0, x)$ to represent the initial velocity for the n th image J_n in the remaining sections.

2.2. Fourier representation of velocity fields

It has been recently shown that the velocity fields generated by the EPDiff (3) can be efficiently captured via a discrete low dimensional bandlimited representation in the Fourier space (Zhang and Fletcher, 2015b), which dramatically speeds up geodesic shooting algorithm. The main idea is that the velocity fields do not develop high frequencies and only a small amount of low frequencies contributes to the transformations (Fig. 1), therefore working in a bandlimited space captures the deformations as accurately as the original algorithm. Here we briefly review the relevant details of the method.

Let $f: \mathbb{R}^d \rightarrow \mathbb{R}$ be a real-valued function. The Fourier transform \mathcal{F} of f is given by

$$\mathcal{F}[f](\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \langle \xi, x \rangle} dx, \quad (4)$$

where $x = (x_1, \dots, x_d)$ is a d -dimensional image coordinate vector, $\xi = (\xi_1, \dots, \xi_d)$ is a d -dimensional frequency vector, and $\langle \cdot, \cdot \rangle$ denotes the inner product operator. The inverse Fourier transform \mathcal{F}^{-1} of a discretized Fourier signal \tilde{f}

$$\mathcal{F}^{-1}[\tilde{f}](x) = \sum_{\xi} \tilde{f}(\xi) e^{2\pi i \langle \xi, x \rangle} \quad (5)$$

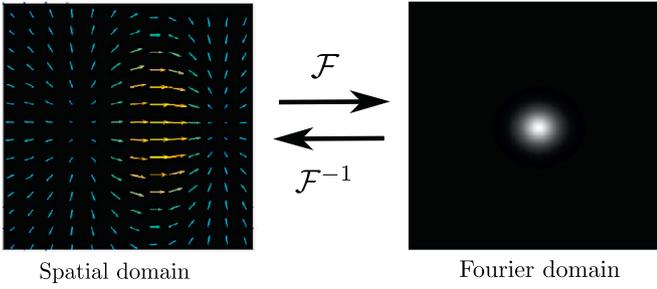


Fig. 1. Velocity fields in spatial and Fourier domain.

is an approximation of the original signal f . For vector-valued functions, such as diffeomorphisms ϕ and velocity fields v , we apply the (inverse) Fourier transform to each vector component separately.

Analogous to the definition of a distance metric in (2), Zhang and Fletcher (2015b) developed a new representation of velocity fields entirely in the frequency domain that leads to an efficient computation of diffeomorphisms in a low dimensional space. In particular, if \tilde{V} is the discrete Fourier space of velocity fields, then for any elements $\tilde{u}, \tilde{v} \in \tilde{V}$, the distance metric at identity is defined as

$$\langle \tilde{u}, \tilde{v} \rangle_{\tilde{V}} = \sum_{\xi} (\tilde{\mathcal{L}}\tilde{u}(\xi), \tilde{v}(\xi)),$$

where $\tilde{\mathcal{L}}: \tilde{V} \mapsto \tilde{V}^*$ is the Fourier transform of a differential operator, e.g., a commonly used Laplacian operator $(-\alpha\Delta + e)^c$ with a positive weight parameter α and a smoothness parameter c , and (\cdot, \cdot) is a dot product in the frequency space. The Fourier transform of the Laplacian operator is given by

$$\tilde{\mathcal{L}}(\xi) = (-2\alpha \sum_{j=1}^d (\cos(2\pi\xi_j) - 1) + 1)^c.$$

The Fourier coefficients of the inverse operator $\mathcal{K}: \tilde{V}^* \mapsto \tilde{V}$ can be easily computed as $\tilde{\mathcal{K}}(\xi) = 1/\tilde{\mathcal{L}}(\xi)$.

Since \mathcal{K} is a smoothing operator that suppresses high frequencies in the Fourier domain, the geodesic evolution Eq. (3) indicates that the velocity field v at each time point can be represented efficiently as a bandlimited signal in the Fourier space as

$$\begin{aligned} \frac{\partial \tilde{v}}{\partial t} &= -\text{ad}_{\tilde{v}}^{\dagger} \tilde{v} \\ &= -\tilde{\mathcal{K}}[(\tilde{D}\tilde{v})^T \star \tilde{m} + \tilde{D}\tilde{m} \star \tilde{v} + \tilde{m} \star \tilde{\nabla} \cdot \tilde{v}], \end{aligned} \quad (6)$$

where $\tilde{m} = \tilde{\mathcal{L}}\tilde{v}$, and \star is the truncated matrix-vector field auto-correlation¹ and $\tilde{D}\tilde{v}$ is a tensor product $\tilde{D} \otimes \tilde{v}$ with $\tilde{D}(\xi) = i \sin(2\pi\xi)$ representing the Fourier frequencies of a central difference Jacobian matrix D . The operator $\tilde{\nabla} \cdot$ is the discrete divergence operator that is computed as the sum of the Fourier coefficients of the central difference operator \tilde{D} along each dimension, i.e., $\tilde{\nabla} \cdot \xi = \sum_{j=1}^d i \sin(2\pi\xi_j)$.

