Low-Dimensional Statistics of Anatomical Variability via Compact Representation of Image Deformations

Miaomiao Zhang[1(✉)], William M. Wells III[1,2], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, USA
miao86@mit.edu
[2] Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Abstract. Using image-based descriptors to investigate clinical hypotheses and therapeutic implications is challenging due to the notorious "curse of dimensionality" coupled with a small sample size. In this paper, we present a low-dimensional analysis of anatomical shape variability in the space of diffeomorphisms and demonstrate its benefits for clinical studies. To combat the high dimensionality of the deformation descriptors, we develop a probabilistic model of principal geodesic analysis in a bandlimited low-dimensional space that still captures the underlying variability of image data. 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 models based on the high-dimensional state-of-the-art approaches such as tangent space PCA (TPCA) and probabilistic principal geodesic analysis (PPGA).

1 Introduction

Shape analysis is critical for image-based studies of disease as it offers characterizations of anatomical variability between different groups, or in the course of a disease. Analysis of shape changes can provide new insights into the nature of the disease and support treatment. For example, brain atrophy has been identified in patients affected by neuro-degenerative diseases such as Parkinson's, Huntington's, and Alzheimer's [5,10]. When combined with other clinical information, characterization of shape differences between clinical cohorts and a healthy population can be useful in predicting disease progression. Landmarks [3], distance transforms [6,9], and medial cores [11] are examples of image-based shape descriptors often used in medical image analysis. Most of these descriptors require informative feature points or a segmented binary image as input to the shape extraction procedure. In this paper, we focus on diffeomorphic transformations estimated from full images as a way to represent shape in a group of images [12,14].

The high-dimensional nature of the data (e.g., a 128^3 displacement grid as a shape descriptor for a 3D brain MRI) presents significant challenges for the

statistical methods when extracting relevant latent structure from image transformations. The barriers for effective statistical analysis include (i) requiring greater computational resources and special programming techniques for statistical inference and (ii) numerous local minima. Two main ways of overcoming this problem via data dimensionality reduction have been recently proposed in the diffeomorphic setting. One is to perform statistical modeling of the transformations as a step that follows the estimation of deformations, for instance, by carrying out principal component analysis (PCA) in the tangent space of diffeomorphisms (TPCA) [14]. An empirical shape distribution can be constructed by using TPCA to estimate the intrinsic dimensionality of the diffeomorphic surface variation [12]. Later, a Bayesian model of shape variability was demonstrated to extract the principal modes after estimating a covariance matrix of transformations [7]. Alternatively, one could infer the principal modes of variation and transformations simultaneously. Principal geodesic analysis (PGA) generalized PCA to finite-dimensional manifolds and estimated the geodesic subspaces by minimizing the sum-of-squared geodesic distances to the data [4]. This enabled factor analysis of diffeomorphisms that treats data variability as a joint inference problem in a probabilistic principal geodesic analysis (PPGA) [17]. While these models were designed to find a concise low-dimensional space to represent the data, the estimation must be performed numerically on dense image grids in a high-dimensional space.

In contrast, we use the finite-dimensional representation of the tangent space of diffeomorphisms [18] to investigate shape variability using bandlimited velocity fields as a representation. We call this approach *low-dimensional probabilistic principal geodesic analysis* (LPPGA). We define a low-dimensional probabilistic framework for factor analysis in the context of diffeomorphic atlas building. Our model dramatically reduces the computational cost by employing a low-dimensional parametrization in the Fourier space. Furthermore, we enforce the orthogonality constraints on the principal modes, which is computationally intractable in high-dimensional models like PPGA [17]. We report estimated principal modes in the ADNI brain MRI dataset [8] and compare them with the results of TPCA and PPGA of diffeomorphisms in the full dimensional space. The experimental results show that the low-dimensional statistics encode the features of interest in the data, better capture the group variation and improve data interpretability. Moreover, our model requires much less computational resources.

2   Diffeomorphic Atlas Building with Geodesic Shooting

We first briefly review the mathematical background of diffeomorphic atlas building in the setting of large deformation diffeomorphic metric mapping (LDDMM) [1] with geodesic shooting [15,16].

