Geodesic Refinement Using James-Stein Estimators

Greg M. Fleishman[1,2], P. Thomas Fletcher[3], Boris A. Gutman[2], Gautam Prasad[2], Yingnian Wu[4], and Paul M. Thompson[2]

[1] UC Los Angeles, Department of Bioengineering
[2] Imaging Genetics Center, University of Southern California
[3] Scientific Computing and Imaging Institute, University of Utah
[4] UC Los Angeles, Department of Statistics

Abstract. In longitudinal imaging studies, geodesic regression in the space of diffeomorphisms [9] can be used to fit a generative model to images over time. The parameters of the model, primarily its initial direction or momentum, are important objects for study that contain biologically meaningful information about the dynamics occurring in the underlying anatomy. Unfortunately, it is common for any given subject to have a very limited number of longitudinal images available, the acquisition of which is corrupted by noise and variability due to scanning conditions. Furthermore, the underlying anatomy is subject to many entangled biological processes, the effect of which on images is in many cases poorly characterized. Hence, the approach must fit a model to few data points with uncharacterized variability. Here, we propose supplementing the lack of longitudinal information for an individual patient with information that can be extracted cross-sectionally from a population of time series to improve the model fit for the individual. To that effect, we propose a probabilistic model that leads to a well established technique from classical statistics: James-Stein estimators. We show that recent work on groupwise registration for improved geodesic estimation is a sub-optimal special case of our proposed model. Finally, we validate the model by showing geodesics refined by the James-Stein estimator extrapolate more accurately on average than raw geodesic estimates.

1 Introduction

In the large deformation diffeomorphic metric mapping (LDDMM) framework for nonlinear image registration [1], interpolation and extrapolation of longitudinal image time series can be accomplished with geodesic regression [9]. In this setting, a geodesic on a manifold of diffeomorphisms is estimated such that it passes maximally close to transformations that optimally map the initial image to all subsequent images in the time series. The geodesic is parameterized by an initial transformation (here fixed at the identity for simplicity) and a single vector field (tangent to the manifold at the identity), which specifies the direction of the geodesic. If one assumes this vector field is everywhere proportional

to the initial image gradient [8], then the geodesic is fully specified by a single scalar-valued image, henceforth referred to as the momentum. The task of geodesic regression can then be formulated as: given the time series of images $I_1(x), ..., I_N(x)$, find the momentum $p(x)$ such that the geodesic parameterized by $p(x)$ passes through $\phi_2(x), ..., \phi_N(x)$ and $\sum_{i=2}^{N} d(I_1 \circ \phi_i, I_i)^2$ is minimal; where $d(I, J)$ is some quantitative assessment of similarity between images I and J.

As in any learning task, our confidence in the ability of the geodesic model to make accurate predictions at unobserved time points increases with the number of observations. Unfortunately however, due to the high cost of collecting anatomical images, many longitudinal studies of brain structure collect images at fewer than 5 time-points per individual, and often at relatively small time intervals. The short time intervals are particularly problematic considering the slow dynamics of many neurodegenerative diseases. Such a small number of observations, prone to noise, over a short time interval may be insufficient to fit a geodesic that generalizes to unobserved time points with an acceptable level of confidence. We address the challenge of improving geodesic model generalization for an individual time series by pooling information from multiple time series cross-sectionally, and using it to regularize the individual geodesic models. Such an approach may have practical implications on study design, wherein a researcher may choose to acquire fewer images over a shorter time period from more individuals, and yet achieve similar confidence in the accuracy of individual geodesics had they collected more images over a longer period of time from fewer individuals.

We find a natural mathematical setting to implement this in the James-Stein estimator. The James-Stein estimator is a classical statistical model that improves upon the maximum-likelihood estimate for the mean of a Gaussian random variable. That is, the James-Stein estimator is closer in Euclidean distance on average to the unobserved ground truth value of the mean than its maximum-likelihood estimate. James-Stein estimators are commonly used for massively parallel data sets where the same inference must be made for many samples. Information is pooled across the samples and used to regularize the inference of each individual sample. This model reflects the case in neuroimaging where only short sparsely sampled time series are available but for many patients. Using James-Stein estimates as opposed to maximum-likelihood estimates can offer substantial improvements on model accuracy on average [3]. We utilize the James-Stein estimator to leverage the information contained cross-sectionally in a population of time series to improve the geodesic fit for each individual time series.