All computational operations in (6) are easy to implement in a truncated low dimensional space by eliminating the high frequencies. To ensure that \tilde{f} represents a real-valued vector field in the spatial domain, we require $\tilde{f}(\xi_1, \dots, \xi_d) = \tilde{f}^*(-\xi_1, \dots, -\xi_d)$, where $*$ denotes the complex conjugate. We build on the fast computation of diffeomorphisms introduced in Zhang and Fletcher (2015b)

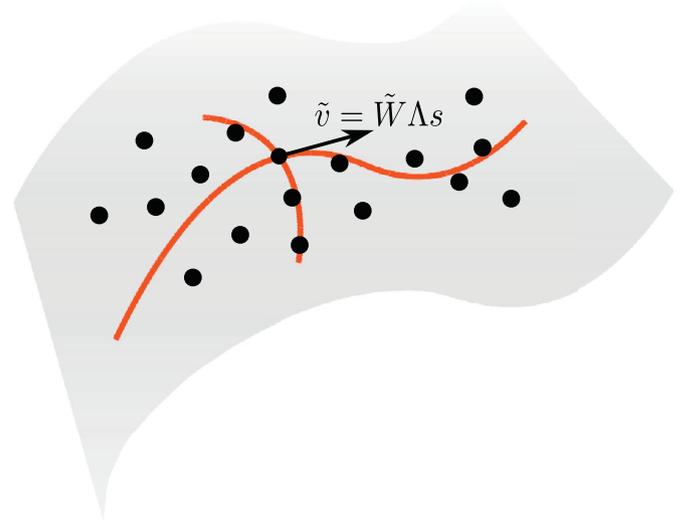


Fig. 2. Principal analysis of diffeomorphisms.

to demonstrate an efficient diffeomorphic shape analysis in the same low dimensional Fourier space.

3. Generative model

We introduce a generative model for principal geodesic analysis of diffeomorphisms represented in the bandlimited velocity space \tilde{V} , with shape variability explicitly encoded as factors of the model.

Let $\tilde{W} \in \mathbb{C}^{p \times q}$ be a matrix in the Fourier space whose q columns ($q < N$) are orthonormal principal initial velocities in a low p -dimensional space with unit length, $\Lambda \in \mathbb{R}^{q \times q}$ be a diagonal matrix of scale factors for the columns of \tilde{W} , and $s \in \mathbb{R}^q$ be a vector of random factors that parameterizes the space of initial velocities. Therefore, each initial velocity is generated as $\tilde{v} = \tilde{W}\Lambda s$ (see Fig. 2).

For subject $n \in \{1, \dots, N\}$, we define a prior on the loading coefficient vector s_n to be a Gaussian distribution whose covariance matrix is a combination of the identity matrix e and a matrix $(\tilde{\mathcal{L}}\Lambda\tilde{W}^T\tilde{W}\Lambda)^{-1}$ that ensures the smoothness of the geodesic path, i.e.,

$$\begin{aligned} p(s_n | \tilde{W}, \Lambda) &= \mathcal{N}(s_n; 0, (\tilde{\mathcal{L}}\Lambda\tilde{W}^T\tilde{W}\Lambda)^{-1} + e) \\ &= \mathcal{N}(s_n; 0, \tilde{\mathcal{L}}^{-1}\Lambda^{-2} + e). \end{aligned}$$

The normalizing constant of $p(s_n | \tilde{W}, \Lambda)$ including the determinant of the covariance matrix is computed as

$$(2\pi)^{q/2} |\tilde{\mathcal{L}}^{-1}\Lambda^{-2} + e|^{1/2} = (2\pi)^{q/2} \cdot \prod_{l=1}^q \left(\frac{1}{\tilde{\mathcal{L}}(l, l)\Lambda^2(l, l)} + 1 \right),$$

where $l \in \{1, \dots, q\}$ denote the diagonal element.

Assuming i.i.d. Gaussian noise on image intensities, we obtain the likelihood

$$p(J_n | s_n; \tilde{W}, \Lambda, I, \sigma) = \mathcal{N}(J_n; I \circ \phi_n^{-1}, \sigma^2),$$

where ϕ_n is a deformation that corresponds to the initial velocity $v_n = \mathcal{F}^{-1}[\tilde{W}\Lambda s_n]$ in the spatial domain, that is,

$$\frac{d\phi_n}{dt} = \mathcal{F}^{-1}[\tilde{W}\Lambda s_n] \circ \phi_n, \quad (7)$$

and σ^2 is the image noise variance.

