Symmetric diffeomorphic modeling of longitudinal structural MRI

John Ashburner * and Gerard R. Ridgway

Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, UK

This technology report describes the longitudinal registration approach that we intend to incorporate into SPM12. It essentially describes a group-wise intra-subject modeling framework, which combines diffeomorphic and rigid-body registration, incorporating a correction for the intensity inhomogeneity artifact usually seen in MRI data. Emphasis is placed on achieving internal consistency and accounting for many of the mathematical subtleties that most implementations overlook. The implementation was evaluated using examples from the *OASIS Longitudinal MRI Data in Non-demented and Demented Older Adults*.

Keywords: diffeomorphisms, geodesic shooting, shape modeling, non-linear registration, longitudinal registration, inverse consistency, symmetry, transitivity

1. INTRODUCTION

Growth, plasticity, aging, and degeneration are inherently longitudinal processes; while much can be learned by studying a cross-sectional sample of subjects at different stages of such processes, longitudinal data have well-established advantages in terms of increasing power and reducing confounds. To give just one example, cross-sectional studies of aging are challenged by the relative subtlety of changes over time compared to the large inter-individual variation, and they can never fully separate true aging effects from confounding effects of birth-year such as changes in nutrition. Longitudinal anatomical MRI provides a framework for characterizing many of the macroscopic brain changes in natural development and in response to disease or injury. A *MEDLINE* search for keywords "longitudinal," "brain," and "MRI" in the topic, gives over 1,500 hits.

Longitudinal data require appropriate statistical models, due to the dependence of repeated measurements within-subject. Beyond this, it is also recognized that there can be substantial gains in power from using image processing and modeling algorithms that are specifically designed for longitudinal data. For example, the boundary shift integral (Freeborough and Fox, 1997; Leung et al., 2012), which provides a direct measure of the difference between two brain volumes, has proven to be substantially more powerful than simple subtraction of the two brain segmentations upon which it is based. However, the reduction in measurement variability comes at a price of increased risk of various forms of bias. In particular, any technique that is not symmetric[1] with respect

to the multiple time-points has the potential to introduce false positive differences.

Concerns about asymmetry in pairwise image registration, related to one image being chosen as a (fixed) reference and the other as a (moving) source, first arose at the turn of the century (Ashburner et al., 1999; Cachier and Rey, 2000; Christensen and Johnson, 2001). In particular, Smith et al. (2001) discuss the potential use of the matrix square root of an affine transformation to derive a "half-way" space between two images[2]. This concept has since been generalized to more than two images, as described further below. Despite this early recognition of the potential problem, many (probably most) studies employing image registration for longitudinal data over the past decade have used methods with some form of asymmetry [usually registering the later time-point(s) to the baseline].

More recently, it has been empirically demonstrated that these theoretical concerns can indeed cause practical problems. Pairwise registration results have been shown to differ depending on which image is chosen as the reference (Thomas et al., 2009; Yushkevich et al., 2010). Results over three time-points appear to exhibit an additive bias, such that the majority of change occurs over the first interval, in a setting where such deceleration is not biologically plausible (Hua et al., 2011; Thompson et al., 2011). Intransitivity (where, for example, the sum of changes from A to B and B to C differs from the change estimated directly from A to C) can also be demonstrated (Leung et al., 2012). The increasingly significant potential of longitudinal MRI for clinical diagnosis, as a biomarker of disease progression and as an outcome measure in treatment trials, has led to an increased focus on this issue (Fox et al., 2011; Reuter and

---

[1] Following Tagare et al. (2009) we prefer the term "symmetric" to alternatives such as "inverse-consistent" because of its formal mathematical definition, which includes not only invariance to swapping a pair of images, but also invariance to permutations of more than two images.

[2] See also the FSL SIENA tool, www.fmrib.ox.ac.uk/fsl/siena/index.html

Fischl, 2011), along with on-going controversy (Holland et al., 2012).