We let $J_1, \cdots, J_N$ be the N input images that are assumed to be square integrable functions defined on d-dimensional torus domain $\Omega = \mathbb{R}^d / \mathbb{Z}^d$ ($J_n \in L^2(\Omega, \mathbb{R}), n \in \{1, \cdots, N\}$). We use I to denote the atlas template and ϕ_n to

denote the deformation from template I to image J_n. The time-varying deformation $\phi_n(t,x) : t \in [0,1], x \in \Omega$ is defined as the integral flow of time-varying velocity field $v_n(t,x)$ in a reproducing kernel Hilbert space V:

$$d/dt\, \phi_n(t,x) = v_n(t, \phi_n(t,x)).$$

The geodesic path of the deformation is uniquely determined by integrating the Euler-Poincaré differential equation (EPDiff) [18] with an initial condition of $v_n(t,x)$ at $t = 0$:

$$\frac{\partial v_n}{\partial t} = -\mathcal{K}\left[(Dv_n)^T m_n + Dm_n\, v_n + m_n\, \mathrm{div}\,(v_n)\right], \tag{1}$$

where 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 $\mathcal{L} : V \to V^*$ 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 $v_n = \mathcal{K}m_n$. This process is known as *geodesic shooting* [15,16].

With a slight abuse of notation, we define $\phi_n = \phi_n(1, \cdot)$, $v_n = v_n(0, \cdot)$, allowing us to drop the time index in the subsequent derivations. Geodesic shooting (1) enables differentiation of the image differences with respect to the initial velocity field, leading to a gradient decent minimization of the energy function

$$E(v_n, I) = \sum_{n=1}^{N} \frac{1}{2\sigma^2}\left\|J_n - I \circ \phi_n^{-1}\right\|_{L^2}^2 + (\mathcal{L}v_n, v_n), \tag{2}$$

where σ^2 is the image noise variance. In this paper, we use $\mathcal{L} = (-\alpha\Delta + e)^c$, where Δ is the discrete Laplacian operator, e is the identity matrix, c is a positive scalar controlling smoothness, and α is a positive regularity parameter. The notation $(\cdot, \cdot)$ denotes the pairing of a momentum vector with a tangent vector, similar to an inner product.

It has been recently demonstrated that the initial velocity v_n can be efficiently captured via a discrete low-dimensional bandlimited representation in the Fourier space [18]. We adopt this low-dimensional representation for statistical shape analysis.

3   Generative Model

We build our generative model in the discrete finite-dimensional space $\tilde{V}$ that represents bandlimited velocity fields. Elements of this space $\tilde{v} \in \tilde{V}$ are complex-valued vector fields in the Fourier domain that represent conjugate frequencies: $v = \mathcal{F}\tilde{v}$, where $\mathcal{F}$ is the Fourier basis that maps from the frequency domain to the image domain.

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 $(p \ll d)$, $\Lambda \in \mathbb{R}^{q \times q}$ be a diagonal matrix of scale factors for the columns of $\tilde{W}$, and $s \in \mathbb{R}^q$ be a vector that parameterizes the space of transformations. The initial velocity

is therefore represented as $\tilde{v} = \tilde{W}\Lambda s$ in the low-dimensional space. Assuming i.i.d. Gaussian noise on image intensities, we obtain

$$p(J_n \,|\, s_n; \tilde{W}, \Lambda, I, \sigma) = \mathcal{N}(J_n \,; I \circ \phi_n^{-1}, \sigma^2), \tag{3}$$

where ϕ_n is a deformation that corresponds to the initial velocity $v_n = \mathcal{F}\tilde{W}\Lambda s_n$ in the image space, that is, $d/dt\,\phi_n = \mathcal{F}\tilde{W}\Lambda s_n$, and $\mathcal{N}(\cdot\,; \mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2.