Defining $\Theta = \{\tilde{W}, \Lambda, I, \sigma\}$, we employ Bayes' rule to arrive at the posterior distribution of s_n :

$$p(s_n | J_n; \Theta) \propto p(J_n | s_n; \Theta) \cdot p(s_n | \tilde{W}, \Lambda) \quad (8)$$

¹ The auto-correlation operates on zero-padded signals followed by truncating back to the bandlimits in each dimension to guarantee the output remains bandlimited.

$$= \mathcal{N}(J_n; I \circ \phi_n^{-1}, \sigma^2) \cdot \mathcal{N}(s_n; 0, \tilde{\mathcal{L}}^{-1} \Lambda^{-2} + e).$$

The log posterior distribution of the loading coefficients s_1, \dots, s_N for the entire image collection is therefore

$$\begin{aligned} Q &\triangleq \log p(s_1, \dots, s_N | J_1, \dots, J_N; \Theta) \\ &= \sum_{n=1}^N \log p(J_n | s_n; \Theta) + \log p(s_n | \tilde{W}, \Lambda) + \text{const.} \\ &= \sum_{n=1}^N \left[-\frac{\|J_n - I \circ \phi_n^{-1}\|_{L^2}^2}{2\sigma^2} - \frac{s_n^T (\tilde{\mathcal{L}}^{-1} \Lambda^{-2} + e) s_n}{2} \right. \\ &\quad \left. - \frac{dN}{2} (\log \sigma) - \frac{N}{2} \sum_{l=1}^q \log \left(\frac{1}{\tilde{\mathcal{L}}_{ll} \Lambda_{ll}^2} + 1 \right) \right] + \text{const.} \end{aligned} \quad (9)$$

4. Inference

We present two alternative ways to estimate the model parameters: the maximum a posteriori (MAP) and the Monte Carlo expectation maximization (MCEM) that treats the loading coefficients $\{s_1, \dots, s_N\}$ as latent variables.

MAP. We use gradient ascent to maximize the log posterior probability (9) with respect to the parameters Θ and latent variables $\{s_n\}$.

By setting the derivative of Q with respect to I and σ to zero, we obtain closed-form updates for the atlas template I and noise variance σ^2 :

$$I = \frac{\sum_{n=1}^N J_n \circ \phi_n |D\phi_n|}{\sum_{n=1}^N |D\phi_n|},$$

$$\sigma^2 = \frac{1}{MN} \sum_{n=1}^N \|J_n - I \circ \phi_n^{-1}\|_{L^2}^2,$$

where M is the number of image voxels.

To estimate the matrix of principal directions \tilde{W} , the scaling factor Λ , and the loading coefficients $\{s_1, \dots, s_N\}$, we follow the derivations in Zhang and Fletcher (2015b) and first obtain the gradient of Q w.r.t. the initial velocity \tilde{v}_n as follows:

- (i) Forward integrate the geodesic evolution Eq. (6) to compute time-varying velocity fields $\{\tilde{v}_n\}$ and then follow the flow Eq. (7) to generate a flow of diffeomorphic transformations $\{\phi_n\}$.
- (ii) Compute the gradient $\nabla_{\tilde{v}_n} Q$ at time point $t = 1$ as

$$\begin{aligned} \delta Q_1 &\triangleq [\nabla_{\tilde{v}_n} Q]_{t=1} \\ &= -\tilde{K} \mathcal{F} \left[\frac{1}{\sigma^2} (J_n - I \circ \phi_n^{-1}) \cdot \nabla (I \circ \phi_n^{-1}) \right]. \end{aligned} \quad (10)$$

- (iii) Backward integrate the gradient (10) to $t = 0$ to obtain $\delta Q_0 \triangleq [\nabla_{\tilde{v}_n} Q]_{t=0}$ by using reduced adjoint Jacobi field equations (Francesco, 1995; Zhang and Fletcher, 2015b)

$$\frac{d\hat{v}}{dt} = -\text{ad}_{\hat{v}}^{\dagger} \hat{h}, \quad \frac{d\hat{h}}{dt} = -\hat{v} - \text{ad}_{\hat{v}} \hat{h} + \text{ad}_{\hat{h}}^{\dagger} \hat{v},$$

where $\text{ad}_{\hat{v}} \hat{h} = \tilde{D}\hat{v} * \hat{h} - \tilde{D}\hat{h} * \hat{v}$ with $*$ being a truncated convolution operator, and $\hat{v}, \hat{h} \in \hat{V}$ are introduced adjoint variables.

After applying the chain rule, we have the gradient of Q for updating the loading factor s_n :

$$\nabla_{s_n} Q = -\Lambda \tilde{W}^T \delta Q_0 - s_n.$$

The gradients of Q w.r.t. \tilde{W} and Λ are given as follows:

$$\nabla_{\tilde{W}} Q = -\sum_{n=1}^N \delta Q_0 s_n^T \Lambda,$$

$$\nabla_{\Lambda} Q = -\sum_{n=1}^N \left(\tilde{W} s_n^T \delta Q_0 - \frac{1}{\tilde{\mathcal{L}} \Lambda^2 (\tilde{\mathcal{L}} \Lambda^2 + 1)} \right).$$

Unlike the PPGA model (Zhang and Fletcher, 2014), we enforce the mutual orthogonality constraint on the columns of \tilde{W} since it is computationally tractable in the low dimensional space. There are two natural ways to satisfy this constraint: first is to treat \tilde{W} as a point on the complex Stiefel manifold $V_n(\mathbb{C}^d)$, which is a set of orthonormal n -frames in \mathbb{C}^d (Edelman et al., 1998). This requires projecting the gradient of \tilde{W} onto the tangent space of $V_n(\mathbb{C}^d)$, and then updating \tilde{W} within a small step along the projected gradient direction. Another way is to use Gram-Schmidt process (Cheney and Kincaid, 2009) for orthonormalizing the column vectors of \tilde{W} in a complex inner product space. We employ the latter scheme in our implementation.