A necessary first step for James-Stein estimators is to estimate a groupwise representation of the samples. Several recent works have proposed methods for constructing a groupwise representation of image time series data, any of which is compatible with our proposal. In [2] the authors proposed a method to register time series of images in both space and time simultaneously; a groupwise representation of the time series, or spatiotemporal atlas, can then be found in the common spatiotemporal coordinate system. In [10] the authors propose a

hierarchical geodesic model in which individual geodesics are estimated, then used to construct a groupwise geodesic. Their proposed probabilistic model allows an extension that is not fully explored in [10], which is to re-estimate the individual geodesics after the groupwise representation has been constructed. If the groupwise representation is used as a prior (which is suggested by the probabilistic model), the new estimates are similar to the James-Stein estimates for the individual trajectories. The James-Stein estimator shows how to do this second inference optimally.

After a groupwise representation is obtained, James-Stein estimators shrink individual estimates toward the groupwise representation. We show below that this is in fact a maximum *a posteriori* (MAP) estimate, where the shape of the prior distribution is inferred from the data itself. This can also be viewed as a groupwise consistency constraint. Other recent works have proposed groupwise consistency to cope with difficulty in estimation of individual models. In [12], the authors propose a hierarchical Markov random field (hMRF) for segmentation of structural MRI images into functional networks based on fMRI time series. Individual segmentations are constrained to be smooth and consistent with the fMRI data for that individual. They are also constrained to be similar to a grouplevel representation of the network which is jointly estimated with the individual networks. The authors show that this cross-sectional constraint improves the recovery of networks in ficticious data and results in smoother networks with more anatomical meaning in real data.

Similarly, in [4], pairs of longitudinal brain images from a population of individuals diagnosed with Alzheimer's disease (AD) were registered simultaneously. The optimal set of transformations was defined not only to map the template images to their references, but also to satisfy a groupwise consistency constraint. The authors showed that the resulting geodesics predicted a third time point image not used in the learning step more accurately on average than geodesics learned without the groupwise consistency constraint. We demonstrate below that their approach is in fact a special case of James-Stein estimators. Establishing the connection with James-Stein estimators grounds that work in a probabilistic model from classical statistics that provides better intuition for the meaning of parameters and how to find their optimal values.

2 Methods

2.1 Derivation of the multivariate James-Stein estimator for momenta: For simplicity, we consider time series with two images. Because the derivation of James-Stein estimators will deal exclusively with momenta, the generalization to time series of arbitrary length is trivial. Let I_i and J_i for $i \in \{1, ..., N\}$ be initial and follow up image acquisitions of the same anatomy for N patients. In order to share information cross sectionally we must have a common coordinate system. So we also assume we're given transformations ψ_i such that $I_i(\psi_i) \sim I_j(\psi_j)$ for all i and j. This can be accomplished by finding a study specific atlas, or minimal deformation template (MDT), for the images I_i. All further formula are assumed

to be in the common coordinate system. (That is, all momenta have been moved to the common coordinate system by co-adjoint transport, which for the scalar field p_i is $D\psi_i \cdot p_i(\psi_i)$, where D is the Jacobian operator.)

Now, suppose p_i specifies a geodesic beginning at identity and passing through an optimal ϕ_i such that $I_i(\phi_i) \sim J_i$ for all i. The true values of the p_i are unknown, but let β_i be a noisy estimate of p_i acquired via geodesic regression. Now, suppose the following probabilistic model:

$$p_i \sim \mathcal{N}(p^*, A), \tag{1}$$

$$\beta_i | p_i \sim \mathcal{N}(p_i, \sigma_0^2 \cdot \mathrm{Id}). \tag{2}$$

Equation (1) indicates the unobservable p_i are independent samples from a normal distribution with mean p^* and covariance A. This distribution models the variability in time series trajectory across individuals due to differing contributions of the underlying processes that affect the dynamics of aging and disease. The mean momentum parameterizes a geodesic representing the average dynamics over time for images in the population. (Hence, any one of the previously discussed methods for construction of a groupwise representation of time series [2, 10] can be taken as a definition for p^*.)

Equation (2) indicates the observable β_i are independent samples from a normal distribution with mean p_i and covariance $\sigma_0^2 \cdot \mathrm{Id}$, where Id is the matrix identity of the appropriate size. This distribution models the variability of the momentum measurement β_i due to image noise and registration inaccuracies. Hence, each β_i is distributed about its (unobserved) ground truth value of p_i with isotropic variability, the extent of which is given by σ_0^2. This is consistent with standard noise assumptions in much of the image registration literature.