The prior on the loading coefficients s_n is the combination of a Gaussian distribution $\mathcal{N}(0, e)$ (e is the identity matrix) with a complex multivariate Gaussian distribution $\mathcal{N}(0, (\tilde{\mathcal{L}}\tilde{W}^T\Lambda^2\tilde{W})^{-1})$ that ensures the smoothness of the geodesic path. Similar to the $\mathcal{L}$ operator, $\tilde{\mathcal{L}} : \tilde{V} \to \tilde{V}^*$ is also a symmetric, positive definite operator that maps a complex tangent vector $\tilde{v} \in \tilde{V}$ in the Fourier domain to its dual momentum vector $\tilde{m} \in \tilde{V}^*$. For a $D_1 \times D_2 \times D_3$ grid, the operator value $\tilde{L}_{d_1 d_2 d_3}$ at location (d_1, d_2, d_3) is given by

$$\tilde{L}_{d_1 d_2 d_3} = \left[-2\alpha \left(\cos \frac{2\pi d_1}{D_1} + \cos \frac{2\pi d_2}{D_2} + \cos \frac{2\pi d_3}{D_3} - 3 \right) + 1 \right]^c,$$

and its inverse is $\tilde{L}_{d_1 d_2 d_3}^{-1} = \tilde{K}_{d_1 d_2 d_3}$. Finally, we formulate the prior as

$$p(s_n \,|\, \tilde{W}, \Lambda) = \mathcal{N}(s_n \,; 0, (\tilde{\mathcal{L}}\tilde{W}^T\Lambda^2\tilde{W})^{-1} + e). \tag{4}$$

We now arrive at the posterior distribution of $s_1, \cdots, s_N$

$$\begin{aligned}
Q &\triangleq \log p(s_1, \cdots, s_N \,|\, J_1, \cdots, J_N; \tilde{W}, \Lambda, I, \sigma^2) \\
&= \sum_{n=1}^{N} \log p(J_n \,|\, s_n; \tilde{W}, \Lambda, I, \sigma) + \log p(s_n \,|\, \tilde{W}, \Lambda) + \text{const.} \\
&= \sum_{n=1}^{N} -\frac{\|J_n - I \circ \phi_n^{-1}\|_{L^2}^2}{2\sigma^2} - \frac{s_n^T(\tilde{\mathcal{L}}\tilde{W}^T\Lambda^2\tilde{W} + e)s_n}{2} - \frac{dN}{2}\log\sigma + \text{const.} \quad (5)
\end{aligned}$$

4 Inference

We use alternating gradient accent to maximize the posterior probability (5) with respect to the model parameters $\theta = \{\tilde{W}, \Lambda, I, \sigma^2\}$ and latent variables $\{s_1, \cdots, s_N\}$.

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:

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

To estimate the principal initial velocity basis $\tilde{W}$, the scaling factor Λ, and the loading coefficients $\{s_n\}$, we follow the derivations in [18] and first obtain the gradient of Q w.r.t. the initial velocity $\tilde{v}_n$ as follows:

(i) Forward integrate the geodesic evolution equation (1) to generate time-dependent diffeomorphic deformation $\phi_n(t, x)$.

(ii) Compute the gradient $\nabla_{\tilde{v}_n} Q$ at time point $t = 1$ as

$$[\nabla_{\tilde{v}_n} Q]_{t=1} = -\tilde{K} \left[\frac{1}{\sigma^2} (J_n - I \circ \phi_n^{-1}) \cdot \nabla(I \circ \phi_n^{-1}) + \tilde{L}\tilde{v}_n \right]. \tag{6}$$

(iii) Backward integrate the gradient (6) to $t = 0$ to obtain $[\nabla_{\tilde{v}_n} Q]_{t=0}$.

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

$$\nabla_{s_n} Q = -\Lambda \tilde{W}^T [\nabla_{\tilde{v}_n} Q]_{t=0} - s_n.$$

The gradients of Q w.r.t. $\tilde{W}, \Lambda$ are given as follows:

$$\nabla_{\tilde{W}} Q = -\sum_{n=1}^{N} [\nabla_{\tilde{v}_n} Q]_{t=0} s_n^T \Lambda, \qquad \nabla_\Lambda Q = -\sum_{n=1}^{N} \tilde{W} s_n^T [\nabla_{\tilde{v}_n} Q]_{t=0}.$$

Unlike the PPGA model [17], we can readily enforce the mutual orthogonality constraint on the columns of $\tilde{W}$. Here, we choose to employ Gram-Schmidt orthogonalization [2] on the column vectors of $\tilde{W}$ in a complex inner product space.