MCEM. To treat the loading coefficients $\{s_n\}$ fully as latent random variables, we integrate them out from the posterior distribution (9) by using a Hamiltonian Monte Carlo (HMC) method (Duane et al., 1987) due to the fact that direct sampling is difficult. This scheme includes two main steps:

- (i) Draw a random sample of size S of the latent variables $\{s_n\}$ via HMC sampling based on current parameters $\Theta^{(k)}$. Let $s_{jn}, j = 1, \dots, S$, denote the j th sample for the subject n . A Hamiltonian function $H(\mathbf{s}, \beta) = U(\mathbf{s}) + V(\beta)$ that consists of a potential energy $U(\mathbf{s}) = -\log p(\mathbf{s} | J; \Theta)$ and a kinetic energy $V(\beta) = -\log g(\beta)$, where $g(\beta)$ is typically an independent Gaussian distribution on an auxiliary variable β , is constructed to simulate the sampling system. Starting from the current point (\mathbf{s}, β) , the Hamiltonian function H produces a candidate point $(\hat{\mathbf{s}}, \hat{\beta})$ that is accepted as a new sample with probability

$$p_{\text{accept}} = \min(1, \exp(-U(\hat{\mathbf{s}}) - V(\hat{\beta}) + U(\mathbf{s}) + V(\beta))).$$

The sample mean is taken to approximate the expectation:

$$\Upsilon(\Theta | \Theta^{(i)}) \approx \frac{1}{S} \sum_{j=1}^S \sum_{n=1}^N \log p(s_{jn} | J_n; \Theta^{(i)}), \quad (11)$$

where the superscript (i) denotes the current state of the parameter set Θ .

- (ii) Maximize the expectation function Υ to update parameters Θ . By setting its derivatives with respect to I and σ^2 to zero, we obtain closed-form updates for the atlas template I and noise variance σ^2 as

$$I = \frac{\sum_{j=1}^S \sum_{n=1}^N I \circ \phi_{jn} |D\phi_{jn}|}{\sum_{j=1}^S \sum_{n=1}^N |D\phi_{jn}|},$$

$$\sigma^2 = \frac{1}{SMN} \sum_{j=1}^S \sum_{n=1}^N \|J_n - I \circ \phi_{jn}^{-1}\|_{L^2}^2.$$

Since there is no closed-form update for \tilde{W} and Λ , we use gradient ascent to estimate the principal initial velocity basis \tilde{W} and the scaling matrix Λ . The gradients w.r.t. \tilde{W}, Λ of (11) are given as follows:

$$\begin{aligned} \nabla_{\tilde{W}} \Upsilon &= -\sum_{j=1}^S \sum_{n=1}^N [\nabla_{\tilde{v}_n} \Upsilon]_{t=0} s_{jn}^T \Lambda, \\ \nabla_{\Lambda} \Upsilon &= -\sum_{j=1}^S \sum_{n=1}^N \left(\tilde{W} s_{jn}^T [\nabla_{\tilde{v}_n} \Upsilon]_{t=0} - \frac{1}{\tilde{\mathcal{L}} \Lambda^2 (\tilde{\mathcal{L}} \Lambda^2 + 1)} \right). \end{aligned}$$

5. Evaluation

To evaluate the effectiveness of the proposed *low-dimensional principal geodesic analysis* (LPPGA) model, we applied the algorithm

to brain MRI scans of 90 subjects from the ADNI study (Jack et al., 2008), aged 60 to 90. Fifty subjects have Alzheimer’s disease (AD) and the remaining 40 subjects are healthy controls. All MRI scans have the same resolution $128 \times 128 \times 128$ with the voxel size of $1.25 \times 1.25 \times 1.25\text{mm}^3$. All images underwent the preprocessing of skull stripping, downsampling, intensity normalization to $[0, 1]$ interval, bias field correction, and co-registration with affine transformations.

We first estimate a full collection of principal modes $q = 89$ for our model, using $\alpha = 3.0$, $c = 3.0$ for the differential operator $\hat{\mathcal{L}}$ with $p = 16^3$ dimensions of the initial velocity field \tilde{v} , which is similar to the settings used in pairwise diffeomorphic image registration (Zhang and Fletcher, 2015b). The number of time steps for integration in geodesic shooting is set to 10. We initialize the atlas I to be the average of image intensities, Λ to be the identity matrix, s_n to be the all-ones vector, and the principal initial velocity matrix \tilde{W} to be the principal components estimated by TPCA (Vaillant et al., 2004) that runs linear PCA in the space of initial velocity fields after atlas building. For the HMC sampling of the MCEM variant of our model, we use the step size of 0.01 for leap-frog integration with 20 units of time discretization in integration of EPDiff equations.

To investigate the ability of our model to capture anatomical variability, we use the loading coefficients $s = \{s_1, \dots, s_N\}$ as a shape descriptor in a statistical study. The idea is to test the hypothesis that the principal modes estimated by our method are correlated significantly with clinical information such as minimal state examination (MMSE), Alzheimer’s Disease Assessment Scale (ADAS), and Clinical Dementia Rating (CDR). We project the transformations that are derived from the estimated atlas I_0 and each individual from a testing dataset with 40 subjects onto the estimated principal modes. We then fit the clinical score MMSE, ADAS, and CDR using a linear regression model on the computed loading coefficients.