These distributions have the form of a prior and likelihood, which enables us to write the posterior distribution for the p_i:

$$P(p_i|\beta_i) = \frac{P(\beta_i|p_i)P(p_i)}{\int P(\beta_i|p_i)P(p_i)dp_i} = \mathcal{N}(\beta_i - \sigma_0^2 B(\beta_i - p^*),\ \sigma_0^2 B), \tag{3}$$

where $B = (A + \sigma_0^2 \cdot \mathrm{Id})^{-1}$. We see from (3) that the MAP estimate of p_i is:

$$p_i^{map} = \beta_i - \sigma_0^2 B(\beta_i - p^*). \tag{4}$$

Equation (4) reveals what we gain by incorporating (1) as a prior to regularize β_i. We see that p_i^{map} is equal to the measurement β_i minus an adjustment: $\sigma_0^2 B(\beta_i - p^*)$. The adjustment is a linear transformation of the difference vector $\beta_i - p^*$. If that transformation were the identity, this would simply move β_i toward p^*. However, the linear transformation is actually the covariance matrix of the posterior distribution: $\sigma_0^2 B$. Hence, (4) begins with the idea of moving β_i toward p^*, but takes into account the shapes of the prior and likelihood distributions and adjusts the direction in which we move the estimate accordingly. The net affect is the rearrangment of the observations β_i such that the scatter of the p_i^{map} is more consistent with the prior covariance structure A. Assuming the prior (1)

is correct, p_i^{map} is guaranteed to be a better estimate of p_i on average than the original measurement β_i [7, 3].

Unfortunately, we cannot use (4) directly, as σ_0^2, p^* and A are unknown. However, with N independent parallel time series at our disposal, we can estimate them directly from the data. First we observe the marginal distribution for β_i:

$$P(\beta_i) = \int P(\beta_i|p_i)P(p_i)dp_i = \mathcal{N}(p^*,\ A + \sigma_0^2 \cdot \mathrm{Id}) \tag{5}$$

The maximum likelihood estimate for the mean of a Gaussian random variable is the sample mean. Hence, the maximum likelihood estimate for p^* is simply: $p^* \sim \hat{\beta} = \frac{1}{N}\sum_{i=1}^{N}\beta_i$. Next, we define the sample covariance matrix for the β_i as: $S = \sum_{i=1}^{N}(\beta_i - \hat{\beta})(\beta_i - \hat{\beta})^T$. Because β_i is a random variable, so too is S; which hence must have a corresponding distribution. In fact, the sample covariance matrix of a multivariate normal random variable (such as β_i) is distributed by the Wishart distribution, a multivariate analog of the χ^2 distribution. We now observe:

$$E\{(N-d-1)\sigma_0^2 S^{-1}\} = \sigma_0^2 B \tag{6}$$

where d is the dimensionality of β_i and the expectation is taken with respect to the Wishart distribution. From (6) then, we see that $(N-d-1)\sigma_0^2 S^{-1}$ is the maximum likelihood estimate for $\sigma_0^2 B$. Combining this with $\hat{\beta}$ (the maximum likelihood estimate for p^*) and equation (4) we arrive at the James-Stein estimator for image time series momenta:

$$p_i^{js} = \beta_i - (N-d-1)\sigma_0^2 S^{-1}(\beta_i - \hat{\beta}). \tag{7}$$

The final ingredient is to estimate σ_0^2. Recall, in this model σ_0^2 does not model any biological variability, which is entirely captured by the prior covariance A in (1). σ_0^2 is the noise in the β_i estimates exclusively due to image noise and registration inaccuracy. Hence, any method for estimating the variability due to noise and registration inaccuracy can be used to estimate σ_0^2.

We note here that if we let d be the number of image voxels (the naive dimensionality of β_i), it is almost certain for image analysis applications that $d >> N$, which is generally prohibited if equation (8) is to be useful. Furthermore if $d >> N$, S is certain to be singular and therefore the estimation of S^{-1} becomes problematic. This is the crux issue to be dealt with if one wants to use p_i^{js} for the proposed application. Below, we make the simplest (and least informative) assumption to contend with this issue and then discuss alternatives that might improve the framework.

