

A Transformation Similarity Constraint for Groupwise Nonlinear Registration in Longitudinal Neuro Imaging Studies

Greg M. Fleishman^{a b}, Boris A. Gutman^b, P. Thomas Fletcher^c, and Paul Thompson^b

^aDept. of Bioengineering, University of California, Los Angeles

^bImaging Genetics Center, INI, University of Southern California

^cScientific Computing and Imaging Institute, University of Utah

ABSTRACT

Patients with Alzheimer's disease and other brain disorders often show a similar spatial distribution of volume change throughout the brain over time,^{1,2} but this information is not yet used in registration algorithms to refine the quantification of change. Here, we develop a mathematical basis to incorporate that prior information into a longitudinal structural neuroimaging study. We modify the canonical minimization problem for non-linear registration to include a term that couples a collection of registrations together to enforce group similarity. More specifically, throughout the computation we maintain a group-level representation of the transformations and constrain updates to individual transformations to be similar to this representation. The derivations necessary to produce the Euler-Lagrange equations for the coupling term are presented and a gradient descent algorithm based on the formulation was implemented. We demonstrate using 57 longitudinal image pairs from the Alzheimer's Disease Neuroimaging Initiative (ADNI) that longitudinal registration with such a groupwise coupling prior is more robust to noise in estimating change, suggesting such change maps may have several important applications.

Keywords: Nonlinear registration, groupwise registration, longitudinal registration, Bayesian registration, Tensor Based Morphometry, ADNI

1. INTRODUCTION

In longitudinal neuroimaging studies, non-linear image registration is often used to quantify the spatial distribution of volume change experienced by the brain over time.³ In such a setting, we are given two images of the same brain acquired at different times and we wish to compute a mapping between them that reveals where tissue has contracted or expanded. To make this notion mathematically precise, take $R(x), T(x) \in L^2(\Omega, \mathbb{R})$ to be finite scalar valued images over the domain $\Omega \subset \mathbb{R}^d$; further suppose $R(x)$ and $T(x)$ are images of the same brain acquired at different times. The problem of non-linear image registration is then typically formulated as follows: find a transformation $\phi : \Omega \rightarrow \Omega$ on the coordinate system x such that*:

$$\mathcal{D}[R, T(\phi)] + \beta \mathcal{S}[\phi] = \min \quad (1)$$

*We will often suppress the positional dependence of transformations and images to make the notation more compact

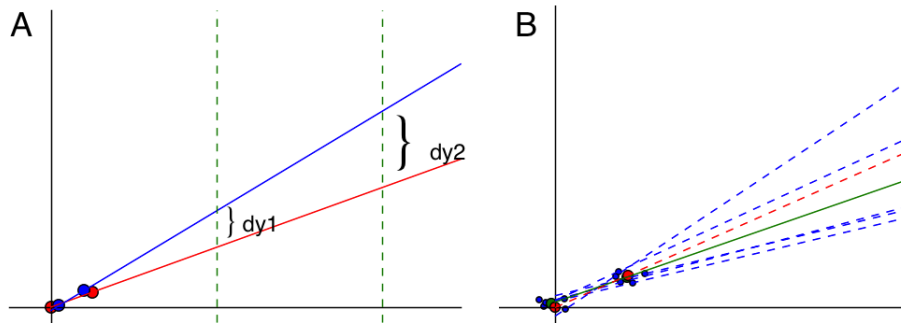


Figure 1. Demonstration of groupwise coupling for improved model estimation in the presence of noise

Here, $\mathcal{D} : L^2(\Omega, \mathbb{R}) \times L^2(\Omega, \mathbb{R}) \rightarrow \mathbb{R}$ is an appropriate image similarity function that measures how well matched two images are, and hence assesses the likelihood that the transformation ϕ is a suitable one. Although there are some specific biological situations where the assumption may be violated (e.g., neoplastic growth), for our current application we assume that transformations that reflect actual biological change are bijective and smooth, so the function $\mathcal{S} : \phi \rightarrow \mathbb{R}$ is a regularizer/smoothing on an appropriate set of transformations ϕ . Finally β is a scalar that regulates the relative importance between image matching and the smoothness of ϕ as regulated by \mathcal{S} . Given an approximate solution to (1), the spatial distribution of tissue change can be estimated from the Jacobian determinant of the transformation, or (less commonly) other metrics derived from the local deformation or strain tensor. This method is often termed tensor-based morphometry (TBM).⁴

In a longitudinal study with n individuals, we are given a collection of such image pairs: $R_0, T_0, \dots, R_{n-1}, T_{n-1}$ and we wish to find transformations $\phi_0, \dots, \phi_{n-1}$ such that:

$$\sum_{i=0}^{n-1} \mathcal{D}[R_i, T_i[\phi_i]] + \beta \mathcal{S}[\phi_i] = \min \quad (2)$$

Each term in this sum can be minimized independently and the task is simply a concatenation of n separate registrations presenting no new challenges. However, consider the addition of a term $\mathcal{P} : \phi^n \rightarrow \mathbb{R}$ to measure some aspect of the collection of transformations $\phi_0, \dots, \phi_{n-1}$ as a whole, producing the objective:

$$\alpha \mathcal{P}[\phi_0, \dots, \phi_{n-1}] + \sum_{i=0}^{n-1} \mathcal{D}[R_i, T_i[\phi_i]] + \beta \mathcal{S}[\phi_i] = \min \quad (3)$$

for some scalar α . The construction of each ϕ_i is no longer independent of the others, and the individual registrations must share information to achieve a minimum.