Many of the solutions proposed for avoiding bias from asymmetry are restricted to two time-points, such as the formulation of Tagare et al. (2009). There are three distinct approaches for attempting to address asymmetry and intransitivity across more than two time-points: all time-points of all subjects can be treated independently, i.e., processed cross-sectionally (e.g., Giedd et al., 1999); each time-point can be registered to every other time-point within-subject, with appropriate adjustments to ensure consistency (e.g., Leung et al., 2012); or all time-points can be registered to some form of within-subject average image (e.g., Skrinjar et al., 2008; Reuter et al., 2012a). The first of these results in low or zero bias but very high variance, while the second is computationally infeasible for high-dimensional diffeomorphic image registration. We therefore propose a solution based on the third approach. Registration to an average is defined as an "indirect" approach by Tagare et al. (2009), in contrast to "direct" approaches which symmetrize an individual pairwise registration; however, in this work, we attempt to address some of the more subtle differential geometric issues raised by Tagare et al. (2009) in their direct setting. A further novelty of our approach is the unification of rigid registration, diffeomorphic registration, and correction of (differential) intensity inhomogeneity in a single generative model (in the same spirit as Ashburner and Friston, 2005). The algorithms described in this Technology Report will be made available in the Statistical Parametric Mapping (SPM)[3] software with the next major release, SPM12.

2. METHODS
Our proposed model involves combining rigid-body registration, intensity inhomogeneity correction, and non-linear diffeomorphic registration. Because of all the conditional dependencies among model parameters, naive approaches involving pipelines of steps are likely to lead to less optimal solutions. For example, accurate inhomogeneity correction requires the images to be aligned, whereas more accurate alignment is only possible after inhomogeneity correction. Our solution involves performing all steps in an interleaved fashion, such that model parameters are optimized together. Rigid-body and diffeomorphic registration are combined because the boundary conditions of the diffeomorphic registration do not permit rotations of the whole image. In addition, the inclusion of translations in the rigid-body model allows the diffeomorphic model to not penalize pure displacements.

In the remainder of this section, the individual components will be introduced separately, before proceeding to describe the combined model and some of the implementational details. First though, we say something about how the within-subject template space is defined.

2.1. TEMPLATE POSITION AND DIMENSIONS
The SPM software uses the concept of voxel-to-world mappings. These are encoded in the image headers, and are used to determine the real world locations ($\mathbf{y}$) of voxel indices ($\mathbf{x}$). Usually, they are

read from the "sform" fields of the NIfTI header[4], and encode affine transformation matrices ($\mathbf{M}$), such that

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix}. \quad (1)$$

Each of the individual images is assumed to have such a voxel-to-world mapping associated with it ($\mathbf{M}_n$). The first step of the algorithm involves computing suitable dimensions and a voxel-to-world mapping ($\mathbf{M}_\mu$) for the within-subject template. The aim is to have this template in some form of average position such that bias introduced by interpolation is minimized.

As mentioned above, some pairwise registration approaches have suggested using square roots of transformation matrices for determining the half-way position for approximately symmetrizing the registration (Smith et al., 2001; Thomas et al., 2009; Yushkevich et al., 2010). Several forms of transformation are members of matrix Lie groups (Woods, 2003), so it is illustrative to consider one of the simplest Lie groups – the positive scalars under multiplication. As emphasized in Leung et al. (2012), the conventional square root of a positive scalar can be seen as the geometric mean of the scalar and unity (the identity for positive scalars). The simple geometric mean of a set of points can be shown to minimize the sum squared distances from the mean to all points under a logarithmic distance metric; this property defines a Fréchet mean, sometimes known as a Karcher mean, where the latter is only required to be at a local minimum (Woods, 2003). The geometric mean is also an exponential barycenter (Pennec and Arsigny, 2012), in the sense that the signed logarithmic discrepancies from the mean to all other points sum to zero. Considering now the Lie group of three-dimensional rotation matrices [known as SO(3), the special orthogonal group], an appropriate Riemannian distance metric and Fréchet mean (also the exponential barycenter) can be analogously defined (Moakher, 2002). Unfortunately, in the more general groups of rigid-body or affine transformations, it can be shown that no bi-invariant Riemannian metric exists[5].

Three different compromises have been used in the literature to circumvent the lack of bi-invariant metric: Reuter et al. (2012b) compute a simple Euclidean mean of rigid-body transformations before using the singular value decomposition to factor out any resultant zooms and shears; (Leung et al., 2012) use Arsigny's log-Euclidean mean (Arsigny, 2006), which is a distance-minimizing Fréchet mean under a sub-optimal metric that is not bi-invariant; (Woods, 2003) use the exponential barycenter (Pennec and Arsigny, 2012) which corresponds to a bi-invariant mean, relaxing the manifold to a semi-Riemannian one without a true metric (and hence without a Fréchet mean), but still having a set of tangent space vectors from the mean that sum to zero.

