Simultaneous Longitudinal Registration with Group-Wise Similarity Prior

Greg M. Fleishman[1,2(✉)], Boris A. Gutman[2], P. Thomas Fletcher[3], and Paul M. Thompson[2]

[1] Department of Bioengineering, UC Los Angeles, Los Angeles, USA
gfleishman@ucla.edu
[2] Imaging Genetics Center, LONI, Univeristy of Southern California, Los Angeles, USA
[3] Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, USA

Abstract. Here we present an algorithm for the simultaneous registration of N longitudinal image pairs such that information acquired by each pair is used to constrain the registration of each other pair. More specifically, in the geodesic shooting setting for Large Deformation Diffeomorphic Metric Mappings (LDDMM) an average of the initial momenta characterizing the N transformations is maintained throughout and updates to individual momenta are constrained to be similar to this average. In this way, the N registrations are coupled and explore the space of diffeomorphisms as a group, the variance of which is constrained to be small. Our approach is motivated by the observation that transformations learned from images in the same diagnostic category share characteristics. The group-wise consistency prior serves to strengthen the contribution of the common signal among the N image pairs to the transformation for a specific pair, relative to features particular to that pair. We tested the algorithm on 57 longitudinal image pairs of Alzheimer's Disease patients from the Alzheimer's Disease Neuroimaging Initiative and evaluated the ability of the algorithm to produce momenta that better represent the long term biological processes occurring in the underlying anatomy. We found that for many image pairs, momenta learned with the group-wise prior better predict a third time point image unobserved in the registration.

1 Introduction

Nonlinear image registration in brain imaging has progressed to an advanced stage with powerful mathematical tools for sensitive and precise measurements with important theoretical properties. The LDDMM framework establishes a setting wherein constructions like the Fréchet mean and geodesic regression in a space of diffeomorphisms are well defined [7,16]. For some lines of work, the availability of such statistical constructs has promoted a more probabilistic view of transformations. Real image data is noisy, and transformations estimated from it are susceptible to over fitting to this noise. For example, given three images

of the same anatomy acquired over time, it is not likely that a geodesic can be drawn in the transformation space that passes through the identity and the optimal transformations for both of the follow up images. (For example, see Fig. 4 in [5].) Hence, an initial momentum characterizing the geodesic between the identity and the optimal transformation for the first or second follow up image does not describe the optimal geodesic that would be obtained from geodesic regression of all three images.

In this paper we attempt to estimate initial momenta from only two images with improved ability to predict future unobserved images, by simultaneously registering many image pairs that share information throughout optimization. Our approach can be viewed in two equivalent ways: we maintain a group level representation of a transformation and constrain individual transformations to be similar to this representation, which is equivalent to compressing the variance of the set of transformations about their mean. Both of these techniques have precedent in the literature. For example in [13], to estimate functional networks from resting state fMRI data, the authors construct a hierarchical Markov Random Field (hMRF) where the highest level of the hierarchy is a group-wise representation of the network estimate. Edges connecting this level to the individual levels represent a group-wise consistency constraint. Shrinkage of the transformations about their mean is also reminiscent of a James-Stein estimator [4], where we have chosen the average momentum as the prior estimate of the true geodesic regression slope. From this perspective, our method can be viewed as an empirical Bayes prior.

This work uses cross-sectional information in a longitudinal study, which also has precedent in the literature. Other works have used statistical information to constrain registration, but more often in the form of a prior learned from a training set as suggested in [8] and implemented in [2]. These authors constrained the strain tensor of an elastic transformation to be similar to an average strain learned from training data. More recently, the authors of [5] use the transformations of normal controls to refine transformations of AD patients for effects due to the disease. Perhaps most similar to our proposal is [9], in which a group level trajectory is jointly estimated with individual trajectories. The group level trajectory is considered a latent generator for the individual trajectories, but unlike the proposed work, deviation from the group level is not explicitly penalized. In all cases, the incorporation of group level information resulted in transformations with features not found without the group level information, and in many cases, these features were shown to be desirable.