Suppose the images R_0, \dots, R_{n-1} were acquired before the same event or process occurred in the n different subjects and images T_0, \dots, T_{n-1} were acquired after. It is reasonable to expect the transformations $\phi_0, \dots, \phi_{n-1}$ to share some characteristics. Performing the registrations independently does not take advantage of knowledge that can be learned by observing similarity amongst the group. We incorporate \mathcal{P} in this work to account for our prior knowledge that the images have undergone transformations generated from the same or similar processes.

To further motivate the use of such a prior, consider the example presented in Figure 1. In Figure 1A, the two red dots represent the unobserved ground truth values of two points in a 2 dimensional independent/dependent variable system. If we assume the model is linear, the optimal linear regression between the two points is the red line passing through both points. The blue dots correspond to noisy measurements of these variables. We attempt to estimate the parameters of the model from the noisy measurements and find the blue line, which clearly differs from the unobserved ground truth. For values of t (the independent variable) close to the observations, the difference dy_1 between the ground truth model and the estimated model may be small. However, for values of t further from the observations, the difference dy_2 will grow substantially. Unfortunately, this is often the situation in longitudinal neuroimaging: we wish to predict the trajectory of a brain's growth and shape change from only two noisy observations acquired temporally close together. In clinical trial design, or to assess treatment effects in an individual, there is often a desire to begin to estimate brain changes at the soonest possible follow-up interval, when changes are harder to estimate.

Figure 1B demonstrates the use of a groupwise similarity prior. Again, the red dots represent the unobserved ground truth values of two points in a linear system, and the dotted red line represents the optimal linear prediction. We make five separate noisy observations of the system (the blue dots), and wish to interpolate between each of the five pairs separately (the dotted blue lines). If we take the average of these models (the green dots and line), and shrink each model toward it, in this case, all the estimated models will approach the ground truth model. Predictions made from these improved individual models will be closer to the ground truth values at distant time points than predictions made from the original estimates. Here, the average estimated model (in green) serves as the groupwise representation, and we couple the estimation of the individual models by enforcing they be similar to their average.

Other works have incorporated cross-sectional or group level information into individual computations to improve sensitivity and group level consistency. In Pennec et al.,⁵ the authors suggest computing statistics of the Hencky strain tensor of the transformation from a training data set of multiple registration pairs. In a second round of registrations, the strain tensor is constrained by the previously computed statistics. Our method differs in that the group level information is determined from the data itself during registration in an empirical Bayes fashion, eliminating the need for a training set and multiple rounds of registration.

A groupwise similarity prior was also used more recently in Liu et al.⁶ for the determination of resting state networks in fMRI analysis. In this work, the authors propose using a hierarchical Markov Random Field (hMRF) to cluster voxels into functional networks in fMRI analysis. The upper level of the hierarchy is an average representation of the resting state network across the population. Edges connecting this upper level to the individual fields represent a constraint that individual networks should be similar to the group level representation. Similar to our work, the group level representation and individual networks are simultaneously estimated in an iterative fashion. However, their work is formulated in a discrete setting and for a different data modality.

2. METHODS

2.1 Mathematical Formulation

Here, we propose \mathcal{P} in (3) be a measurement of the disagreement among the patterns of tissue change determined by the transformations $\phi_0, \dots, \phi_{n-1}$. If we want to compare volumetric change maps, we have to do it in a common coordinate system. So along with the images R_i and T_i , we assume we're given transformations ψ_i such that $T_j(\psi_j) \sim T_k(\psi_k)$ for all $j, k \in [0, 1, \dots, n-1]$. This is achieved by computing a Minimal Deformation Template (MDT) for the initial time point images T_i prior to longitudinal registration using established techniques.