---

[3]Available from www.fil.ion.ucl.ac.uk/spm/

[4]See http://nifti.nimh.nih.gov/nifti-1

[5]Bi-invariance requires that composing both transformations with a third should not change their distance, and that inverting both should also preserve the distance; in the simple positive scalar case, the logarithmic distance between 1/2 and 1/4 is the same as that between 1/6 and 1/12, and the same as that between 2 and 4.

Our implementation is also based on the exponential barycenter, which in theory could be directly used to define the average position template. Unfortunately, because voxel sizes are not necessarily isotropic, there are some potential difficulties with using this. In particular, it is possible for the exponential barycenter to be a transform that can not be encoded by axis-aligned scaling (to account for voxel sizes), rotating, and translating (i.e., nine parameters). It would not be possible to encode such a matrix in the "qform" fields of a NIfTI header, and it would also make some of the subsequent processing slightly more involved. Therefore, the closest nine-parameter affine transform to the exponential barycenter is estimated via a Gauss-Newton optimization strategy that minimizes the Euclidean distance between the matrices.

Finally, the dimensions of the template are determined such that it easily covers the field of view of the images. This is achieved by projecting the locations of the corners of the images into the template space and finding the maximum and minimum values. Corresponding adjustments are also needed for $\mathbf{M}_\mu$, which involves changing the translations.

Once the dimensions, etc., of the template are defined, the algorithm can proceed to fit the combined model. The next three subsections outline the individual components, and how they would be optimized in isolation. This will be followed by a subsection about combining the components together into the unified model.

2.2. GROUP-WISE RIGID-BODY REGISTRATION

Each image is considered as a scalar function of space, such that $f_n : \Omega_{f_n} \to \mathbb{R}$, where $\Omega_{f_n} \subseteq \mathbb{R}^3$. A series of N images $\{f_n\}_{n=1}^N$ may be modeled as a common template mean ($\mu : \Omega_\mu \to \mathbb{R}$), which is rotated and translated by a rigid-body transformation. Additive Gaussian noise (of variance $1/\lambda_n$) is assumed to be constant over each image and known, but may vary from image to image.

When dealing with rigid-body transforms, it is useful to consider them in terms of their membership of the special Euclidean group in three dimensions $[SE(3)]$. Within this framework, a transformation matrix ($\mathbf{R_q}$) is constructed via an exponential mapping of the six parameters ($\mathbf{q}$) that constitute the Lie algebra of the group.

$$\mathbf{R_q} = \exp \begin{bmatrix} 0 & q_4 & -q_5 & q_1 \\ -q_4 & 0 & q_6 & q_2 \\ q_5 & -q_6 & 0 & q_3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (2)$$

This involves a matrix exponential, rather than computing the exponential of each element. Although there are many ways to compute this (Moler and Van Loan, 2003), it is usually defined as

$$\exp \mathbf{Q} = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbf{Q}^n \quad (3)$$

Rigid alignment among a set of images can be performed by estimating the optimal mapping between the template and each of the images. A mapping from voxel indices in the template, to those in the nth image, is given by the following affine transform:

$$\boldsymbol{\xi}_{\mathbf{q}_n}(\mathbf{x}) = \mathbf{I}_{3,4} \mathbf{M}_n^{-1} \mathbf{R}_{\mathbf{q}_n} \mathbf{M}_\mu \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}, \text{ where } \mathbf{I}_{3,4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (4)$$

This leads to the following objective function for group-wise rigid-body alignment:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{\lambda_n}{2} \left\| f_n - \mu \left(\boldsymbol{\xi}_{\mathbf{q}_n}^{-1} \right) \right\|^2. \quad (5)$$

Without constraints, this simple approach may lead to a variety of equivalent solutions in which the template is rotated and translated by some arbitrary amount. The Lie algebra of the $SE(3)$ group is used to parameterize the transforms to make is easier factor out the Fréchet mean of all the rigid-body transforms, thus ensuring that the template remains in the average position. All that is required to achieve this is to ensure that $\sum_{n=1}^{N} \mathbf{q}_n = 0$, which can be achieved simply by subtracting the mean after all the $\mathbf{q}_n$ have been re-estimated.

For reasons that should become apparent later, we formulate the objective function as

$$\mathcal{E} = \sum_{n=1}^{N} \frac{\lambda_n}{2} \left| \mathbf{D}\boldsymbol{\xi}_{\mathbf{q}_n} \right| \int_{\mathbf{x} \in \Omega'_n} \left(f_n \left(\boldsymbol{\xi}_{\mathbf{q}_n}(\mathbf{x}) \right) - \mu(\mathbf{x}) \right)^2 d\mathbf{x},$$
$$\text{such that } \sum_{n=1}^{N} \mathbf{q}_n = 0. \quad (6)$$

In the above equation, the $\mathbf{D}$ operator refers to computing the Jacobian, the determinants of which are included to account for a change of variables. The field of view common to both the nth image and the template is denoted by $\Omega'_n = \boldsymbol{\xi}_{\mathbf{q}_n}^{-1}(\Omega_{f_n}) \cap \Omega_\mu$.