2   Methods

Background, the LDDMM Framework: We begin with a brief review of the LDDMM framework for nonlinear image registration [1]. Given $I_0, I_1 \in L^2(\Omega, \mathbb{R})$, the LDDMM energy functional is defined as:

$$E(v, I_0, I_1) = \int_0^1 \|v\|_V dt + \|I_0 \circ \phi_1^{-1} - I_1\|_{L_2} \tag{1}$$

where $v \in L^2([0,1], V)$ is a time dependent velocity field drawn from the reproducing kernel Hilbert space (RKHS) V. The RKHS is specified by the choice of kernel K, and the inner product in V is then given by $\langle K^{-1}u, u \rangle_{L_2}$ for any $u \in V$. The transformation $\phi(t,x)$ is given by the flow of the velocity $v(t,x)$ through the ODE: $(d/dt)\phi(t,x) = v(t, \phi(t,x))$, with initial condition $\phi(0,x) = x$. $v(t,x)$ and $\phi(t,x)$ will be written as $v_t(x)$ and $\phi_t(x)$. The minimizer of (1) is considered the optimal ϕ for the registration of I_0 and I_1.

The second term on the right hand side of (1) is a quantitative assessment of the similarity between the images $I_0 \circ \phi_1^{-1}$ and I_1, whereas the first term is the geodesic energy of the flow of $v_t(x)$. For suitable choices of K, $\phi_t(x)$ is always a diffeomorphism [10]; hence, (1) defines ϕ_1 to be the transformation that best matches I_0 and I_1 such that ϕ_t is a geodesic in a space of diffeomorphisms specified by the choice of K. As ϕ_t is a geodesic when $E(v, I_0, I_1)$ is optimal, $E(v, I_0, I_1)$ defines a metric distance $d(Id, \phi_1)^2$ in the space of diffeomorphisms. This can also be considered a metric $d(I_0, I_1)^2$ on the orbit given by the group action of the space of diffeomorphisms on the template image I_0.

Background, Geodesic Shooting Algorithm: Several approaches to optimizing (1) have been proposed. In this paper we use the geodesic shooting approach [6,12], which we now review. The kernel K can also be considered a mapping between V^*, the space of linear functionals on V, and V itself. Note that V^* is also a Hilbert space. An Element of V^* is called a momentum. Hence for any momentum $m \in V^*$ there is some $v \in V$ such that $Km = v$ and $K^{-1}v = m$.

An optimal solution to (1) specifies a geodesic, which is uniquely determined by its initial velocity $v_0(x)$, or equivalently, its initial momentum $m_0(x)$. m_t for all t, and hence v_t and ϕ_t, can then be determined by solving the co-adjoint equation [6]: $(\partial/\partial t)m_t = -ad_V^* m_t = -(Dv)^T m_t - Dm_t v - div(v)m_t$, where D denotes the Jacobian operator and $div(.)$ the divergence operator. If the initial momentum is assumed to be proportional to the template image gradient, that is $m_0(x) = p_0(x)\nabla I_0(x)$ for some scalar field p_0, the adjoint equation can be separated into a disjoint system of differential equations for $I_{0,t}$ and p_t respectively [12], where $I_{0,t} = I_0 \circ \phi_t^{-1}$. Considering these equations and the gradient of (1) with respect to v_t, we arrive at a system of partial differential equations that completely specifies ϕ_t given initial conditions I_0 and p_0 ($\star$ denotes convolution):

$$\begin{cases} (\partial/\partial t)p + \nabla \cdot (pv) = 0 \\ (\partial/\partial t)I + \nabla I \cdot v = 0 \\ (\partial/\partial t)v + K \star \nabla I p = 0 \end{cases} \qquad (2)$$