First, let $J_{\phi_i}^{mdt}(x) = J_{\phi_i}(\psi_i(x))$ be the Jacobian matrix of transformation ϕ_i at position x in the common coordinate system, and let $\log(\det[J_{\phi_i}^{mdt}(x)])$ be its log determinant. We take the sample mean of these to be $\mu(x) = (1/n) \sum_{i=0}^{n-1} \log(\det[J_{\phi_i}^{mdt}(x)])$, and use it to center the log Jacobian determinants: $\log(\det[J_{\phi_i}^{mdt}(x)])_{cen} = \log(\det[J_{\phi_i}^{mdt}(x)]) - \mu(x)$. Next, let \mathbf{X} be a uniform grid sampling of the domain $\Omega \subset \mathbb{R}^d$ and $a_i = \log(\det[J_{\phi_i}^{mdt}(\mathbf{X})])_{cen}$ be a column vector of the mean centered log Jacobian determinant in the common coordinate system evaluated at all spatial positions in \mathbf{X} . Finally, we define the mean centered design matrix of the log Jacobian determinant in the common coordinate system to be:

$$A = [a_0, a_1, \dots, a_{n-1}]^T \quad (4)$$

and choose for \mathcal{P} in (3):

$$\mathcal{P}[\phi_0, \dots, \phi_{n-1}] = \text{Trace}(AA^T) \quad (5)$$

Note that $(1/n)A^T A$ is the sample covariance matrix⁷ for the $\log(\det[J_{\phi_i}^{mdt}(\mathbf{X})])_{cen}$. Sample covariance matrices are symmetric positive semi-definite, and thus have a canonical representation $A^T A = V\Lambda V^T$ for unitary/rotation matrix V and diagonal matrix Λ .⁷ The columns of V are the eigenvectors, or principal axes, of the covariance and the elements of Λ are the corresponding eigenvalues, λ_i . Because $A^T A$ is symmetric positive semi-definite, the λ_i are real and non-negative. Each λ_i is a measure of the magnitude of the corresponding principal axis. Finally, the inner-product matrix AA^T has the same non-zero eigenvalues as $A^T A$.⁷

The trace of a matrix is invariant to rotation, hence $\text{Trace}(AA^T) = \text{Trace}(\Lambda) = \sum_{i=0}^{n-1} \lambda_i$. From this point of view, we see that minimizing (5) minimizes the magnitude of the principal axes of the covariance. That is, (5) is a scalar measure of the spread of the a_i , and minimizing it compresses the log Jacobian determinant images about their mean. We also have $\text{Trace}(AA^T) = \sum_{i=0}^{n-1} \|\log(\det[J_{\phi_i}^{mdt}(\mathbf{X})])_{cen} - \mu(\mathbf{X})\|^2$. From this point of view, we can say that throughout the computation of the ϕ_i , we maintain a group representation of the volumetric change in the brain, $\mu(\mathbf{X})$, and constrain individual volumetric change maps to be similar to it.

To simplify notation, we return to the continuous setting and omit positional dependence of transformations and images. To optimize (3) using (5), we make use of the gradient:

$$\nabla_{\phi_k} \mathcal{P}_1 = -\nabla \cdot [(\log(\det[J_{\phi_k}^{mdt}]) - \mu) J_{\phi_k}^{-1}] \quad (6)$$

in a gradient descent strategy detailed below. The derivation of (6) from (5) is presented in the appendix. This gradient is proportional to the distance between the log Jacobian determinant for the transformation we wish to update and the mean log Jacobian determinant

across the data set. Hence, when this distance is large, this term more forcefully pushes the transformation in the direction of the average log Jacobian determinant image. Note also that this gradient is in the MDT coordinate system and must be composed with ψ_k^{-1} before it can be used to update ϕ_k .

2.2 Algorithm Formulation

Denote the Euler-Lagrange equations corresponding to \mathcal{P} , \mathcal{D} , and \mathcal{S} as p , d , and s respectively. Optimization of (3) with respect to the ϕ_i is done iteratively by gradient descent. One approach to this optimization is to solve:

$$s[\phi_k^{l+1}] = \frac{1}{\beta}d[\phi_k^l] + \frac{\alpha}{\beta}p[\phi_0^{l+1}, \dots, \phi_{k-1}^{l+1}, \phi_k^l, \phi_{k+1}^l, \dots, \phi_{n-1}^l] \quad (7)$$

for ϕ_k^{l+1} . Here, d and p apply forces to the template image: d in a direction that registers the template to the reference image and p such that the log Jacobian determinant of the transformation is pushed toward its average. α and β control the compromise between these two forces, representing individual matching accuracy and group level consistency. Finally, s regularizes or smooths this force field.

In (7), the ϕ_i are updated one at a time, taking advantage of the maximum amount of information available at all times. This approach is exceedingly time consuming. A more tractable approach is to solve:

$$s[\phi_k^{l+1}] = \frac{1}{\beta}d[\phi_k^l] + \frac{\alpha}{\beta}p[\phi_0^l, \dots, \phi_{k-1}^l, \phi_k^l, \phi_{k+1}^l, \dots, \phi_{n-1}^l] \quad (8)$$

in parallel for all n image pairs holding the average μ fixed, then update μ using all of the ϕ_i^{l+1} for use in the next iteration. We used this method in our experiments.

