Bayesian Principal Geodesic Analysis in Diffeomorphic Image Registration

Miaomiao Zhang and P. Thomas Fletcher

School of Computing, University of Utah, Salt Lake City, USA

Abstract. Computing a concise representation of the anatomical variability found in large sets of images is an important first step in many statistical shape analyses. In this paper, we present a generative Bayesian approach for automatic dimensionality reduction of shape variability represented through diffeomorphic mappings. To achieve this, we develop a latent variable model for principal geodesic analysis (PGA) that provides a probabilistic framework for factor analysis on diffeomorphisms. Our key contribution is a Bayesian inference procedure for model parameter estimation and simultaneous detection of the effective dimensionality of the latent space. We evaluate our proposed model for atlas and principal geodesic estimation on the OASIS brain database of magnetic resonance images. We show that the automatically selected latent dimensions from our model are able to reconstruct unseen brain images with lower error than equivalent linear principal components analysis (LPCA) models in the image space, and it also outperforms tangent space PCA (TPCA) models in the diffeomorphism setting.

1 Introduction

Diffeomorphic image registration plays an important role in understanding anatomical shape variability in medical image analysis. For example, analysis of diffeomorphic shape changes can be linked to disease processes and changes in cognitive and behavioral measures. In this setting, the high dimensionality of the deformations, combined with the relatively small sample sizes available, make statistical analysis challenging. However, the intrinsic dimensionality of brain shape variability is much lower. Extracting these intrinsic dimensions before further statistical analysis can improve the statistical power and interpretability of results.

Motivated by Bayesian reasoning, current approaches in diffeomorphic atlas building [6,12,14] are formulated as *maximum a posteriori* (MAP) optimization problems. A set of input images are registered to a template, which is simultaneously estimated with the unknown deformations in an alternating optimization strategy. In these approaches, the likelihood is defined by an image match term which is a sum squared distance function between deformed atlas and input image, and a prior on transformations that enforces smoothness. Allassonnière et al. [1] proposed a fully generative Bayesian model of elastic deformation in which estimation proceeds by marginalization over the latent image transformations.

P. Golland et al. (Eds.): MICCAI 2014, Part III, LNCS 8675, pp. 121–128, 2014.

Ma et al. [7] introduced a Bayesian formulation of the diffeomorphic image atlas problem by adding fixed hypertemplate information. Simpson et al. [10] inferred the level of regularization in small deformation registration by a hierarchical Bayesian model. Zhang et al. [18] develop a generative model for diffeomorphic atlas formulation and regularization parameter estimation by using a Monte Carlo Expectation Maximization (MCEM) algorithm.

Beyond estimation of an atlas, or mean image, several dimensionality reduction methods have been proposed for modeling shape variability in the diffeomorphism setting. Vaillant et al. [13] compute a PCA in the tangent space to the atlas image. Later, Qiu et al. [9] used TPCA as an empirical shape prior in diffeomorphic surface matching. Gori et al. [5] formulate a Bayesian model of shape variability using diffeomorphic matching of currents. Their model includes estimation of a covariance matrix of the deformations, from which they then extract PCA modes of shape variability. Even though these methods formulate the atlas and covariance estimation as probabilistic inference problems, the dimensionality reduction is done after the fact, i.e., as a singular value decomposition of the covariance as a second stage after the estimation step. We propose instead to treat the dimensionality reduction step as a probabilistic inference problem on discrete images, in a model called Bayesian principal geodesic analysis (BPGA), which jointly estimates the image atlas and principal geodesic modes of variation. This Bayesian formulation has two advantages. First, computing a PCA after the fact in the tangent space does not explicitly optimize the fit of the principal modes to the data (this is due to the nonlinearity of the space of diffeomorphisms), whereas we explicitly optimize this criteria intrinsically in the space of diffeomorphisms, resulting in better fits to the data. Second, by formulating dimensionality reduction as a Bayesian model, we can also infer the inherent dimensionality directly from the data.