2.2 Connection to groupwise registration with similarity constraint: To incorporate cross sectional information into the registration of a population of N time series, recent works [4] proposed an objective function of the form:

$$\alpha\mathcal{P}[\phi_1, ..., \phi_N] + \sum_{i=1}^{N}\mathcal{D}[J_i, I_i[\phi_i]] + \gamma\mathcal{S}[\phi_i] = \min \tag{8}$$

Here, the typical image similarity term $\mathcal{D}$ and smoothing prior $\mathcal{S}$ are summed over the N pairs of images. The objective is augmented by a new term $\mathcal{P}$ that is a function of the full set of N transformations, or in the diffeomorphic case, of the estimated transformation momenta in MDT coordinates β_i. Specifically, for $\mathcal{P}$ those works proposed:

$$\mathcal{P}[\beta_1, ..., \beta_N] = \sum_{i=1}^{N} \|\beta_i - \hat{\beta}\|^2 \tag{9}$$

which is the sum of squared difference of the N momenta from their sample average. The Euler-Lagrange equations for this term are: $\nabla_{\beta_i}\mathcal{P}[\beta_1, ..., \beta_N] = 2\alpha(\beta_i - \hat{\beta})$ such that at every iteration the estimate for β_i is updated according to:

$$\beta_i^{t+1} = \beta_i^t - 2\alpha(\beta_i - \hat{\beta}) - \nabla_{\beta_i}\mathcal{D} - \nabla_{\beta_i}\mathcal{S} \tag{10}$$

The first two terms in equation (10) are very similar to equation (4). In fact, if B in (4) were proportional to the identity matrix then the first two terms in (10) would be identical to (4): a shrinkage of the estimate β_i directly toward $\hat{\beta}$ proportional to some scalar value. B is proportional to the identity if and only if A in (1) is proportional to the identity. This reveals two things: the simultaneous registration with groupwise consistency is equivalent to using p_i^{js} with an isotropic prior distribution instead of β_i at every iteration, and that the parameter α in (10) is a function of A and σ_0^2. The perspective of James-Stein estimators thus enables us to generalize the groupwise consistency to anisotropic prior structures and provides an interpretation of the groupwise consistency parameter α.

3 Experiments and Results

3.1 Images: We downloaded screening, 1 year follow up, and 2 year follow up 1.5 Tesla T1-weighted images for 57 participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI). All 57 participants had been diagnosed with Alzheimer's Disease (AD) prior to the acquisition of their screening image. The population consisted of 32 males mean age 75.91 +/- 7.85 years and 25 females mean age 75.08 +/- 8.15 years. This was the maximum number of individuals we could download from the ADNI 1 cohort that were in the AD group and had screening, year 1, and year 2 follow up images available. All images were corrected for geometric distortion and bias in the static field with GradWarp and N3 before downloading as part of the ADNI preprocessing protocol. Subsequent to downloading, the images were linearly registered to the ICBM template and skull stripped using ROBEX [6]. Transformations ψ^i mapping the template images I_i into a MDT coordinate system were computed using a preexisting implementation of [13]. We then registered the screening (I_i) to the follow up images (J_i) to acquire the β_i using our own implementation of the geodesic

shooting algorithm proposed in [11].

3.2 Experimental design: The multivariate James-Stein estimator, equation (7), presents some computational challenges for image data. The full image resolution for most image data sets (a total of d voxels) is very large. Hence S and S^{-1} may be computationally intractable to compute or store. The easiest way to avoid this problem is to assume A and thus S and S^{-1} are proportional to the identity. In that case, the coefficient in front of the second term in (7) reduces to a scalar value:

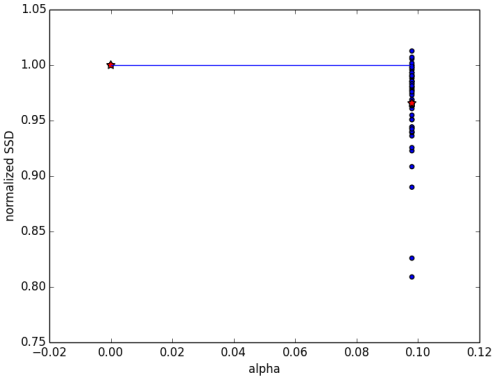
$$p_i^{js} = \beta_i - \alpha(\beta_i - \hat{\beta}) \tag{11}$$

The scalar α can then be estimated empirically using cross-validation, which is what we've done for our first tier validation experiments. This assumption is permitted in the context of James-Stein estimators, and more accurate assumptions about the prior structure can only improve results. More elegant solutions that would allow for anisotropic prior densities are explored in the discussion.