Finally, for \mathcal{S} we use the elastic potential energy, the EL equations for which are the Navier-Lame equations for linear elasticity:

$$s[\phi] = \mu \nabla^2 u + (\lambda + \mu) \nabla(\nabla \cdot u) \quad (9)$$

for lame constants λ and μ , which control the relative contributions of the Laplacian and gradient of divergence terms. u is the displacement vector component of $\phi(x) = x + u(x)$.

For \mathcal{D} we used the cross correlation (CC). Let $\mu(I)$ be the intensity average for image I . Then $v_1 = \int_{\Omega} [T(\phi) - \mu(T)]^2 dx$ is the intensity variance of T , $v_2 = \int_{\Omega} [R - \mu(R)]^2 dx$ is the intensity variance of R , and $v_{1,2} = \int_{\Omega} [R - \mu(R)][T(\phi) - \mu(T)] dx$ is the intensity covariance of images R and T , assuming Ω is the unit cube. The CC is then:

$$\mathcal{D}[R, T(\phi)] = \frac{v_{1,2}^2}{v_1 v_2} \quad (10)$$

This formula ranges from 0 to 1 and equals 1 if the statistically normalized images are perfectly aligned, hence we attempt to maximize rather than minimize it. The EL equations of the CC are:⁸

$$d[R, T(\phi)] = \frac{2(Rv_1 v_2 v_{1,2} - T(\phi)v_{1,2}^2 v_2)}{(v_1 v_2)^2} \nabla T(\phi) \quad (11)$$

2.3 Experimental Formulation

We downloaded screening and 1 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 men of mean age 75.91 +/- 7.85 years and 25 women, with mean age 75.08 +/- 8.15 years. This was the maximum number of individuals we could download from the ADNI 1 cohort who 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.⁹ Subsequent to downloading, the images were linearly registered to the ICBM template and skull stripped using ROBEX.¹⁰ Transformations ψ_i mapping the template images T_i into a MDT coordinate system were computed using a preexisting implementation of the registration method in Yanovsky et al.¹¹

A second set of images was created by adding isotropic Gaussian noise to every voxel within the skull stripped mask of the initial 57 image pairs. The average standard deviation in intensity of the original images was 0.164. We used 0.082 for the standard deviation of the Gaussian noise, which made for an artificial SNR of 4 for the noisy image set. Both the initial and noisy image sets were registered for 100 iterations using (8) with $\beta = 1$ and $\alpha \in \{0, 25, 50, 75, 100, 150, 200\}$. If J_α is the Jacobian determinant map for the transformation of a given image pair from the initial no noise added images at a given value of α , then let J_α^{nz} be the corresponding Jacobian determinant map for the images with noise added. We computed:

$$SSD = \int_{\Omega} (J_\alpha - J_\alpha^{nz})^2 dx \quad (12)$$

for all image pairs and values of α . We hypothesized that as α increased, the SSD would decrease. That is, as the 57 image pairs shared more information during registration, their sensitivity to noise in the estimation of change would decrease.

3. RESULTS

The total cross-correlation (CC) throughout optimization summed over all 57 image pairs for all tested values of α are presented in Figure 2. Figure 2A shows the values for the initial image set, and figure 2B shows the values for the noisy image set. The algorithm succeeds in improving the CC for all values of α . As α increases the final CC value decreases. This is entirely consistent with expectations as we have compromised exact matching for groupwise consistency. The final CC values are lower for the noisy image set, indicating that exact matching is harder in the presence of this substantial noise.

A coronal cross-section of the mean and variance of the Jacobian determinant images in MDT coordinates for all values of α are shown in Figure 3. Figure 3A shows the statistics for the initial image set, and Figure 3B shows the statistics for the noisy image set. The mean images are in the top rows of A and B, and the variance images are in the bottom rows. α increases from left to right, the first column corresponding to $\alpha = 0$ and subsequent columns to α equaling 25, 50, 75, 100, 150, and 200 respectively. The left colorbars correspond to the mean images, the right colorbars correspond to the variance images.