With this in mind, (1) is replaced with a functional of the initial momentum exclusively:

$$\mathcal{E}(p_0, I_0, I_1) = \langle p_0 \nabla I_0, K \star p_0 \nabla I_0 \rangle_{L_2} + \|I_{0,1} - I_1\|_{L_2}^2 \qquad (3)$$

and optimization proceeds within V^* only. In order to optimize (3) by gradient descent, we need the gradient of (3) with respect to p_0, subject to the geodesic

shooting constraints (2). This naturally gives way to an optimal control problem. Time dependent Lagrange multipliers $\hat{p}_t$, $\hat{I}_{0,t}$, and $\hat{v}_t$ enable us to write an augmented functional for (3) incorporating the constraints (2):

$$\tilde{\mathcal{E}}(p_0, I_0, I_1) = \mathcal{E} + \int_0^1 \langle \hat{p}_t, (\partial/\partial t)p + \nabla \cdot (pv) \rangle dt \; +$$

$$\int_0^1 \langle \hat{I}_{0,t}, (\partial/\partial t)I + \nabla I \cdot v \rangle dt \; + \qquad (4)$$

$$\int_0^1 \langle \hat{v}_t, (\partial/\partial t)v + K \star \nabla I p \rangle dt$$

The first variation of (4) gives the gradient of (3) subject to (2):

$$\nabla_{p_0} \mathcal{E} = \nabla I_0 \cdot K \star p_0 \nabla I_0 - \hat{p}_0 \qquad (5)$$

where $\hat{p}_0$ is specified by a system of partial differential equations solved backward in time termed the adjoint system:

$$\begin{cases} (\partial/\partial t)\hat{p} + \nabla \hat{p} \cdot v - \nabla I \cdot K \star \hat{v} = 0 \\ (\partial/\partial t)\hat{I} + \nabla \cdot (Iv) + \nabla \cdot pK \star \hat{v} = 0 \\ (\partial/\partial t)\hat{v} + \hat{I}\nabla I - p\nabla \hat{p} = 0 \end{cases} \qquad (6)$$

with initial conditions $\hat{I}_1 = I_1 - I_{0,1}$ and $\hat{p}_1 = 0$. The gradient descent proceeds by solving the system (2) forward in time to acquire p_t, $I_{0,1}$, and v_t for a sufficiently dense sampling of $t \in [0,1]$, then solving (6) backward in time to acquire $\hat{p}_0$. p_0 is then updated with (5), and the process is repeated until convergence.

Group-wise Similarity Prior: We consider the case where we are given N longitudinal image pairs $I_0^i, I_1^i \in L^2(\Omega, \mathbb{R})$, $i \in [1, 2, ..., N]$, all taken approximately the same time interval apart. We take Ω to be the unit cube with periodic boundary conditions, and the time interval to be $[0,1]$. Additionally, we are given N transformations ψ^i mapping the initial images I_0^i to a Minimal Deformation Template (MDT) coordinate system, that is, $I_0^k \circ \psi^k \sim I_0^j \circ \psi^j$ for all k and j. To consider all N registrations simultaneously with no modification to the geodesic shooting approach, we could write $\tilde{\mathcal{E}}_{tot} = \sum_{i=1}^N \tilde{\mathcal{E}}_i$ where $\tilde{\mathcal{E}}_i$ is eq. (3) for the ith image pair. The first variation of $\tilde{\mathcal{E}}_{tot}$ with respect to an initial momentum p_0^i will only include terms for the ith pair, that is, the N transformations are decoupled. However, we would like the N transformations to explore the space of diffeomorphisms as a group. We couple them by considering equations of the form:

$$\tilde{\mathcal{E}}_{tot} = \alpha \mathcal{G}(p_1, p_2, ..., p_N) \; + \sum_{i=1}^N \tilde{\mathcal{E}}_i \qquad (7)$$