We use the previous state of PPGA (Zhang and Fletcher, 2014) in a high dimensional image space and TPCA (Vaillant et al., 2004) as two baseline methods. In order to conduct a fair comparison, we keep all the parameters including regularization and time steps for numerical integration fixed across the three algorithms. To evaluate the model stability, we rerun the entire experiment 50 times on randomly sampled subsets of 50 images.

6. Results

Fig. 3 reports the cumulative variance explained by the model as a function of the model size. Both variants of our approach LPPGA-MCEM and LPPGA-MAP achieve higher representation accuracy than the two state-of-the-art baseline algorithms across the entire range of model sizes. This is mainly because that conducting statistical analysis in the low dimensional space improves the gradient-based optimization landscape, where local minima often occur in a high dimensional image space. The Monte Carlo sampling of MCEM algorithm further reduces the risk of getting stuck in local minima by allowing random steps away from the current minimal solution.

Table 1 reports the number of principal modes required to achieve the same level of shape variation across the entire dataset. Our model LPPGA-MCEM / LPPGA-MAP captures better shape changes while using fewer number of principal modes, which also means that our model estimates more compact representation of the image data.

Fig. 4 visualizes the first three modes of variation in this cohort by shooting the estimated atlas I along the initial velocities $\tilde{v} = a_i \tilde{W}_i \Lambda_i$ ($a_i = \{-2, -1, 0, 1, 2\}$, $i = 1, 2, 3$). We also show the log determinant of the Jacobian at $a_i = 2$. The first mode of variation clearly reflects that changes in the ventricle size, which is the

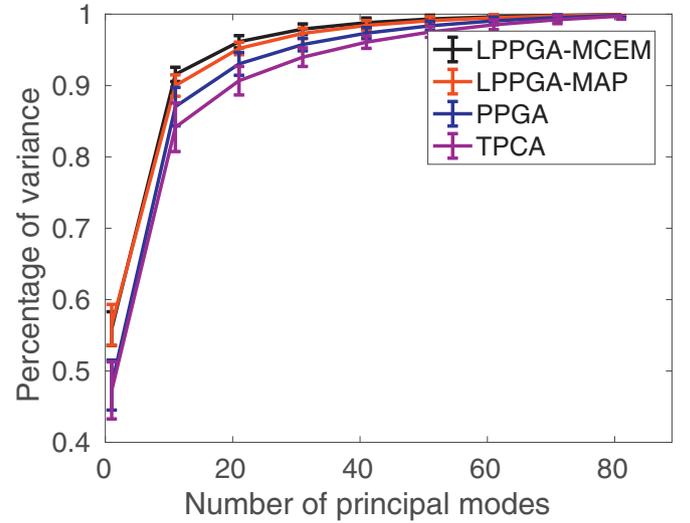


Fig. 3. Cumulative variance explained by principal modes estimated from our model (LPPGA-MCEM and LPPGA-MAP) and baseline algorithms (PPGA-MAP and TPCA).

Table 1

Number of principal modes that achieves 90% and 95% of total variance.

| Method | 90% | 95% |
|------------|-----|-----|
| LPPGA-MCEM | 9 | 17 |
| LPPGA-MAP | 11 | 20 |
| PPGA-MAP | 15 | 27 |
| TPCA | 19 | 35 |

Table 2

Comparison of linear regression models on the first two principal mode for our model (LPPGA-MCEM / LPPGA-MAP) and the baseline algorithms (PPGA and TPCA) on 40 brain MRIs from ADNI.

| MMSE | | | | |
|------------|----------|-------|-------|---------------|
| Model | Residual | R^2 | F | p -value |
| LPPGA-MCEM | 4.42 | 0.19 | 21.68 | $1.13e^{-5}$ |
| LPPGA-MAP | 4.45 | 0.18 | 19.47 | $2.18e^{-5}$ |
| PPGA | 4.49 | 0.16 | 17.96 | $5.54e^{-5}$ |
| TPCA | 4.53 | 0.14 | 16.34 | $1.10e^{-4}$ |
| ADAS | | | | |
| Model | Residual | R^2 | F | p -value |
| LPPGA-MCEM | 8.25 | 0.21 | 13.14 | $1.033e^{-5}$ |
| LPPGA-MAP | 8.36 | 0.19 | 11.68 | $3.20e^{-5}$ |
| PPGA | 8.41 | 0.18 | 11.10 | $5.09e^{-5}$ |
| TPCA | 8.65 | 0.17 | 10.75 | $1.03e^{-4}$ |
| CDR | | | | |
| Model | Residual | R^2 | F | p -value |
| LPPGA-MCEM | 2.21 | 0.22 | 24.78 | $3.16e^{-6}$ |
| LPPGA-MAP | 2.22 | 0.20 | 23.99 | $4.37e^{-6}$ |
| PPGA | 2.23 | 0.19 | 22.92 | $6.77e^{-6}$ |
| TPCA | 2.25 | 0.17 | 21.54 | $2.88e^{-5}$ |

dominant source of variability in the brain shape. The algorithm estimates standard deviation of the image noise to be $\sigma = 0.02$.