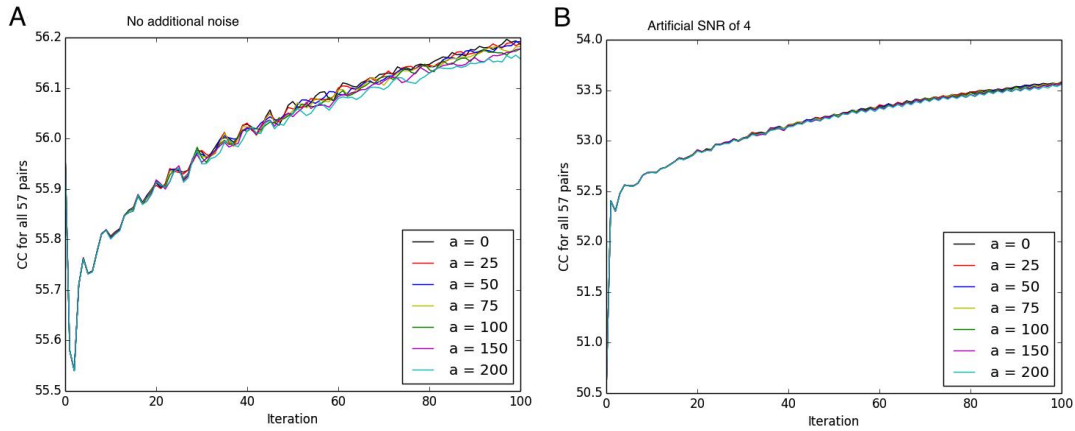


Figure 2. Total Cross Correlation for all 57 image pairs throughout optimization. A: the initial image set, B: with additional Gaussian noise. Note the different scales for A and B.

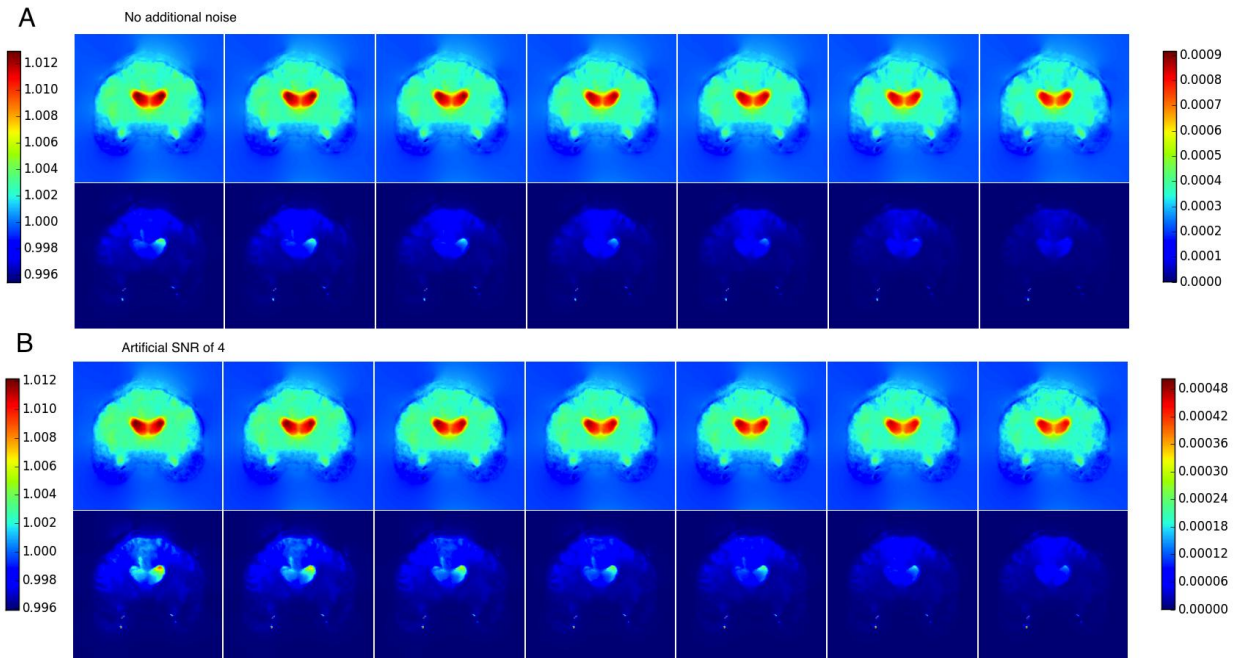


Figure 3. Statistics of Jacobian determinant maps. Columns from left to right correspond to α equals 0, 25, 50, 75, 100, 150, and 200 respectively. The top row of both A and B are the mean images and the bottom row of both A and B are the variance images. The left colorbars are for the mean images, the right colorbars are for the variance images.

The algorithm successfully minimizes the variance of the Jacobian determinant maps at every voxel as α increases. We also see that the variance at many voxels increases somewhat under noisy conditions relative to the original data set, as perfect matching is more challenging in the presence of noise. The averages are consistent across α values, however have slightly more defined regions and smaller magnitude as α increases.