$\mathcal{G}(.)$ is intended to enforce some criteria that we may think all p_0^i must satisfy. In this paper, we consider longitudinal studies where all N image pairs come from patients in the same diagnostic group, where a predictable distribution of

volume change is known to occur. Because V^* is a Hilbert space, we can calculate statistical moments in this space in an ordinary manner, being careful to spatially normalize the p_0^i to a MDT coordinate system using coadjoint transport [15]. First, let $p_0^{mdt,i} = |D\psi^i|p_0^i \circ \psi^i$, be the ith initial momentum in the MDT coordinate system. Let $p_0^{mdt,avg} = (1/N)\sum_{i=1}^N p_0^{mdt,i}$ be the sample average initial momentum in MDT coordinates. Let $p_0^{mdt,cen,i} = p_0^{mdt,i} - p_0^{mdt,avg}$ be the mean centered initial momentum for image pair i in MDT coordinates, and let $A = [p_0^{mdt,cen,1}, p_0^{mdt,cen,2}, ..., p_0^{mdt,cen,N}]^T$ be the mean centered design matrix for all initial momenta in MDT coordinates. We take $\mathcal{G}(.)$ to be:

$$\mathcal{G}(p_1, p_2, ..., p_N) = Trace(AA^T) = \sum_{i=1}^N \|p_0^{mdt,i} - p_0^{mdt,avg}\|_{L_2}^2 \qquad (8)$$

the trace of the sample inner-product matrix for p_0.

First we consider the rightmost form of (8). We see that this term maintains a group-wise average of the initial momentum, and requires that all momenta be close to this average. This is similar to hierarchical latent variable models that maintain a group-wise representation of the data and constrain updates to predictions to be similar to this representation.

Now consider the middle form of (8). The covariance matrix $A^T A$ has the same eigenvalues as the inner-product matrix AA^T. Covariance matrices are symmetric positive-definite, and therefore have all real non-negative eigenvalues. Finally, the trace of a matrix is invariant to rotation. So, considering the canonical form of $A^T A$, we see that $Trace(AA^T) = \sum_{i=1}^N \lambda_i$ where λ_i is the ith eigenvalue of the sample covariance matrix. Each λ_i is a measurement of the magnitude of the corresponding principal axis of the covariance. Hence, by minimizing $\sum_{i=1}^N \lambda_i$, we are compressing the covariance about the mean.

To minimize (8) we need to consider the contribution of $\mathcal{G}(.)$ to the gradient (5). In our implementation, $p_0^{mdt,avg}$ is considered to be constant during any given iteration (see section on gradient descent strategy). Hence, the gradient of $\mathcal{G}(p_1, p_2, ..., p_N)$ with respect to p_0^k for some k in MDT coordinates is simply found to be:

$$\nabla_{p_0^k}\mathcal{G}(p_1, p_2, ..., p_N) = 2\alpha(p_0^{mdt,k} - p_0^{mdt,avg}) \qquad (9)$$

to put this back into individual coordinates, we compose with the appropriate inverse transformation:

$$\nabla_{p_0^k}\mathcal{G}(p_1, p_2, ..., p_N) = 2\alpha|D(\psi^k)^{-1}|(p_0^{mdt,k} - p_0^{mdt,avg}) \circ (\psi^k)^{-1} \qquad (10)$$

and so the complete gradient of (7) with respect to an initial momentum p_0^k in the coordinate system for the kth template image is the sum of Eqs. (5) and (10). The result is that for every update of p_0^k it is pulled in such a way as to map I_0^k to I_1^k by (5), but it is also held close to the group representation of p_0 by (10).

Gradient Descent Algorithms for Optimization of (7): We now consider optimizing (7) with respect to each p_0^i one at a time. Multiple strategies are available for the order in which we update the p_0^i. The most rigorous update

would be to use the maximum amount of information possible at each update. That is, for the $(l+1)$st update of the kth momentum, p_0^{avg} in (10) equals $(1/n)\sum_{i=1}^{k-1} p_0^{i,l+1} \circ \psi^i + (1/n)\sum_{i=k}^{N} p_0^{i,l} \circ \psi^i$. This approach requires the N registrations to be done in series for every iteration and is exceedingly costly in both time and memory.