5   Results

Data. To evaluate the effectiveness of the proposed *low-dimensional principal geodesic analysis* (LPPGA), we applied the algorithm to brain MRI scans of 90 subjects from the ADNI [8] study, aged 60 to 90. Fifty subjects have Alzheimer's disease and the remaining 40 subjects are healthy controls. All MRIs 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 skull-stripping, downsampling, intensity normalization, bias field correction, and co-registration with affine transformations.

Experiments. We first estimate a full collection of principal modes $q = 89$ for our model, using $\alpha = 3.0, c = 3.0$ for the operator $\tilde{\mathcal{L}}$ with $p = 16^3$ dimensions of initial velocity $\tilde{v}$, similar to the settings used in pairwise diffeomorphic image registration [18]. 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 initial velocity matrix $\tilde{W}$ to be the principal components estimated by TPCA [14]. We then compare the results with PPGA [17] and TPCA on the same dataset. 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 randomly select 50 images out of 90 and rerun the entire experiment 50 times.

To investigate the ability of our model to capture anatomical variability, we use the loading coefficients $s = \{s_1, \cdots, s_N\}$ as a shape descriptor in a
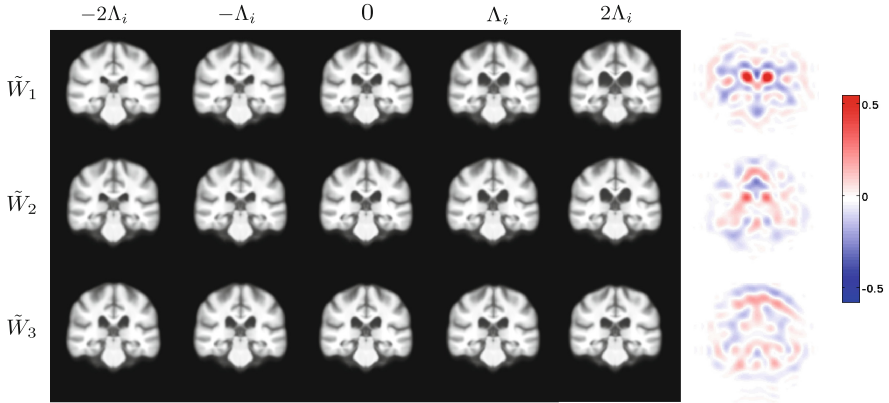
Fig. 1. Top to bottom: first, second and third principal modes of brain shape variation evaluated for varying amounts of the corresponding principal mode, and log determinant of Jacobians at $2\Lambda_i$. Coronal views are shown.

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 mini-mental state examination (MMSE), Alzheimer's Disease Assessment Scale (ADAS), and Clinical Dementia Rating (CDR). We focus on MMSE and fit it to a linear regression model using the loadings for all 90 subjects in the training dataset as predictors. Similar analysis is performed on the results of PPGA and TPCA.

Experimental Results. Figure 1 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 Jacobians at $a_i = 2$ (regions of expansion in red and contraction in blue). The first mode of variation clearly shows that changes in ventricle size is the dominant source of variability in the brain shape. The algorithm estimates standard deviation of the image noise to be $\sigma = 0.02$.

Figure 2 reports the cumulative variance explained by the model as a function of the model size. Our approach achieves higher representation accuracy than the two state-of-the-art baseline algorithms across the entire range of model sizes.

Table 1 compares the regression results of our model and the two baseline algorithms using the first principal mode. The higher F and R^2 statistics indicate that our approach captures more variation of the MMSE score than the other models. Table 1 also reports run time and memory consumption for building the full model of anatomical variability. It demonstrates that our approach offers an order of magnitude improvement in both the run time and memory requirements while providing a more powerful model of variability.
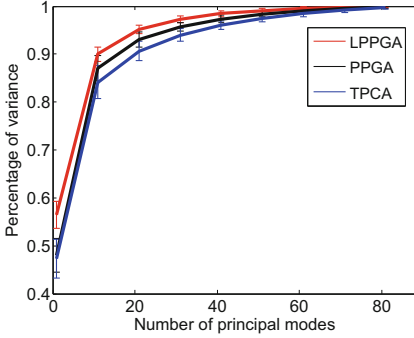
Fig. 2. Left: cumulative variance explained by principal modes estimated through our method (LPPGA) and baseline algorithms (PPGA and TPCA). Right: number of principal modes that explain 90 % and 95 % of total variance respectively.