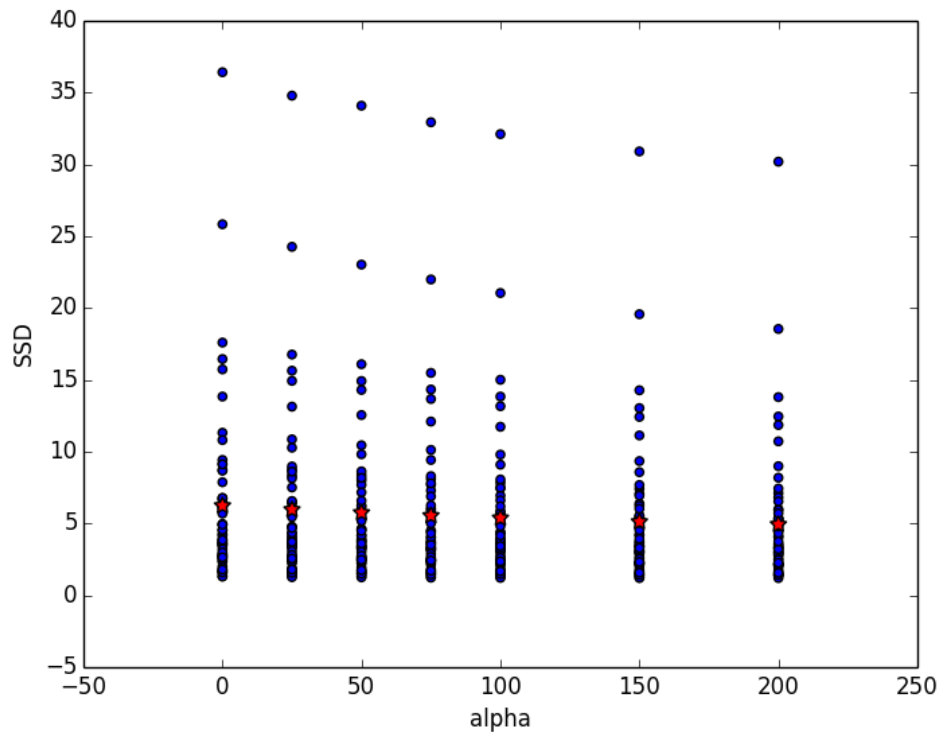


Figure 4. SSD between Jacobian maps learned under noisy and initial conditions for all α values

Finally, the results of calculating (12) for all image pairs at all α levels are shown in Figure 4. Each blue dot corresponds to one image pair registered at the corresponding α value. Most voxels in a Jacobian determinant image are very close to 1.0, so small differences such as 0.05 often correspond to clinically relevant differences in volumetric change in brain tissue (a 5 percent tissue loss would also be important to detect). There is a visibly clear reduction in this SSD as α increases.

To test if the reduction in SSD values with increasing α was due to chance, a one-sided Student's t-test was performed between the $\alpha = 0$ data and the data sets for all other α values respectively. The results are presented in Table 1. The improvement in Jacobian determinant map estimation in the presence of noise was statistically significant for all values of α with p-values less than $1e-8$.

α	25	50	75	100	150	200
μ	0.27	0.48	0.69	0.86	1.12	1.38
σ	0.32	0.55	0.77	0.95	1.23	1.43
T	6.36	6.65	6.78	6.80	6.83	6.90
p	2e-8	6.5e-9	4e-9	3.7e-9	3.3e-9	2.5e-9

Table 1. t-test results for the improvement in estimation of Jacobian determinant maps under noisy conditions; significant results are bold.

4. DISCUSSION

First we address the reduction in final cross-correlation value with increasing α . This is expected, as we compromise the exact matching of the template to the reference image in favor of some level of groupwise consistency among the log Jacobian determinant maps. In our model, this is the appropriate thing to do, as an exact matching between template and reference images would have registered not only the anatomy, but whatever noise was corrupting it as well. We hypothesized that most of the CC lost due to the groupwise consistency term was overfitting to the noise in the images. Figure 4 and Table 1 suggest this hypothesis was reasonable and warrant its further exploration.

Considering the results in Figure 4 and Table 1, we interpret the change in the mean Jacobian determinant images with α , presented in Figure 3, as an enrichment for change caused by actual biological processes present in most of the 57 image pairs. The slight reduction in magnitude may be due to the loss of components of the transformation that were solely due to noise. Also, the boundaries of anatomy are somewhat more clear in the mean images corresponding to increased α , so those images may be a better measure of biological change.

It is very important to mention that Jacobian determinant maps estimated using this method should only be used for voxel-based analysis studies with extreme care. Equation (5) intentionally minimizes the variance in the log Jacobian determinant values at every voxel, and statistics dependent on that variance are biased.¹² It is still possible to use Jacobian determinant maps learned in this way, if one empirically determines the null distribution of the variance values, and uses it to correct any statistics for the bias. However, the most obvious applications for registration via this method lie elsewhere.

The most immediate extension of the work is to put the registration framework into the diffeomorphic setting.^{13,14} In that case, for each image pair we truly estimate a geodesic on a manifold of diffeomorphisms (specified by the choice of metric) that ideally passes through the identity and a transformation that perfectly matches the template to the reference image. In that case, our example presented in figure 1 is even more applicable, as we are estimating the “slope” of the geodesic, given by the initial momentum. In that case, predictions for future unobserved time points can be made by shooting the geodesic given by the estimated momentum further in time. For predictions far in time from the initial data, the accuracy of the momentum estimation is very important, as shown in Figure 1. This method may make such predictions more accurate and clinically useful, and ease the inference of clinically relevant information from data collected sooner or closer together in time.