The optimization strategy involves alternating between re-estimating the mean (μ), and then sequentially using this to re-estimate the registration parameters. The update of μ is achieved by differentiating equation (6) with respect to μ and solving.

$$\mu(\mathbf{x}) = \frac{\sum_{n=1}^{N} w_n(\mathbf{x}) f_n \left(\boldsymbol{\xi}_{q_n}(\mathbf{x}) \right)}{\sum_{n=1}^{N} w_n(\mathbf{x})}, \text{ where}$$
$$w_n(\mathbf{x}) = \begin{cases} \lambda_n \left| \mathbf{D}\boldsymbol{\xi}_{\mathbf{q}_n} \right| & \text{if } \mathbf{x} \in \Omega'_n \\ 0 & \text{otherwise} \end{cases}. \quad (7)$$

A Gauss-Newton optimization iteration is then done for each of the N scans. Dropping the n subscripts for notational simplicity, this involves the following:

$$\mathbf{q} \leftarrow \mathbf{q} - \left(\left. \frac{\partial^2 \mathcal{E}}{\partial \mathbf{q}^2} \right|_{\mathbf{q}} \right)^{-1} \left(\left. \frac{\partial \mathcal{E}}{\partial \mathbf{q}} \right|_{\mathbf{q}} \right). \quad (8)$$

It requires the vector of first derivatives, which are computed by differentiating equation (5) with respect to $\mathbf{q}$, and applying the same change of variables as earlier (see Appendix A), giving:

$$\frac{\partial \mathcal{E}}{\partial q_i} = \int_{\mathbf{x} \in \Omega'} a(\mathbf{x}) \mathbf{g}(\mathbf{x}) \cdot \mathbf{h}_i(\mathbf{x}) \, d\mathbf{x} \quad (9)$$

where $a = \lambda |\mathbf{D}\boldsymbol{\xi}_{\mathbf{q}}| (f(\boldsymbol{\xi}_{\mathbf{q}}) - \mu)$, $\mathbf{g} = \nabla\mu$ and

$$\mathbf{h}_i(\mathbf{x}) = \mathbf{I}_{3,4} \mathbf{M}_{\mu}^{-1} \mathbf{R}_{\mathbf{q}}^{-1} \frac{\partial \mathbf{R}_{\mathbf{q}}}{\partial q_i} \mathbf{M}_{\mu} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \quad (10)$$

For a more stable algorithm, the matrix of second derivatives (Hessian matrix) should be positive definite. A suitable positive definite approximation to the Hessian (see Appendix A) is:

$$\frac{\partial^2 \mathcal{E}}{\partial q_i \partial q_j} \simeq \int_{\mathbf{x} \in \Omega'} w(\mathbf{x}) \left(\mathbf{g}(\mathbf{x}) \cdot \mathbf{h}_i(\mathbf{x}) \right) \left(\mathbf{g}(\mathbf{x}) \cdot \mathbf{h}_j(\mathbf{x}) \right) d\mathbf{x}. \quad (11)$$

Our implementation makes use of $\frac{\partial \mathbf{R}_{\mathbf{q}}}{\partial q_i}$, obtained by differentiating equation (3), although this could also have been computed numerically using finite differences or by more elegant methods (Al-Mohy and Higham, 2009). The overall optimization scheme is presented in Algorithm 1.

Algorithm 1 | Rigid-body registration.

---

for $n = 1...N$ **do**
   Initialize $\mathbf{q}_n$.
end for
repeat
   Compute μ and $\nabla\mu$ (equation 7).
   for $n = 1...N$ **do**
      Compute 1st derivatives (equation 9).
      Compute Hessian (equation 11).
      Gauss-Newton update of $\mathbf{q}_n$ (equation 8).
   end for

   $\bar{\mathbf{q}} \leftarrow \frac{1}{N} \sum_{n=1}^{N} \mathbf{q}_n$.

   for $n = 1...N$ **do**
      $\mathbf{q}_n \leftarrow \mathbf{q}_n - \bar{\mathbf{q}}$.
   end for
until convergence

2.3. GROUP-WISE INHOMOGENEITY CORRECTION

MRI scans are usually corrupted by a spatially smooth intensity non-uniformity (also known as "inhomogeneity" and sometimes referred to as "bias"), which is often corrected prior to image registration using a procedure such as N3 (Sled et al., 1998). Instead, we propose that the additional internal consistency from incorporating the non-uniformity correction within the longitudinal registration scheme should provide potentially more accurate results. Studholme et al. (2004), Modersitzki (2006), and Modat et al. (2010) previously used registration-based non-uniformity corrections, although those approaches did not consider aspects of inverse (or group-wise) consistency.