Fig. 5 reports run time and memory consumption for building the full model of anatomical variability. Our approach LPPGA-MAP offers an order of magnitude improvement in both the run time and memory requirements while providing a more powerful model of variability. While the MCEM variant is computationally more expensive than all baseline methods due to the sampling procedure, it provides better statistical analysis of regression (as reported in Table 2) than the two baseline algorithms using the first two prin-

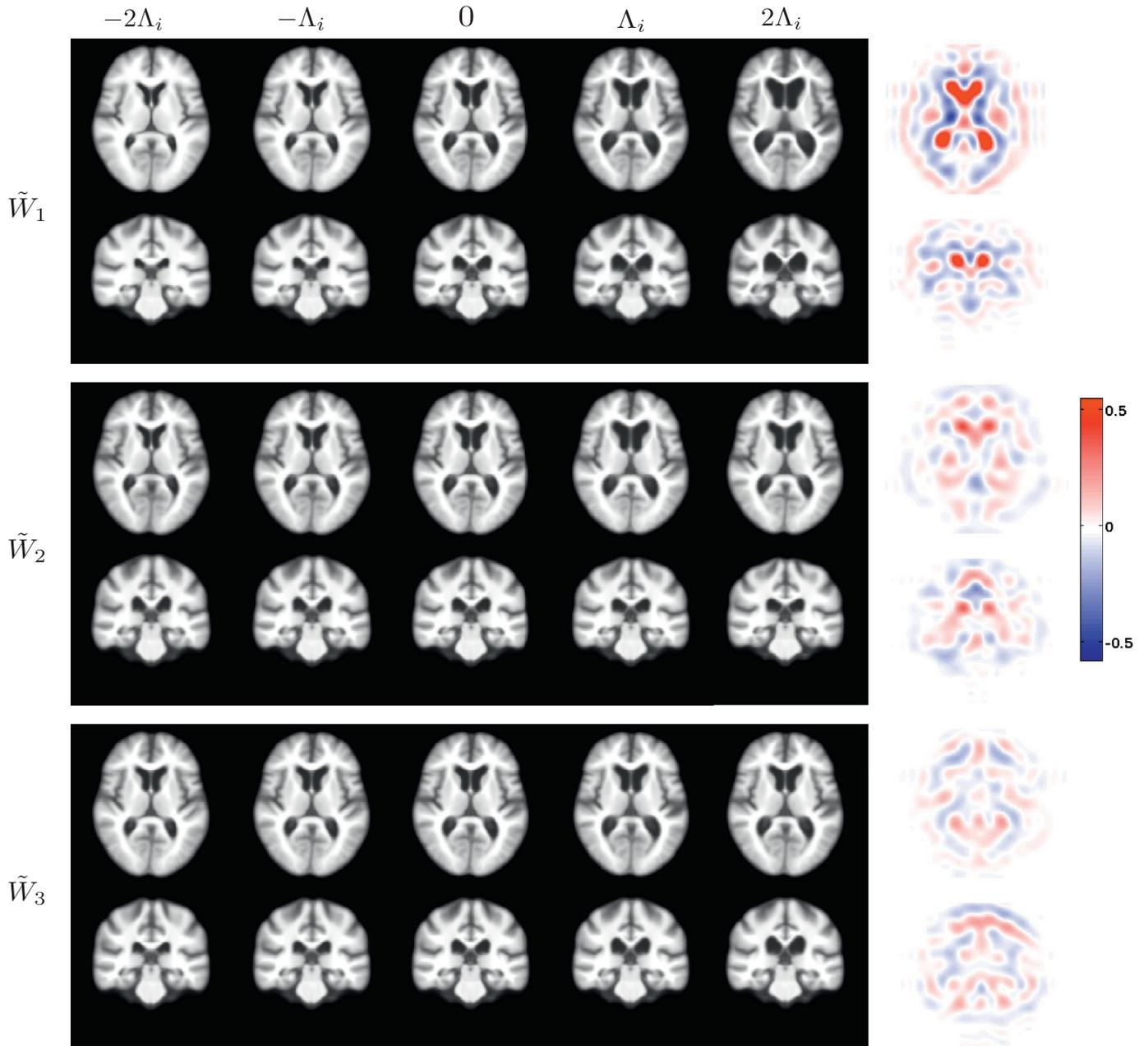


Fig. 4. Top to bottom: first, second and third principal modes of brain shape variation estimated by our model LPPGA-MCEM for varying amounts of the corresponding principal mode, and log determinant of the transformation Jacobians at $2\Lambda_i$ (regions of expansion in red and contraction in blue). Axial and coronal views are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

principal modes. The higher F and R^2 statistics indicate that our approach captures more variation of the MMSE scores than the other models. Another advantage of such Monte Carlo approach is that it provides consistent statistics in noisy case (Allasonnière et al., 2007) and better model selection.