Our work is inspired by the Bayesian PCA model introduced in Euclidean space by Bishop (BPCA) [2]. Recently, Zhang and Fletcher [17] introduced a probabilistic principal geodesic analysis (PPGA) to finite-dimensional manifolds based on PGA [3]. This work goes beyond the PPGA model by introducing the automatic dimensionality reduction, as well as extending from finite-dimensional manifolds to the infinite-dimensional case of diffeomorphic image registration. We also mention the relationship of our work to manifold learning approaches to dimensionality reduction [4]. The main advantage of the Bayesian approach we present is that it is fully *generative*, and the principal modes of variation can reconstruct shape deformation of individuals, information that is lost when mapping to a Euclidean parameter space in manifold learning. We show experimental results of principal geodesics and parameters estimated from OASIS brain dataset. To validate the advantages of our model, we reconstruct images from our estimation and compare the reconstruction errors with TPCA of diffeomorphisms and LPCA based on image intensity. Our results indicate that intrinsic modeling of the principal geodesics, estimated jointly with the image atlas, provides a better description of brain image data than computing PCA in the tangent space after atlas estimation.

2 Background

We define a generative probabilistic model for principal geodesic analysis in the setting of diffeomorphic atlas building. Before introducing our model, we first briefly review the mathematical background of diffeomorphic atlas building and its computations for geodesic shooting [11,15,16]. We use vector-valued momenta [11], which unlike scalar momenta, decouple the deformations from the atlas, leading to more efficient and stable estimation procedures.

In this framework, given input images $I_1, \ldots, I_N \in L^2(\Omega, \mathbb{R})$, a minimization problem is solved to estimate the template image and the diffeomorphic transformations between the template and each input image as

$$E(v^k, I) = \sum_{k=1}^{N} \frac{1}{2\sigma^2} \left\| I \circ (\phi_k)^{-1} - I_k \right\|^2 + \int_0^1 (Lv_t^k, v_t^k)\, dt. \tag{1}$$

Here σ^2 represents noise variance, and the $v^k \in L^2([0,1], V)$ are time-varying velocity fields in a reproducing kernel Hilbert space, V, equipped with a metric, $L : V \to V^*$, a positive-definite, self-adjoint, differential operator, mapping to the dual space, V^*. The dual to the vector v^k is a momentum, $m^k \in V^*$, such that $m^k = Lv^k$ and $v^k = Km^k$. The operator K is the inverse of L. The notation (m^k, v^k) denotes the pairing of a momentum vector $m^k \in V^*$ with a tangent vector $v^k \in V$. The deformation ϕ_k is defined as the integral flow of v^k, that is, $(d/dt)\phi_k(t,x) = v^k(t, \phi_k(t,x))$. We use subscripts for the time variable, i.e., $v_t(x) = v(t,x)$, and $\phi_t(x) = \phi(t,x)$. When the energy above is minimized over the initial momenta m^k, the geodesic path ϕ_k is constructed via integration of the following EPDiff equation [8]:

$$\frac{\partial m^k}{\partial t} = -\mathrm{ad}^*_{v^k} m^k = -(Dv^k)^T m^k - Dm^k\, v^k - m^k \operatorname{div}(v^k), \tag{2}$$

where D denotes the Jacobian matrix, and the operator ad^* is the dual of the negative Lie bracket of vector fields, $\mathrm{ad}_v w = -[v, w] = Dvw - Dwv$.

3 Bayesian Principal Geodesic Analysis

3.1 Probability Model

We formulate the random momentum for the kth individual as $m^k = Wx^k$, where W is a matrix with q columns of principal initial momenta, $x \in \mathbb{R}^q$ is a latent variable that lies in a low-dimensional space, with $x \sim N(0, I)$. Our noise model is i.i.d. Gaussian at each image voxel, with likelihood given by

$$p(I_k \mid W, I, \sigma) = \frac{1}{(2\pi)^{M/2}\sigma^M} \exp\left(-\frac{\|I \circ (\phi_k)^{-1} - I_k\|^2}{2\sigma^2} \right), \tag{3}$$

where M is the number of voxels, and the norm inside the exponent is the $L^2(\Omega, \mathbb{R})$ norm. Note that for a continuous domain, this is not a well-defined probability distribution due to its infinite measure on images. Therefore, we consider the input images as well as diffeomorphisms to be defined on a finite discretized grid.