An alternative is use $(1/n)\sum_{i=1}^{N} p_0^{i,l} \circ \psi^i$ for p_0^{avg} for all N at the $(l+1)$st iteration. This way, for a given iteration, the N $p_0^{i,l}$ can be updated in parallel. Subsequently, each pair shares its updated value $p_0^{i,l+1}$ to compute $p_0^{avg,l+1} = (1/n)\sum_{i=1}^{N} p_0^{i,l+1} \circ \psi^i$ to be used in the $(l+2)$nd iteration. We used this strategy to compute the results presented in the next section.

3   Results

Experimental Setup: 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 [3]. Transformations ψ^i mapping the template images I_0^i into a MDT coordinate system were computed using a preexisting implementation of [14].

We used a multi-resolution approach for 50, 30, 20, and 5 iterations at 64^3, 80^3, 96^3, and 128^3 resolutions respectively to register the screening images to the year 1 follow up images. We used the second strategy described in the above section to optimize (7) with respect to the initial momenta for the 57 pairs. To test the influence of the group-wise term $\mathcal{G}(.)$, we ran the algorithm over a range of values for α including 0.0 (control), 0.01, 0.025, 0.05, 0.075, 0.1, and 0.5. After completion, we computed and compared the average and variance of the initial momenta for each value of α. We then solved the system (2) over the interval $[0,2]$, which in this case represents 2 years, and compared the computed image $I_{0,2}$ to the year 2 follow up images for all values of α.

Mean and Variance Images: Coronal slices for the final mean and variance of the initial momenta are shown in Fig. 1 for all values of α. As α increases, both the mean and variance become smaller in magnitude, however the variance falls off at a much faster pace. The primary features of the mean image, including the change in the ventricles and temporal lobes, remain the strongest with increasing α, while individual features fade away with increasing group-wise influence.
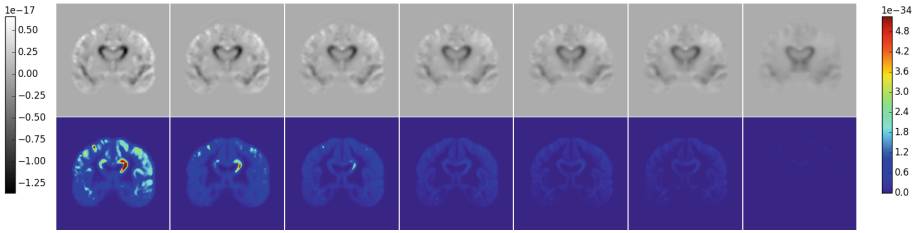
Fig. 1. Mean and variance images for different values of α. Top: Mean images, Bottom: Variance images, Columns correspond to α values from left to right: 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, and 0.5.

The Sum of Squared Difference During Registration: The initial sum of squared difference (SSD) between the screening and year 1 image was retained and used to normalize the SSD at every iteration. This normalized SSD was summed over all 57 image pairs. The results for all values of α are shown in Fig. 2. Clearly, as α increases the total normalized SSD increases for every iteration, which is expected considering the forms of Eqs. (7) and (10). As alpha increases, exact matching to the target is compromised for more coherence with the group-wise representation. The spikes occur where the resolution changes in the multi-scale registration approach.

Prediction of Year 2 Images from Initial Momenta: The momenta learned for all values of α were integrated from $[0, 2]$, representing a 2 year period, and the screening images were transformed with the resulting diffeomorphisms. These images were quantitatively compared to actual year 2 follow up acquisitions. The SSD between the year 2 prediction and actual year 2 image, normalized by its value for $\alpha = 0.0$, is presented in Fig. 3. A value less than one indicates the prediction at a particular α level is closer by SSD than the prediction for $\alpha = 0.0$. Clearly, for many images the prediction improves with increasing α. These images are those for which the true, unobserved, initial momenta lies closer to the group-wise mean. For some images, the prediction becomes worse with increasing α. These images are those for which the true, unobserved, initial momenta does not lie closer to the group-wise mean. An immediate extension of this work to address this issue is to modify (8) to allow for multiple subgroup-wise representations and/or to accommodate outliers.