3.3 Results: Using the empirically determined value $\alpha = 0.098$, we computed p_i^{js} according to equation (11). We then compared the ability of the β_i and the p_i^{js} to predict the year 2 follow up images (K_i) by extrapolating their geodesics forward to the year 2 time point and composing the initial image I_i with the resulting transformations. This produced two predictions for each K_i, which we label K_i^{β} and K_i^{js} respectively. We calculated the square Euclidean distances $d(K_i, K_i^{\beta})^2$ and $d(K_i, K_i^{js})^2$ between the ground truth year 2 images and those predictions. In Figure 1 we present $\frac{d(K_i,K_i^{js})^2}{d(K_i,K_i^{\beta})^2}$ for all 57 patients.

Figure 1 shows that by measure of sum of squared differences, the p_i^{js} make better predictions of the third time point image for nearly all patients by about 5% on average. In the best case, an improvement of 20% is achieved. We also subjected the differences $d(K_i, K_i^{\beta})^2 - d(K_i, K_i^{js})^2$ to a pairwise one sided Student's t-test to evaluate the likelihood of achieving these improvements by chance. The p-value of 0.0002 suggests that these results are significant, and that the improvements are due to the use of the James-Stein estimates.

We also inspected the predicted images K_i^{β} and K_i^{js} for any qualitative differences. While the majority of gains due to p_i^{js} are spread thinly throughout the whole image, some improvements clearly correspond to an anatomical interpretation. Figure 2 shows one such case, where β overestimates the expansion of the posterior horn of the left ventricle. The top row is the time series of images I, J, and K from left to right. The bottom row are the predictions corresponding to β and p^{js}. The heat map shows $|\frac{K-\mu}{\sigma} - \frac{K^{js}-\mu^{js}}{\sigma^{js}}| - |\frac{K-\mu}{\sigma} - \frac{K^{\beta}-\mu^{\beta}}{\sigma^{\beta}}|$. That is, it is the difference of the absolute values of the difference images, normalized to their own intensity distributions. This reveals, in cool colors, the locations where p^{js} provided a better estimate of K. The boxed areas show β overestimates the expansion of the ventricle more severely than p^{js}.



Student T test results for $d(K_i, K_i^{\beta})^2 - d(K_i, K_i^{js})^2$

μ	187.23
σ	341.84
T	4.135
p	**.0002**

Fig. 1. Square euclidean distance between ground truth year 2 images and predictions made with p_i^{js} for $\alpha = 0.098$. For each i the distance is normalized by the distance between the ground truth year 2 image and the prediction made with the unrefined β_i. This reveals (by the distance under the red line) the percent improvement earned by using p_i^{js} instead of β_i. The pairwise one sided student's T test shows the improved predictions are due to the use of p_i^{js}.

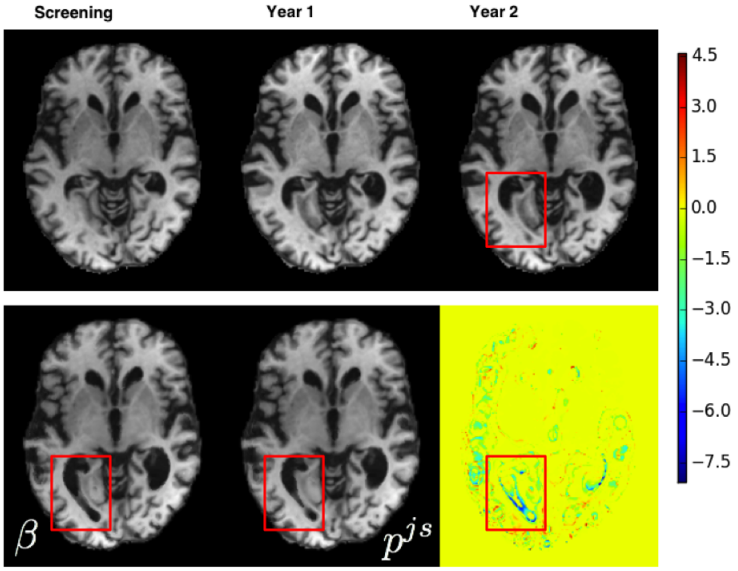


Fig. 2. A time series of images from one patient is shown in the top row. The predictions for the year 2 image derived from β and p^{js} are in the bottom row. The heat map shows in cool colors areas where the p^{js} improved the prediction over β. For this patient, p^{js} reduced an over estimation of ventrical expansion.

4 Discussion

Consistent with expectations, the results indicate the James-Stein estimates p_i^{js} provide geodesics that extrapolate more accurately on average. Hence, our choice of an isotropic prior covariance (that is, $A = a \cdot \mathrm{Id}$ for some scalar a) to cope with the high dimensionality of the β_i is sufficient to gain some improvement in trajectory estimates. A more accurate prior model can only provide more information to improve results.