Here, we outline a general strategy for dealing with non-uniformity fields in aligned images. This part of the model assumes that a series of N aligned images (f_n) may be modeled as a common template mean (μ), scaled by non-uniformity fields. Because these fields should be positive, they are modeled using an exponential of

a Gaussian process (e^{b_n}), where $b_n : \Omega_{f_n} \to \mathbb{R}$. This model leads to the following objective function.

$$\mathcal{E} = \sum_{n=1}^{N} \left(\frac{\lambda_n}{2} \left\| f_n - \mu e^{b_n} \right\|^2 + \frac{1}{2} \left\| L_b b_n \right\|^2 \right) \quad (12)$$

Note that L_b is a differential operator that penalizes the roughness of the (logarithms of the) estimated non-uniformity fields. A Laplacian is used in practice, although the optimal choice of differential operator will depend on the nature of the artifacts in the image data.

Regularized maximum likelihood (or maximum *a posteriori*) optimization of the inhomogeneity fields may be achieved by minimizing the above function. As in the rigid registration case, we propose alternating between re-estimating the mean and then using this to re-estimate the inhomogeneity fields.

The update of μ is achieved by differentiating equation (12) with respect to μ, and solving.

$$\mu = \frac{\sum_{n=1}^{N} \lambda_n f_n e^{b_n}}{\sum_{n=1}^{N} \lambda_n e^{2b_n}} \quad (13)$$

Gradient descent could then be used to update the inhomogeneity fields, which requires the first derivatives of the objective function. These are computed via their Gâteaux differential.

$$d\mathcal{E}(b_n; h) = \frac{d}{d\tau} \frac{1}{2} \left(\lambda_n \left\| f_n - \mu e^{b_n + \tau h} \right\|^2 + \left\| L_b (b_n + \tau h) \right\|^2 \right) \Bigg|_{\tau=0}$$

$$= \int_{\mathbf{x} \in \Omega_{\mu}} -a_n(\mathbf{x}) \mu(\mathbf{x}) h(\mathbf{x}) d\mathbf{x} + \left\langle L_b^{\dagger} L_b b_n, h \right\rangle \quad (14)$$

where $a_n = \lambda_n e^{b_n} \left(f_n - \mu e^{b_n} \right)$.

In our implementation, the fields are updated via a Gauss-Newton step, which makes use of a positive definite approximation to the second derivatives (derived using similar principles to those in Appendix A).

$$d^2 \mathcal{E}(b_n; h_1, h_2) = \frac{d^2}{d\tau_1 d\tau_2} \frac{1}{2} \left(\lambda_n \left\| f_n - \mu e^{b_n + \tau_1 h_1 + \tau_2 h_2} \right\|^2 \right.$$

$$\left. + \left\| L_b \left(b_n + \tau_1 h_1 + \tau_2 h_2 \right) \right\|^2 \right) \Bigg|_{\tau_1=0, \tau_2=0}$$

$$\simeq \int_{\mathbf{x} \in \Omega_{\mu}} w_n(\mathbf{x}) \mu(\mathbf{x})^2 h_1(\mathbf{x}) h_2(\mathbf{x}) d\mathbf{x} + \left\langle L_b^{\dagger} L_b h_1, h_2 \right\rangle \quad (15)$$

where $w_n = \lambda_n e^{2b_n}$.

The overall procedure is summarized in Algorithm 2, which has a similar overall structure to Algorithm 1. The non-uniformity fields are encoded by a parameter at each voxel, such that continuous representations [of $b(\mathbf{x})$] may be obtained via tri-linear interpolation. This parameterization results in a diagonal matrix for the first term in equation (15), making it relatively straightforward to solve the system of linear equations required for the update via a full multi-grid (FMG) method.