The prior on W is given by the combination of a multivariate Gaussian distribution on the initial momenta m that guarantees smoothness of the geodesic shooting path, and a Gaussian distribution on W to suppress small principal initial momenta to zero. This second term is analogous to the automatic relevance determination (ARD) prior used in BPCA [2], with the difference that we use the natural Hilbert space norm for the momenta. This prior induces sparsity in the columns of W and automatically selects the dimensionality. The formulation is given by

$$p(W \mid x, \gamma) \propto \left(\prod_{i=1}^{q} \left(\frac{\gamma_i}{2\pi} \right)^{\frac{d}{2}} \right) \exp\left(-\frac{1}{2} \sum_{k=1}^{N} \|m^k\|_K^2 - \sum_{i=1}^{q} \frac{\gamma_i}{2} \|W_i\|_K^2 \right), \quad (4)$$

where i denotes the ith principal initial momentum, and γ_i is a hyperparameter which controls the precision of the corresponding W_i. Estimating γ_i induces sparsity so that if it has a large value, then the corresponding W_i will become small, and will be effectively removed in the latent space. In this work, we use a metric of the form $K = (-\alpha\Delta + \beta I)^{-2}$, where Δ is the discrete Laplacian. In this operator, α controls the smoothness of diffeomorphisms, and β is a small positive number to ensure that the K operator is nonsingular.

3.2 Inference

After defining the likelihood (3) and prior (4) in the previous section, we now arrive at the log joint posterior for the diffeomorphisms as

$$\begin{aligned} \log \prod_{k=1}^{N} p\left(W \mid I_k; I, \sigma^2, \gamma\right) \propto & -\frac{1}{2} \sum_{k=1}^{N} \|m^k\|_K^2 - \frac{1}{2\sigma^2} \sum_{k=1}^{N} \|I \circ (\phi_k)^{-1} - I_k\|^2 \\ & - \frac{MN}{2} \log \sigma - \sum_{i=1}^{q} \left[\frac{\gamma_i}{2} \|W_i\|_K^2 - \log \gamma_i \right] - \frac{1}{2} \|x^k\|^2. \end{aligned} \quad (5)$$

We use MAP estimation to determine the model parameters $\theta = I, \sigma$. In order to treat the x^k as latent random variables with log posterior given by (5), we would ideally integrate out the latent variables, which is intractable in closed form. Instead, we use a mode approximation to the posterior distribution. Next, we introduce a gradient ascent scheme to estimate W, x^k, $\theta = (I, \sigma)$ and γ simultaneously.

Gradient Ascent for W, x^k: We need to compute the gradient with respect to initial momentum m^k of the diffeomorphic image matching problem in (1), and then apply the chain rule to obtain the gradient term w.r.t. W and x^k. Following the optimal control theory approach in [15], we add Lagrange multipliers to constrain the kth diffeomorphism $\phi_k(t)$ to be a geodesic path. The following equations are equivalent for the geodesic paths of each of the subjects, so for notational simplicity, we drop the subject index k from the notation momentarily. This is done by introducing time-dependent adjoint variables, $\hat{m}, \hat{I}$ and $\hat{v}$, and writing the augmented energy

$$\tilde{E}(m) = E(Km, I, I_k) + \int_0^1 \left[\langle \hat{m}, \dot{m} + \mathrm{ad}_v^* m \rangle + \langle \hat{I}, \dot{I} + \nabla I \cdot v \rangle + \langle \hat{v}, m - Lv \rangle \right] dt,$$

where E is the diffeomorphic image matching energy from (1), and the other terms correspond to Lagrange multipliers enforcing: a) the geodesic constraint, which comes from the EPDiff equation (2), b) the image transport equation, $\dot{I} = -\nabla I \cdot v$, and c) the constraint that $m = Lv$, respectively.

The optimality conditions for m, I, v are given by the following time-dependent system of ODEs, termed the *adjoint equations*:

$$-\dot{\hat{m}} + \mathrm{ad}_v \hat{m} + \hat{v} = 0, \qquad -\dot{\hat{I}} - \nabla \cdot (\hat{I} v) = 0, \qquad -\mathrm{ad}_{\hat{m}}^* m + \hat{I} \nabla I - L\hat{v} = 0,$$

subject to initial conditions $\hat{m}(1) = 0, \hat{I}(1) = \frac{1}{\sigma^2}(I(1) - I_k)$. Finally, after integrating these adjoint equations backwards in time to $t = 0$, the gradient of $\tilde{E}$ with respect to the initial momentum is $\nabla_m \tilde{E} = Km - \hat{m}$.