7. Discussion and conclusion

We presented a low dimensional probabilistic framework for factor analysis in the space of diffeomorphisms. Our model explicitly optimizes the fit of the principal modes to the data in a low dimensional space of bandlimited velocity fields, which results in (1) better data fitting, and (2) dramatically lower computational cost with more powerful statistical analysis. We developed an inference strategy based on MAP to estimate parameters, including the prin-

incipal modes, noise variance, and image atlas simultaneously. Our model also enables Monte Carlo sampling because of the efficient low dimensional parametrization. We demonstrated that the estimated low dimensional latent loading coefficients provide a compact representation of the anatomical variability and yield a better statistical analysis of anatomical changes associated with clinical variables.

This work represents the first step towards *efficient* probabilistic models of shape variability based on high-dimensional diffeomorphisms. There are several avenues for future work to build upon our model. We will explore Bayesian variants of shape analysis that infer the inherent dimensionality directly from the data by formulating dimensionality reduction with a sparsity prior. Reducing the dimensionality to the inherent modes of shape variability

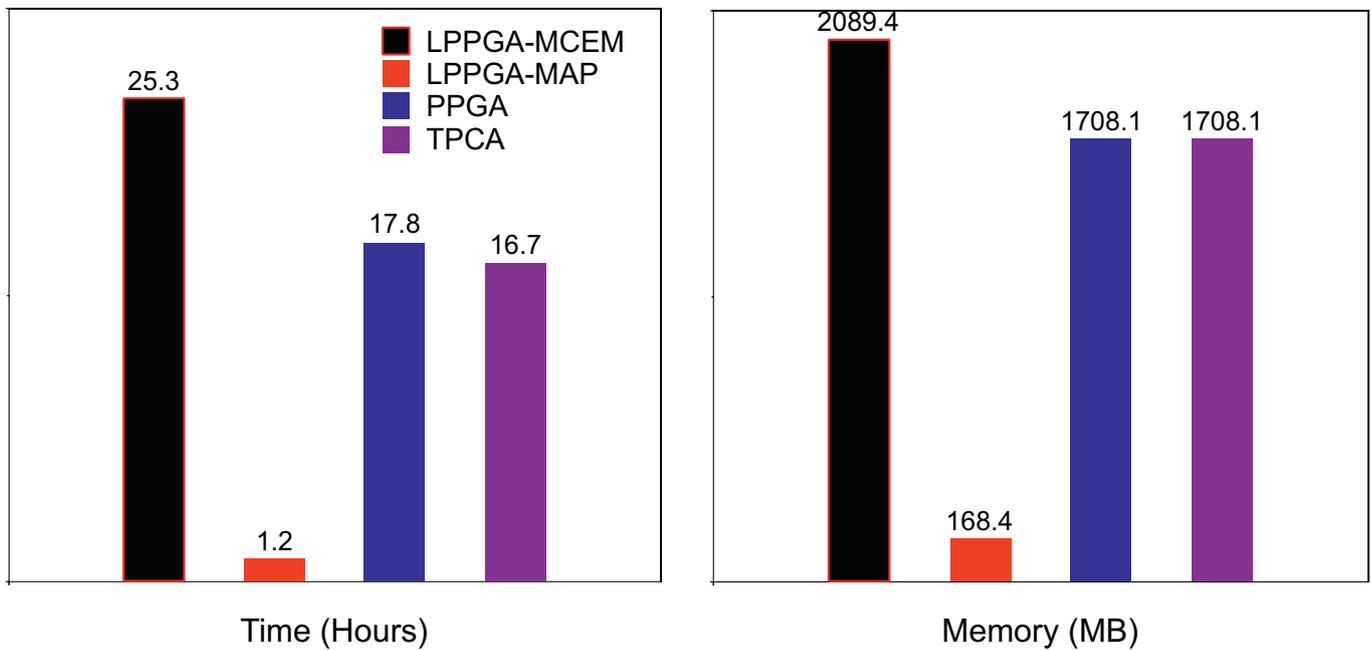


Fig. 5. Comparison of run time and memory consumption. The implementation employed a message passing interface (MPI) parallel programming for all methods and distributed 90 subjects to 10 processors.

has the potential to improve hypothesis testing, classification, and mixture models. A multiscale strategy like that of Sommer et al. (2013) can be added to our model to make the inference even faster. Moreover, since Monte Carlo sampling is computationally more tractable in our model, we can automatically estimate the regularization parameter jointly with the shape variability model. This eliminates the effort of hand-tuning on parameters and enables uncertainty quantification of the hidden variables. Another interesting avenue is to estimate an even more sharp atlas that has clearer details of brain structures such as sulci. Since the atlas is essentially the average over intensities of all inter-subjects, it is possible that structures with relatively large differences across subjects get smoothed out under the spatially-invariant smoothness constraints. Therefore, developing a spatially-varying kernel that penalizes local smoothness is desirable for the problem of atlas estimation.

Acknowledgments

This work was supported by NIH NIBIB NAC P41EB015902, NIH NINDS R01NS086905, NIH NICHD U01HD087211, NCI/NIH P41EB015898, and Wistron Corporation. The data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904). All the investigators within the ADNI provided data but did not participate in the analysis or writing of this paper.