We performed a one-sided student's t-test to determine if the SSD for predictions with α not equal to zero were significantly different from those with α equal to zero. All values of α except $\alpha = 0.5$ have significantly different SSD values (at a standard significance level of p = 0.05) for their predictions. The relevant values are presented in Table 1.

Prediction of Year 2 Images from Average Momenta: The average momentum for all α in MDT coordinates was transformed into individual coordinates for the ith image pair using coadjoint transport through $(\psi^i)^{-1}$. The resulting
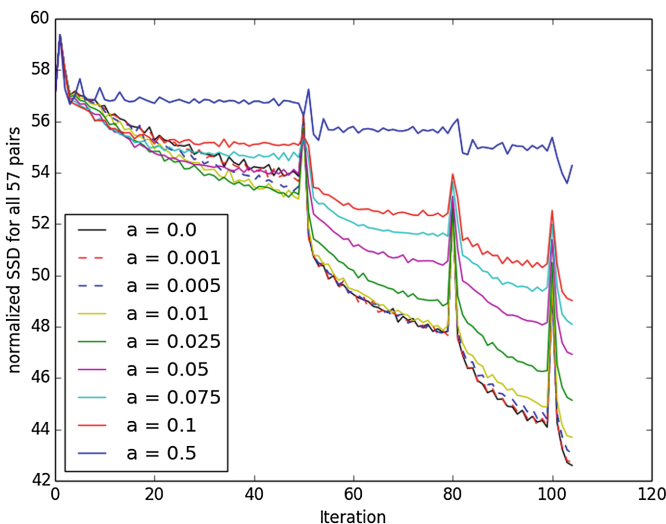
Fig. 2. Normalized SSD throughout optimization for all values of α. The Spikes occur when the resolution changes.

average momenta in individual coordinates were integrated over $[0, 2]$, representing a 2 year period. The screening images were transformed with the resulting diffeomorphisms and the resulting images were compared to the actual year 2 acquisitions. The results are presented in Fig. 4.

We performed a one-sided student's t-test to determine if the SSD for predictions from the average momenta with α not equal to zero were significantly different from those with α equal to zero. All values of α have significantly different SSD values (at a standard significance level of $p = 0.05$) for their predictions. The relevant values are presented in Table 2.

Table 1. t-test results comparing all α not equal to zero with $\alpha = 0$ for SSD between year 2 prediction and acquired year 2 image. μ is the average difference between SSD values for $\alpha = 0$ and $\alpha \neq 0$, σ is the standard deviation, T is the t-statistic, and p is the p-value. Recall, there were 57 image pairs. Significant results are bold.