Applying the chain rule, the gradient term of (5) for updating W is

$$\nabla_W \tilde{E} = -\sum_{k=1}^{N} (Km^k - \hat{m^k})[x^k]^T - KW\gamma,$$

where γ is a diagonal matrix with diagonal element γ_i. The gradient with respect to x^k is

$$\nabla_{x^k} \tilde{E} = -W^T (Km^k - \hat{m^k}) - x^k.$$

Closed-Form Solution for θ, γ: We now derive the maximization for updating the parameters θ. This turns out to be a closed-form update for the atlas I, noise variance σ^2, and dimensionality control parameter γ. For updating the atlas image I, we set the derivative of the log posterior with respect to I to zero. The solution for I, σ^2 gives an update

$$I = \frac{\sum_{k=1}^N I_k \circ \phi_k |D\phi_k|}{\sum_{k=1}^N |D\phi_k|}, \qquad \sigma^2 = \frac{1}{MN} \sum_{k=1}^{N} \|I \circ (\phi_k)^{-1} - I_k\|^2.$$

We use the similar approximation on ARD prior in BPCA [2] to get a closed-form update for γ_i, as $\gamma_i = q / \|W_i\|_K^2$.

4 Results

4.1 OASIS Brain Dataset

To demonstrate the effectiveness of our proposed model and MAP estimation, we applied our BPGA model to a set of axial slices of brain magnetic resonance

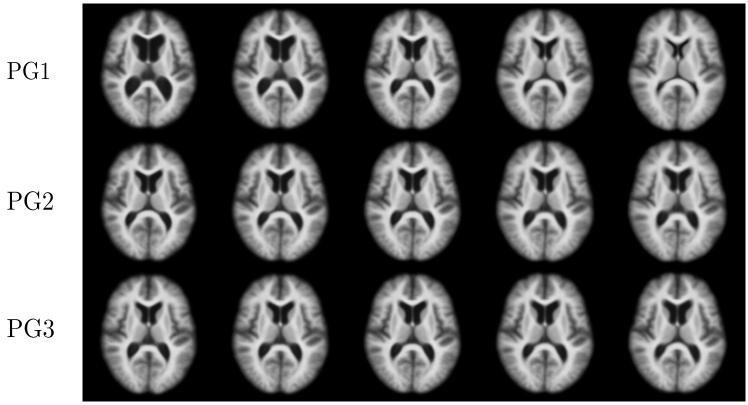
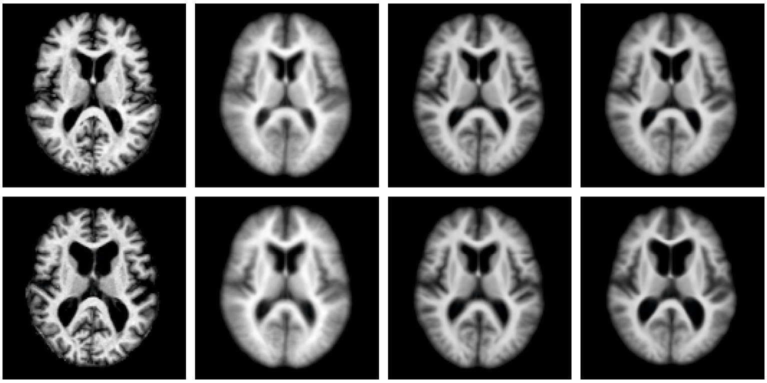


Fig. 1. Top to bottom: shooting atlas by the first, second and third principal modes. Left to right: BPGA model of image variation evaluated at $a_i = -3, -1.5, 0, 1.5, 3$.

images (MRI) from the OASIS brain database. The data consists of MRI from 40 healthy subjects between the age of 60 to 95. The MRI have resolution $108 \times 128 \times 128$ and are skull-stripped, intensity normalized, and co-registered with rigid transforms. We use $\alpha = 0.8$, $\beta = 0.4$ estimated using the procedure in [18] with 15 time-steps in geodesic shooting, and initialize the template I as the average of image intensities, while W as the matrix of principal components from TPCA.

The proposed BPGA model automatically determined that the latent dimensionality of the data was three. Figure 1 displays the automatic estimated modes, $i = 1, 2, 3$, of the brain MRI variation. We forward shoot the constructed atlas, I, by the estimated principal momentum $a_i W_i$ along geodesics. For visualization purpose, here we demonstrate the brain variation from the atlas by $a_i = -3, -1.5, 0, 1.5, 3$. The first mode of variation clearly shows that ventricle size change is a dominant source variability in brain shape. The algorithm also jointly estimated the image noise standard deviation parameter as $\sigma = 0.04$.