The simplest relaxation is to let A be diagonal but not necessarily proportional to the identity. In that case, we only have to estimate d variables, an independent variance at every voxel. Different parts of the brain are more or less likely to change over time depending on age and pathology, hence this is more biologically plausible than $A = a \cdot \mathrm{Id}$. More plausible still is to allow A to be non diagonal, but assume that it is sparse. The spatial dependence between voxels is likely to fall off after some appropriate distance, hence many entries in A are likely to be zero or near zero. In that case, many recent methods for learning with sparsity constraints may be brought to bear.

Possibly the most elegant solution would be to use a low dimensional parameterization for the β_i. One option would be to use a subset of the principal components. First one would have to determine an optimal number of components that retains the fine scale variability inherent to longitudinal deformations while reducing the dimension to an acceptable level. A second possibility is to use a band limited Fourier basis. It was recently shown that geodesics for cross-sectional image registration can be parameterized with as few as eight Fourier coefficients per spatial dimension without compromising registration accuracy [14].

Above, we used the estimate $p^* \sim \hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i$, which is the maximum likelihood estimate of p^* under the marginal distribution for β_i. However, for many groupwise representations of time series, p^* is a function of time. Hence, the β_i would need to be normalized in time (as well as in space) before averaging. Similarly when computing p_i^{js}, which involves a term $(\beta_i - p^*)$, p^* should be normalized in the time domain to β_i. Propagating p^* along a geodesic is a simple matter of parallel transport, however finding the appropriate correspondence in time between subjects is not trivial. The naive solution is to use nominal time, however aging and pathological effects may not have constant velocity in time. Also, the age of onset of pathological affects is not known for most patients. Hence, a method that infers temporal correspondence directly from the data independent of the acquisition times of the images such as those in [5, 2] would be needed.

5 Conclusion

We have presented the derivation of multivariate James-Stein estimators in the context of image time series regression. We have established a previously pub-

lished method as a sub-optimal special case of the current model. Further, we have demonstrated that the use of James-Stein estimators can improve the extrapolation of individual geodesics in a population of time series, even with the most naive prior structure. We conclude that for the purpose of interpolation and extrapolation of individual time series within a population, the James-Stein estimate of the geodesic is a more accurate representation of the underlying biological dynamics than the raw measurement.

References

1. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vision 61(2), 139–157 (Feb 2005)
2. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. pp. 297–304. MICCAI '09, Springer-Verlag, Berlin, Heidelberg (2009)
3. Efron, B.: Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics Monographs, Cambridge University Press, Leiden (2010)
4. Fleishman, G.M., Gutman, B.A., Fletcher, P., Thompson, P.M.: Simultaneous longitudinal registration with group-wise similarity prior. In: Ourselin, S., Alexander, D.C., Westin, C.F., Cardoso, M.J. (eds.) Information Processing in Medical Imaging, Lecture Notes in Computer Science, vol. 9123, pp. 746–757 (2015)
5. Hong, Y., Singh, N., Kwitt, R., Niethammer, M.: Time-warped geodesic regression. In: Medical Image Computing and Computer-Assisted Intervention MICCAI 2014, Lecture Notes in Computer Science, vol. 8674, pp. 105–112. Springer International Publishing (2014)
6. Iglesias, J., Liu, C., Thompson, P., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. IEEE Transactions on Medical Imaging 30(9), 1617–1634 (2011)
7. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. pp. 361–379. University of California Press, Berkeley, Calif. (1961)
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. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: Fichtinger, G., Martel, A.L., Peters, T.M. (eds.) MICCAI (2). Lecture Notes in Computer Science, vol. 6892, pp. 655–662. Springer (2011)
10. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI). Lecture Notes in Computer Science (LNCS) (2013)
11. 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. Vision 97(2), 229–241 (Apr 2012)
12. Wei, L., Awate, S., Anderson, J., Fletcher, T.: A functional network estimation method of resting-state fMRI using a hierarchical Markov random field. Neuroimage In press (June, 2014)

13. Yanovsky, I., Thompson, P.M., Osher, S., Leow, A.D.: Topology preserving log-unbiased nonlinear image registration: Theory and implementation. In: CVPR. IEEE Computer Society (2007)
14. Zhang, M., Fletcher, P.: Finite-dimensional lie algebras for fast diffeomorphic image registration. In: Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI). Lecture Notes in Computer Science (LNCS) (2015)