Algorithm 2 | Non-uniformity field estimation.

```
for n = 1...N do
    Initialize b_n.
end for
repeat
    Compute μ (equation 13).
    for n = 1...N do
        Compute 1st derivatives (equation 14).
        Compute Hessian (equation 15).
        Gauss-Newton update of b_n (via FMG).
    end for
    b̄ ← (1/N) Σ_{n=1}^{N} b_n.
    for n = 1...N do
        b_n ← b_n − b̄.
    end for
until convergence
```

Although the model appears to incorporate a number of unnecessary parameters (i.e., N non-uniformity fields plus a mean image), it effectively involves only $N − 1$ fields. The mean image is incorporated for convenience, and the geometric mean over all the estimated non-uniformity fields converges to one (given appropriate regularization). In practice, this constraint may be incorporated in the optimization (as shown in Algorithm 2), which enhances convergence. When there are two images, we may assume $b_2 = −b_1$, in which case the objective function reduces to

$$\mathcal{E} = \frac{\lambda_1 \lambda_2}{2} \int_{\mathbf{x} \in \Omega} \frac{\left(f_1(\mathbf{x}) e^{b_1(\mathbf{x})} − f_2(\mathbf{x}) e^{−b_1(\mathbf{x})}\right)^2}{\left(\lambda_1 e^{2b_1(\mathbf{x})} + \lambda_2 e^{−2b_1(\mathbf{x})}\right)} d\mathbf{x} + \|L_b b_1\|^2. \tag{16}$$

This special case may have other applications, such as the computation of smooth ratios of MR scans for mapping of RF transmit fields (Lutti et al., 2010).

2.4. GROUP-WISE DIFFEOMORPHIC REGISTRATION

From a generative modeling perspective, the objective function for group-wise registration may be written as

$$\mathcal{E} = \sum_{n=1}^{N} \left(\frac{\lambda_n}{2} \left\| f_n − \mu \circ \boldsymbol{\phi}_{\mathbf{v}_n}^{-1} \right\| \right)^2 + \frac{1}{2} \left\| \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n \right\|^2 \tag{17}$$

where $\boldsymbol{\phi}_{\mathbf{v}_n} : \Omega_\mu \to \Omega_\mu$ is a diffeomorphic mapping, constructed from the initial velocity field $\mathbf{v}_n : \Omega_\mu \to \mathbb{R}^3$. Each image ($f_n$) is assumed to be a warped version of a template (μ), with added noise. Smoothness of the diffeomorphisms are achieved via the differential operator $\mathbf{L}_{\mathbf{v}}$, which may differ from image to image.

It is now reasonably well known that a diffeomorphism ($\boldsymbol{\phi}_{\mathbf{v}}$) may be computed from an initial velocity field ($\mathbf{v}$) using a procedure known as geodesic shooting (Miller et al., 2006). Essentially, this procedure is based on integrating a particular form of dynamical system over unit time, and relies upon the principle of conservation of momentum. The procedure begins by initializing $\boldsymbol{\phi}_{\mathbf{v}}$

to the identity transform and computing the initial momentum from the initial velocity via

$$\mathbf{u} = \mathbf{L}_{\mathbf{v}}^\dagger \mathbf{L}_{\mathbf{v}} \mathbf{v}. \tag{18}$$

Then the following dynamical system is integrated over unit time.

$$\dot{\boldsymbol{\phi}}_{\mathbf{v}} = \left(\mathbf{K}_{\mathbf{v}} \left(\left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}}^{-1}\right| \left(\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}}^{-1}\right)^T \left(\mathbf{u} \circ \boldsymbol{\phi}_{\mathbf{v}}^{-1}\right)\right)\right) \circ \boldsymbol{\phi}_{\mathbf{v}} \tag{19}$$

Briefly, the initial momentum is re-sampled according to the inverse of the current estimate of $\boldsymbol{\phi}_{\mathbf{v}}$. Each point in the resulting field is matrix-multiplied by the transpose of the Jacobian tensor at that point of $\boldsymbol{\phi}_{\mathbf{v}}^{-1}$ and rescaled by the determinant of the Jacobian. Finally, the result is smoothed by applying the $\mathbf{K}_{\mathbf{v}}$ operator, which is the Green's function of $\mathbf{L}_{\mathbf{v}}^\dagger \mathbf{L}_{\mathbf{v}}$ (see Bro-Nielsen and Gramkow, 1996), to give the velocity field that provides the next update for $\boldsymbol{\phi}_{\mathbf{v}}$. The $\mathbf{K}_{\mathbf{v}}$ operator may be viewed as a low pass filter, which is the (pseudo-) inverse of $\mathbf{L}_{\mathbf{v}}^\dagger \mathbf{L}_{\mathbf{v}}$, such that $\mathbf{K}_{\mathbf{v}} \mathbf{L}_{\mathbf{v}}^\dagger \mathbf{L}_{\mathbf{v}} \mathbf{v} = \mathbf{v}$. In practice, we use a slightly different integration scheme, which was described in Ashburner and Friston (2011).