Image Registration Accuracy. We validated the image registration accuracy of our BPGA model. After estimating the principal initial momenta and parameters from the training subjects above, we used these estimates to reconstruct another 20 testing subjects from the same OASIS database that were not included in the training. We then measured the discrepancy between the reconstructed images and the testing images. Note that our reconstruction only used the first three principal modes, which were automatically selected by our algorithm. We also compared our model with LPCA and TPCA, also using the first three dimensions. Examples of the reconstructed images from these models are shown in Figure 2. Table 1 shows the comparison of the registration accuracy as measured by the average and standard deviation of the mean squared error (MSE). It indicates that our model outperforms both LPCA and TPCA in the diffeomorphic setting.



(a) Observed (b) LPCA (c) TPCA (d) BPGA

Fig. 2. Left to right: original data, reconstruction by LPCA, TPCA, and BPGA

Table 1. Comparison of mean squared reconstruction error between LPCA, TPCA and BPGA models. Average and standard deviation over 20 test images.

	LPCA	TPCA	BPGA
Average MSE	2.8×10^{-2}	1.6×10^{-2}	1.1×10^{-2}
Std of MSE	7.5×10^{-3}	2.3×10^{-3}	2.0×10^{-3}

5 Conclusion

We presented a generative Bayesian model of principal geodesic analysis in diffeomorphic image registration. Our method is the first probabilistic model for automatic dimensionality reduction for diffeomorphisms. We developed an inference strategy based on MAP to estimate parameters, including the noise variance and image atlas, simultaneously. The estimated low-dimensional latent variables provide a compact representation of the anatomical variability in a large image database, and they can be used for further statistical analysis of anatomical shape in clinical studies. Reducing the dimensionality to the inherent modes of shape variability has the potential to improve hypothesis testing, classification, mixture models, etc.

Acknowledgments. This work was supported by NIH Grant 5R01EB007688 and NSF CAREER Grant 1054057.

References

1. Allassonnière, S., Kuhn, E.: Stochastic algorithm for parameter estimation for dense deformable template mixture model. ESAIM-PS 14, 382–408 (2010)
2. Bishop, C.M.: Bayesian pca. In: Advances in Neural Information Processing Systems, pp. 382–388 (1999)

3. Fletcher, P.T., Lu, C., Joshi, S.: Statistics of shape via principal geodesic analysis on Lie groups. In: Computer Vision and Pattern Recognition, pp. 95–101 (2003)
4. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R.: Manifold modeling for brain population analysis. Medical Image Analysis 14(5), 643–653 (2010)
5. Gori, P., et al.: 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)
6. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23(suppl. 1), 151–160 (2004)
7. Ma, J., Miller, M.I., Trouvé, A., Younes, L.: Bayesian template estimation in computational anatomy. NeuroImage 42, 252–261 (2008)
8. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (2006)
9. Qiu, A., Younes, L., Miller, M.I.: Principal component based diffeomorphic surface mapping. IEEE Transactions on Medical Imaging 31(2), 302–311 (2012)
10. Simpson, I.J.A., Schnabel, J.A., Groves, A.R., Andersson, J.L.R., Woolrich, M.W.: Probabilistic inference of regularisation in non-rigid registration. NeuroImage 59, 2438–2451 (2012)
11. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In: International Symposium on Biomedical Imaging, pp. 1219–1222 (2013)
12. Twining, C.J., Cootes, T.F., Marsland, S., Petrovic, V., Schestowitz, R., Taylor, C.J.: A unified information-theoretic approach to groupwise non-rigid registration and model building. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 1–14. Springer, Heidelberg (2005)
13. Vaillant, M., Miller, M.I., Younes, L., Trouvé, A.: Statistics on diffeomorphisms via tangent space representations. NeuroImage 23, S161–S169 (2004)
14. Vialard, F.X., Risser, L., Holm, D., Rueckert, D.: Diffeomorphic atlas estimation using Kärcher mean and geodesic shooting on volumetric images. In: MIUA (2011)
15. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision, 229–241 (2012)
16. Younes, L., Arrate, F., Miller, M.: Evolutions equations in computational anatomy. NeuroImage, 40–50 (2009)
17. Zhang, M., Fletcher, P.T.: Probabilistic principal geodesic analysis. In: Advances in Neural Information Processing Systems, pp. 1178–1186 (2013)
18. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 37–48. Springer, Heidelberg (2013)