References

- Allasonnière, S., Amit, Y., Trouvé, A., 2007. Toward a coherent statistical framework for dense deformable template estimation. *J. R. Stat. Soc. Series B* 69 (1), 3–29.
- Arnol'd, V.I., 1966. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Ann. Inst. Fourier* 16, 319–361.
- Beg, M., Miller, M., Trouvé, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* 61 (2), 139–157.
- Bookstein, F.L., 1997. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press.
- Cheney, W., Kincaid, D., 2009. *Linear algebra: theory and applications*. Aust. Math. Soc. 110.

- Christensen, G.E., Rabbitt, R.D., Miller, M.I., 1993. A deformable neuroanatomy textbook based on viscous fluid mechanics. In: 27th Annual Conference on Information Sciences and Systems, Citeseer, pp. 211–216.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., et al., 1995. Active shape models—their training and application. *Comput. Vis. Image Understand.* 61 (1), 38–59.
- Duane, S., Kennedy, A., Pendleton, B., Roweth, D., 1987. Hybrid Monte Carlo. *Phys. Lett. B* 195 (2), 216–222.
- Edelman, A., Arias, T., Smith, S., 1998. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* 20 (2), 303–353.
- Fletcher, P.T., Lu, C., Joshi, S., 2003. Statistics of shape via principal geodesic analysis on Lie groups. In: *Computer Vision and Pattern Recognition*, 1. IEEE, pp. 1–95.
- Francesco, B., 1995. *Invariant affine connections and controllability on Lie groups*. Technical Report, technical Report for Geometric Mechanics, California Institute of Technology.
- Gerig, G., Styner, M., Shenton, M.E., Lieberman, J.A., 2001. Shape versus size: improved understanding of the morphology of brain structures. In: *Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 24–32.
- Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S., 2013. Bayesian atlas estimation for the variability analysis of shape complexes. In: *Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 267–274.
- Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., et al., 2008. The alzheimer's disease neuroimaging initiative (adni): MRI methods. *J. Magn. Reson. Imaging* 27 (4), 685–691.
- Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 23, S151–S160.
- Miller, M.I., Trouvé, A., Younes, L., 2006. Geodesic shooting for computational anatomy. *J. Math. Imag. Vis.* 24 (2), 209–228.
- Nemmi, F., Sabatini, U., Rascol, O., Péran, P., 2015. Parkinson's disease and local atrophy in subcortical nuclei: insight from shape analysis. *Neurobiol. Aging* 36 (1), 424–433.
- Onofrey, J.A., Staib, L.H., Papademetris, X., 2013. Semi-supervised learning of non-rigid deformations for image registration. In: *International MICCAI Workshop on Medical Computer Vision*, Springer, pp. 13–23.
- Pizer, S.M., Fritsch, D.S., Yushkevich, P.A., Johnson, V.E., Chaney, E.L., 1999. Segmentation, registration, and measurement of shape variation via image object shape. *Med. Imag. IEEE Trans.* 18 (10), 851–865.
- Qiu, A., Younes, L., Miller, M.I., 2012. Principal component based diffeomorphic surface mapping. *Med. Imag. IEEE Trans.* 31 (2), 302–311.
- Sommer, S., Lauze, F., Nielsen, M., Pennec, X., 2013. Sparse multi-scale diffeomorphic registration: the kernel bundle framework. *J. Math. Imag. Vis.* 46 (3), 292–308.
- Sweet, A., Pennec, X., 2010. A log-euclidean statistical analysis of dti brain deformations. *MICCAI 2010 Workshop on Computational Diffusion MRI*.
- Twining, C., Cootes, T., Marsland, S., Petrovic, V., Schestowitz, R., Taylor, C., 2005. A unified information-theoretic approach to groupwise non-rigid registration and model building. In: *Information Processing in Medical Imaging*, Springer, pp. 1–14.
- Vaillant, M., Miller, M.I., Younes, L., Trouvé, A., 2004. Statistics on diffeomorphisms via tangent space representations. *Neuroimage* 23, S161–S169.

- Vialard, F.-X., Risser, L., Holm, D., Rueckert, D., 2011. Diffeomorphic atlas estimation using Kärcher mean and geodesic shooting on volumetric images. MIUA.
- Vialard, F.-X., Risser, L., Rueckert, D., Cotter, C.J., 2012. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *Int. J. Comput. Vis.* 97 (2), 229–241.
- Younes, L., Arrate, F., Miller, M., 2009. Evolutions equations in computational anatomy. *Neuroimage* 45 (1), S40–S50.
- Zhang, M., Fletcher, P.T., 2014. Bayesian principal geodesic analysis in diffeomorphic image registration. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*. Springer, pp. 121–128.
- Zhang, M., Fletcher, P.T., 2015. Bayesian principal geodesic analysis for estimating intrinsic diffeomorphic image variability. *Med. Image Anal.* 25 (1), 37–44.
- Zhang, M., Fletcher, P.T., 2015. Finite-dimensional lie algebras for fast diffeomorphic image registration. In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 249–260.
- Zhang, M., Wells III, W.M., Golland, P., 2016. Low-dimensional statistics of anatomical variability via compact representation of image deformations. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 166–173.