Finally, there is a close connection between the proposed method and empirical Bayes (EB) methods.¹⁵ In EB methods, we know the form of a prior distribution over our model parameters, but we do not know the parameters for the prior itself. Those parameters are instead estimated from the data, and the estimated prior is then used to constrain the fitting of individual models. In such methods, the prior estimation is independent of the model fitting. In our case, we have assumed there is a prior distribution for log Jacobian determinant maps (or momenta in the diffeomorphic setting) for the images we are working with. We would like to constrain the estimation of our Jacobian determinant maps with this prior, but we do not have an explicit representation of it. In lieu of having a given prior, we use the average log Jacobian determinant map, and constrain the estimation of individual log Jacobian maps to be similar to it. Future work will explore the connection between the proposed method and the EB framework.

5. CONCLUSIONS

We demonstrated a mathematical framework to couple the registration of N longitudinal image pairs. We hypothesized that by sharing information, the estimation of Jacobian determinant maps would be less prone to over fitting due to noise. We conducted an experiment, the results of which support that hypothesis and suggest further investigation into groupwise methods and priors for longitudinal registration.

6. ACKNOWLEDGEMENTS

The authors thank Dr. James Ralston for help with the derivation of (6). This work is supported in part by NIH grants including the U54 Big Data to Knowledge program, and funding from the NIA and NIMH

REFERENCES

- [1] Thompson, P., Hayashi, K., de Zubicaray, G., Janke, A., Rose, S., Semple, J., Herman, D., Hong, M., Dittmer, S., Doddrell, D., and Toga, A., “Dynamics of gray matter loss in alzheimer’s disease,” *Journal of Neuroscience* **23**(3), 994–1005 (2003).
- [2] Braak, H. and Braak, E., “Neuropathological staging of alzheimer-related changes,” *Acta Neuropathologica* **82**(4), 239–259 (1991).
- [3] Hua, X., Hibar, D., Ching, C., Boyle, C., Rajagopalan, P., Gutman, B., Leow, A., Toga, A., Jack, C. J., Harvey, D., Weiner, M., Thompson, P., and Initiative, A. D. N., “Unbiased tensor-based morphometry: improved robustness and sample size estimates for alzheimer’s disease clinical trials,” *Neuroimage* **66**, 648–661 (2013).
- [4] Hua, X., Leow, A., Parikshak, N., Lee, S., Chiang, M., Toga, A., Jack, C. J., Weiner, M., Thompson, P., and Initiative, A. D. N., “Tensor-based morphometry as a neuroimaging biomarker for alzheimer’s disease: an mri study of 676 ad, mci, and normal subjects,” *Neuroimage* **43**(3), 458–469 (2008).
- [5] Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., and Ayache, N., “Riemannian elasticity: A statistical regularization framework for non-linear registration,” **8**(2), 943–950, MICCAI (2005).
- [6] Wei, L., Awate, S., Anderson, J., and Fletcher, T., “A functional network estimation method of resting-state fmri using a hierarchical markov random field,” *Neuroimage* **In press** (June, 2014).
- [7] Jolliffe, I., [*Principal Component Analysis*], Springer Verlag (1986).
- [8] Fletcher, E., Knaack, A., Singh, B., Lloyd, E., Wu, E., Carmichael, O. T., and DeCarli, C., “Combining boundary-based methods with tensor-based morphometry in the measurement of longitudinal brain change,” *IEEE Trans. Med. Imaging* **32**(2), 223–236 (2013).
- [9] “Adni preprocessing protocol description.” <http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing>.
- [10] Iglesias, J., Liu, C., Thompson, P., and Tu, Z., “Robust brain extraction across datasets and comparison with publicly available methods,” *IEEE Transactions on Medical Imaging* **30**(9), 1617–1634 (2011).
- [11] Yanovsky, I., Thompson, P. M., Osher, S., and Leow, A. D., “Topology preserving log-unbiased nonlinear image registration: Theory and implementation,” in [*CVPR*], IEEE Computer Society (2007).

- [12] Tustison, N. J., Avants, B. B., Cook, P. A., Kim, J., Whyte, J., Gee, J. C., and Stone, J. R., “Logical circularity in voxel-based analysis: Normalization strategy may induce statistical bias,” *Hum Brain Mapp* (Nov 2012).
- [13] Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J., “Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation,” *Int. J. Comput. Vision* **97**, 229–241 (Apr. 2012).
- [14] Niethammer, M., Huang, Y., and Vialard, F.-X., “Geodesic regression for image time-series,” in [*MICCAI (2)*], Fichtinger, G., Martel, A. L., and Peters, T. M., eds., *Lecture Notes in Computer Science* **6892**, 655–662, Springer (2011).
- [15] James, W. and 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*], 361–379, University of California Press, Berkeley, Calif. (1961).