Table 1. Left: Comparison of linear regression models on the first principal mode for our model (LPPGA) and the baseline algorithms (PPGA and TPCA) on 90 brain MRIs from ADNI. Right: 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.

Model	Residual	R^2	F	p-value
LPPGA	**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	$1.10e^{-4}$

Time (Hours)	Memory (MB)
1.2	**168.4**
17.8	1708.1
16.7	1708.1

6   Conclusion

We presented a low-dimensional probabilistic framework for factor analysis in the space of diffeomorphisms. Our model reduces the computational cost and amplifies the statistical power of shape analysis by using a low-dimensional parametrization. This work represents the first step towards *efficient* probabilistic models of shape variability based on high-dimensional diffeomorphisms. Future work will explore Bayesian variants of shape analysis. A multiscale strategy like [13] can be added to our model to make the inference even faster.

References

1. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. **61**(2), 139–157 (2005)

2. Cheney, W., Kincaid, D.: Linear Algebra: Theory and Applications, vol. 110. The Australian Mathematical Society, Canberra (2009)
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)
4. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. Comput. Vis. Pattern Recogn. **1**, 1–95 (2003). IEEE
5. Gerig, G., Styner, M., Shenton, M.E., Lieberman, J.A.: Shape versus size: improved understanding of the morphology of brain structures. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 24–32. Springer, Heidelberg (2001). doi:10.1007/3-540-45468-3_4
6. Golland, P., Grimson, W.E.L., Shenton, M.E., Kikinis, R.: Small sample size learning for shape analysis of anatomical structures. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 72–82. Springer, Heidelberg (2000). doi:10.1007/978-3-540-40899-4_8
7. Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S.: Bayesian atlas estimation for the variability analysis of shape complexes. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 267–274. Springer, Heidelberg (2013)
8. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging **27**(4), 685–691 (2008)
9. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 316–323. IEEE (2000)
10. Nemmi, F., Sabatini, U., Rascol, O., Péran, P.: Parkinson's disease and local atrophy in subcortical nuclei: insight from shape analysis. Neurobiol. Aging **36**(1), 424–433 (2015)
11. Pizer, S.M., Fritsch, D.S., Yushkevich, P.A., Johnson, V.E., Chaney, E.L.: Segmentation, registration, and measurement of shape variation via image object shape. IEEE Trans. Med. Imaging **18**(10), 851–865 (1999)
12. Qiu, A., Younes, L., Miller, M.I.: Principal component based diffeomorphic surface mapping. IEEE Trans. Med. Imaging **31**(2), 302–311 (2012)
13. Sommer, S., Lauze, F., Nielsen, M., Pennec, X.: Sparse multi-scale diffeomorphic registration: the kernel bundle framework. J. Math. Imaging Vis. **46**(3), 292–308 (2013)
14. Vaillant, M., Miller, M.I., Younes, L., Trouvé, A.: Statistics on diffeomorphisms via tangent space representations. NeuroImage **23**, S161–S169 (2004)
15. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. Int. J. Comput. Vis. **97**(2), 229–241 (2012)
16. Younes, L., Arrate, F., Miller, M.: Evolutions equations in computational anatomy. NeuroImage **45**(1), S40–S50 (2009)
17. Zhang, M., Fletcher, P.T.: Bayesian principal geodesic analysis in diffeomorphic image registration. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8675, pp. 121–128. Springer, Heidelberg (2014). doi:10.1007/978-3-319-10443-0_16
18. Zhang, M., Fletcher, P.T.: Finite-dimensional Lie algebras for fast diffeomorphic image registration. In: Ourselin, S., Alexander, D.C., Westin, C.-F., Cardoso, M.J. (eds.) IPMI 2015. LNCS, vol. 9123, pp. 249–260. Springer, Heidelberg (2015)