We now re-write Equation (17) via integration by substitution.

$$\mathcal{E} = \sum_{n=1}^{N} \frac{\lambda_n}{2} \int_{\mathbf{x} \in \Omega_\mu} \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x})\right| \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x}) − \mu(\mathbf{x})\right)^2 d\mathbf{x}$$
$$+ \frac{1}{2} \sum_{n=1}^{N} \left\| \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n \right\|^2 \tag{20}$$

The template update equation is obtained by differentiating equation (20) with respect to μ and solving.

$$\mu = \frac{\sum_{n=1}^{N} \lambda_n \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}\right| \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n}\right)}{\sum_{n=1}^{N} \lambda_n \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}\right|} \tag{21}$$

Registration is treated as an optimization procedure, using both first and second derivatives. The first Gâteaux differential is computed via

$$d\mathcal{E}(\mathbf{v}_n; \mathbf{h}) = \int_{\mathbf{x} \in \Omega_\mu} a_n(\mathbf{x}) \mathbf{g}(\mathbf{x}) \cdot \mathbf{h}(\mathbf{x}) d\mathbf{x} + \left\langle \mathbf{L}_{\mathbf{v}_n}^\dagger \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n, \mathbf{h} \right\rangle \tag{22}$$

where $a_n = \lambda_n \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}\right| \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n} − \mu\right)$. The template gradients ($\mathbf{g}$) used for the registration are computed (see Appendix B) as follows.

$$\mathbf{g} = \frac{\sum_{n=1}^{N} \lambda_n \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}\right| \nabla \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n}\right)}{\sum_{n=1}^{N} \lambda_n \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}\right|} \tag{23}$$

A Gauss-Newton step is used to update the estimates of the initial velocity, which also requires a positive definite approximation to the second derivatives.

$$d^2\mathcal{E}(\mathbf{v}_n; \mathbf{h}_1, \mathbf{h}_2) \simeq \int_{\mathbf{x} \in \Omega_\mu} w_n(\mathbf{x}) \left(\mathbf{g}(\mathbf{x}) \cdot \mathbf{h}_1(\mathbf{x})\right) \left(\mathbf{g}(\mathbf{x}) \cdot \mathbf{h}_2(\mathbf{x})\right) d\mathbf{x}$$
$$+ \left\langle \mathbf{L}_{\mathbf{v}_n}^\dagger \mathbf{L}_{\mathbf{v}_n} \mathbf{h}_1, \mathbf{h}_2 \right\rangle \tag{24}$$

where $w_n = \lambda_n \left|\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}\right|$.

Further practical details about the implementation may be found in Ashburner and Friston (2011), with some explanation of how the system of linear equations are solved via a full multi-grid approach in Ashburner (2007).

Algorithm 3 | Diffeomorphic registration.

```
for n = 1...N do
    Initialize vₙ.
end for
repeat
    for n = 1...N do
        Compute φ_{vₙ} and Dφ_{vₙ} from vₙ using
    geodesic shooting.
    end for
    Compute mean and gradients
    [equations (21) and (23)].
    for n = 1...N do
        Compute 1st derivatives (equation 22).
        Compute Hessian (equation 24).
        Gauss-Newton update of vₙ (via FMG).
    end for
    ū ← 1/N ∑_{n=1}^N (L†_{vₙ}L_{vₙ}) vₙ.
    for n = 1...N do
        vₙ ← vₙ − K_{vₙ}ū.
    end for
until convergence
```