7. APPENDIX

Derivation of equation (6): We take the first variation of (5) w.r.t. the k th transformation ϕ_k :

$$\frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \mathcal{P}[\phi_0, \dots, \phi_k + \epsilon h, \dots, \phi_{n-1}] = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \text{Trace}(AA^T) =$$

$$\frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \underbrace{\sum_{i=0}^{n-1} \int_{\Omega} [\log(\det(J_{\phi_i}^{m dt})) - \underbrace{\frac{1}{n} \sum_{j=0}^{n-1} \log(\det(J_{\phi_j}^{m dt}))}_{\mu}]^2 dx}_C$$

$$\frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \sum_{i=0}^{n-1} \int_{\Omega} [C]^2 dx = \sum_{i=0}^{n-1} \int_{\Omega} 2C \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} C dx \quad (13)$$

$$\frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} C = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} [\log(\det(J_{\phi_i}^{m dt})) - \mu] =$$

$$\frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \log(\det(J_{\phi_i}^{m dt})) + \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \mu \quad (14)$$

$$\frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \mu = \frac{\partial}{\partial \epsilon} \Big|_{\epsilon=0} \frac{1}{n} \sum_{j=0}^{n-1} \log(\det(J_{\phi_j}^{m dt})) = \quad (15)$$

$$\frac{1}{n} \frac{1}{\det(J_{\phi_k}^{m dt})} \det(J_{\phi_k}^{m dt}) \text{Trace}[J_{\phi_k}^{m dt} J_h] = \frac{1}{n} \text{Trace}[J_{\phi_k}^{m dt} J_h] \quad (16)$$

If $i = k$, then:

$$\left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} \log(\det(J_{\phi_i}^{mdt})) = \text{Trace}[J_{\phi_k}^{-1} J_h] \quad (17)$$

If $i \neq k$, then:

$$\left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} \log(\det(J_{\phi_i}^{mdt})) = 0$$

Substituting (16) and (17) back into (14) and in turn substituting the result into (13), we get:

$$\begin{aligned} \left. \frac{\partial}{\partial \epsilon} \right|_{\epsilon=0} \text{Trace}(AA^T) &= \frac{-1}{n} \sum_{i=0, i \neq k}^{n-1} \int_{\Omega} 2C \text{Trace}[J_{\phi_i}^{mdt} J_h] dx + \int_{\Omega} 2C_k \left(1 - \frac{1}{n}\right) \text{Trace}[J_{\phi_k}^{mdt} J_h] dx \\ &= \frac{-1}{n} \sum_{i=0}^{n-1} \int_{\Omega} 2C \text{Trace}[J_{\phi_i}^{mdt} J_h] dx + \int_{\Omega} 2C_k \text{Trace}[J_{\phi_k}^{mdt} J_h] dx \end{aligned}$$

Here, we require the multidimensional integration by parts formula:

$$\int_{\Omega} \nabla u \cdot \mathbf{v} \, d\Omega = \int_{\Gamma} u(\mathbf{v} \cdot \mathbf{n}) \, d\Gamma - \int_{\Omega} u \nabla \cdot \mathbf{v} \, d\Omega \quad (18)$$

where Γ is the boundary of Ω and \mathbf{n} is the outward normal. In this case,

$$\begin{aligned} \nabla u &= J_h \implies u = h \\ \mathbf{v} &= \frac{-1}{n} \sum_{i=0}^{n-1} 2C J_{\phi_i}^{mdt} + 2C J_{\phi_k}^{mdt} \end{aligned}$$

we enforce the transformation to be zero on Γ , making the first term on the right hand side of (18) equal to zero. Given that $u = h$, the Euler-Lagrange equations we seek are $-\nabla \cdot \mathbf{v}$:

$$\begin{aligned} -\nabla \cdot \mathbf{v} &= -\nabla \cdot \left[2[\log(\det(J_{\phi_k}^{mdt})) - \mu - \frac{1}{n} \sum_{i=0}^{n-1} (\log(\det(J_{\phi_i}^{mdt})) - \mu)] J_{\phi_k}^{mdt} \right] \\ &\propto -\nabla \cdot \left[[\log(\det(J_{\phi_k}^{mdt})) - \frac{1}{n} \sum_{i=0}^{n-1} \log(\det(J_{\phi_i}^{mdt}))] J_{\phi_k}^{mdt} \right] \\ &= -\nabla \cdot \left[[\log(\det(J_{\phi_k}^{mdt})) - \mu] J_{\phi_k}^{mdt} \right] \end{aligned}$$