α	0.001	0.005	0.01	0.025	0.05	0.075	0.1	0.5
μ	13.70	70.21	120.51	222.10	287.75	306.79	324.23	157.03
σ	52.25	143.67	278.01	586.49	953.82	1177.70	1335.04	2066.4
T	1.98	3.70	3.27	2.86	2.28	1.97	1.83	0.57
p	**.026**	**.00025**	**.00092**	**.003**	**.013**	**0.027**	**0.036**	0.29
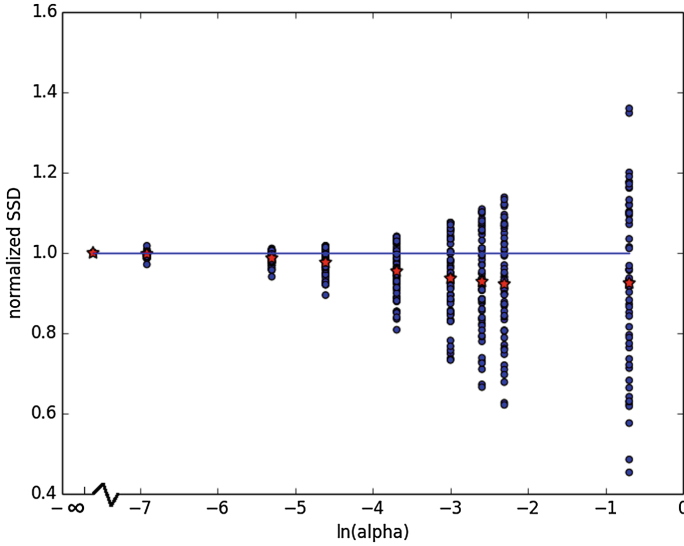
Fig. 3. SSD between year 2 images predicted by integration of initial momenta and actual year 2 image acquisitions for all 57 image pairs and all ln(α) values. The red stars represent the mean.
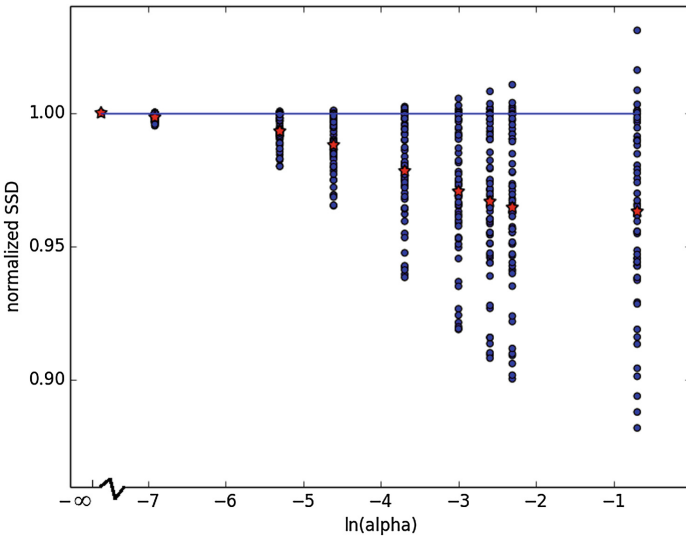
Fig. 4. SSD between year 2 images predicted by integration of average momenta and actual year 2 image acquisitions for all 57 image pairs and all ln(α) values. The red stars represent the mean.

4   Discussion

The first feature of the above presented methods and results to discuss is the obvious compromise between exact image pair matching and group-wise consistency represented by the parameter α. Figure 2 demonstrates the sensitivity of the exact image matching to this parameter. It's interesting to note that the solution trajectory over iterations is very similar in shape for all values of α, though we get less exact matching as α increases as expected. Figure 3 and Table 1 demonstrate the potential advantage to group-wise consistency in learning momenta that more accurately reflect the unobserved long term change. Figure 3 and Table 1 both suggest that there are some values of α that strike a potentially desirable compromise between exact image matching and improved prediction of long term change.

Of course, for some images, the momenta learned with coupling are worse predictors of long term change. As mentioned previously, one avenue to address this is to allow multiple sub-group representations and assign each image pair to the sub-group representation that best approximates it. One interesting question that arises is what is the optimal number of sub-groups for a given population? Additionally, what differences will the sub-group representations encode in them after convergence? Alternatively, the mean is sensitive to outliers, and we could consider replacing the group representation with a different statistic more robust to such variation.

The proposed work has made no effort to normalize temporal misalignment in disease progression across patients. The experimental results suggest that AD disease progression is sufficiently similar at different stages of the disease for group level information to be applicable to individual trajectory estimation. However, this may not be the case for other populations such as the Mild Cognitively Impaired (MCI) or other Neurodegenerative disorders with less well characterized structural changes. Hence, explicit modeling of temporal misalignment in age and disease progression as done in [9] may improve results.

It is important to mention that momenta learned with this technique should not be naively used for statistical tests. We have explicitly minimized the trace covariance of these momenta, so any voxel-wise statistics computed from them

Table 2. t-test results comparing all α not equal to zero with $\alpha = 0$ for SSD between year 2 prediction from average momenta and acquired year 2 image. μ is the average difference between SSD values for $\alpha = 0$ and $\alpha! = 0$, σ is the standard deviation, T is the t-statistic, p is the p-value. Recall, there were 57 image pairs. Significant results are bold.

α	0.001	0.005	0.01	0.025	0.05	0.075	0.1	0.5
μ	10.91	46.26	81.55	148.97	203.29	230.76	246.85	259.22
σ	9.04	37.91	67.28	127.07	179.47	207.67	224.56	249.41
T	9.11	9.21	9.15	8.85	8.55	8.39	8.30	7.85
p	**2.22e-12**	**1.53e-12**	**1.92e-12**	**5.89e-12**	**1.82e-11**	**3.34e-11**	**4.69e-11**	**2.59e-10**

are biased [11]. This issue can be compensated for by determining the null distribution for a particular statistic and establishing significance relative to this learned distribution. However, non-statistical inference applications such as momenta or change map atlas construction and shooting of individual templates via the learned momenta are not affected by this problem.

5 Conclusions

We have presented a mathematical framework for coupling the registration of N image pairs in the geodesic shooting approach for the optimization of the LDDMM energy functional. Individual registrations are coupled by maintaining a group-wise representation of their initial momenta and constraining updates to stay close to this representation. This is an explicit minimization of the variance of the initial momenta in the Lie algebra for the space of diffeomorphisms specified by the choice of metric K. This establishes a trade-off between exact image matching for individual image pairs and group-wise consistency. We've shown that increasing group-wise consistency can improve the prediction of long term change encoded within individual momenta. Finally, we have described some of the strengths and weaknesses of our initial choice for the coupling term $\mathcal{G}(.)$ and suggested methods to address those weaknesses.

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. Vis. **61**(2), 139–157 (2005). doi:10.1023/B:VISI.0000043755.93987.aa
2. Brun, C.C., Lepore, N., Pennec, X., Chou, Y.Y., Lee, A.D., de Zubicaray, G.I., McMahon, K., Wright, M.J., Gee, J.C., Thompson, P.M.: A nonconservative lagrangian framework for statistical fluid registration - safira. IEEE Trans. Med. Imaging **30**(2), 184–202 (2011)
3. Iglesias, J., Liu, C., Thompson, P., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans. Med. Imaging **30**(9), 1617–1634 (2011)
4. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1. Contributions to the Theory of Statistics, pp. 361–379. University of California Press, Berkeley (1961). http://projecteuclid.org/euclid.bsmsp/1200512173
5. Lorenzi, M., Pennec, X., Frisoni, G.B., Ayache, N.: Disentangling normal aging from Alzheimer's disease in structural MR images. Neurobiol. Aging **36**, S42–S52 (2014)
6. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. J. Math. Imaging Vis. **24**(2), 209–228 (2006)
7. Niethammer, M., Huang, Y., Vialard, F.-X.: Geodesic regression for image time-series. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 655–662. Springer, Heidelberg (2011)

8. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian elasticity: a statistical regularization framework for non-linear registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 943–950. Springer, Heidelberg (2005)

9. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) IPMI 2013. LNCS, vol. 7917, pp. 560–571. Springer, Heidelberg (2013)

10. Trouve, A.: Diffeomorphisms groups and pattern matching in image analysis. Int. J. Comput. Vis. **28**, 213–221 (1998)

11. Tustison, N.J., Avants, B.B., Cook, P.A., Kim, J., Whyte, J., Gee, J.C., Stone, J.R.: Logical circularity in voxel-based analysis: normalization strategy may induce statistical bias. Hum. Brain Mapp. **35**, 745–759 (2012)

12. 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). doi:10.1007/s11263-011-0481-8

13. Wei, L., Awate, S., Anderson, J., Fletcher, T.: A functional network estimation method of resting-state fMRI using a hierarchical markov random field. NeuroImage **100**, 520–534 (2014)

14. 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)

15. Younes, L., Qiu, A., Winslow, R.L., Miller, M.I.: Transport of relational structures in groups of diffeomorphisms. J. Math. Imaging Vis. **32**(1), 41–56 (2008)

16. 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)