2.5. COMBINING THE COMPONENTS

The registration procedure combines group-wise rigid-body and diffeomorphic registration with intensity non-uniformity correction. The model assumes that each image (f_n) is a deformed version of a template image (μ), scaled by the exponential of an inhomogeneity field (b_n), with a known amount of additive i.i.d. Gaussian noise (precision λ_n). Fitting this model (see **Figure 1**) involves minimizing the following objective function:

$$\mathcal{E} = \sum_{n=1}^N \frac{1}{2} \left(\int_{\mathbf{x} \in \Omega'_n} \lambda_n \left| \mathbf{D}\boldsymbol{\varphi}_n(\mathbf{x}) \right| \left(f'_n(\mathbf{x}) - \mu(\mathbf{x}) e^{b'_n(\mathbf{x})} \right)^2 d\mathbf{x} \right.$$
$$\left. + \left\| \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n \right\|^2 + \left\| L_b b_n \right\|^2 \right) \tag{25}$$

with the following definitions:

$$\boldsymbol{\varphi}_n = \boldsymbol{\xi}_{\mathbf{q}_n} \circ \boldsymbol{\phi}_{\mathbf{v}_n}$$
$$f'_n = f_n(\boldsymbol{\varphi}_n)$$
$$b'_n = b_n(\boldsymbol{\varphi}_n)$$
$$\Omega'_n = \boldsymbol{\varphi}_n^{-1}(\Omega_{f_n}) \cap \Omega_\mu$$

The images (f) and template (μ) are treated as continuous functions, but are actually encoded via B-spline basis functions. This is done to achieve continuous spatial gradients, which are needed for the registration. Velocity ($\mathbf{v}$) and logarithms of inhomogeneity fields (b) are also treated as spatially continuous, and represented using tri-linear interpolation.
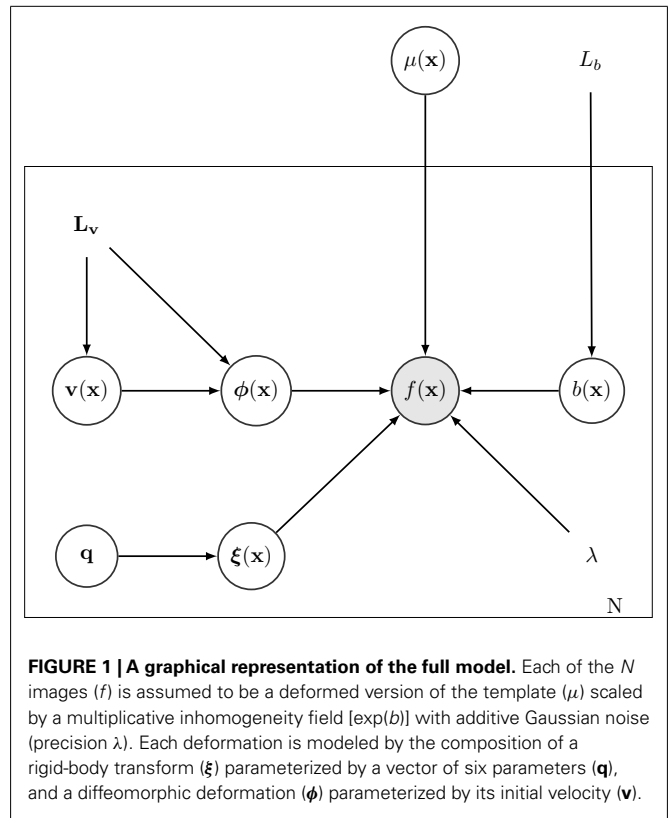
FIGURE 1 | A graphical representation of the full model. Each of the N images (f) is assumed to be a deformed version of the template (μ) scaled by a multiplicative inhomogeneity field [exp(b)] with additive Gaussian noise (precision λ). Each deformation is modeled by the composition of a rigid-body transform ($\boldsymbol{\xi}$) parameterized by a vector of six parameters ($\mathbf{q}$), and a diffeomorphic deformation ($\boldsymbol{\phi}$) parameterized by its initial velocity ($\mathbf{v}$).

We now introduce a few more definitions that will be useful later. The mean image is needed for all steps, and is computed by

$$\mu = \frac{\sum_{n=1}^N w_n e^{-b'_n} f'_n}{\sum_{n=1}^N w_n}, \tag{26}$$

where

$$w_n(\mathbf{x}) = \begin{cases} \lambda_n \left| \mathbf{D}\boldsymbol{\varphi}_n(\mathbf{x}) \right| e^{2b'_n(\mathbf{x})} & \text{if } \mathbf{x} \in \Omega'_n \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

In addition, the following gradients are required for driving the image registration

$$\mathbf{g} = \frac{\sum_{n=1}^N w_n e^{-b'_n} \left(\nabla f'_n + f'_n \nabla b'_n \right)}{\sum_{n=1}^N w_n}$$
$$- \frac{2 \left(\sum_{n=1}^N w_n e^{-b'_n} f'_n \right) \left(\sum_{n=1}^N w_n \nabla b'_n \right)}{\left(\sum_{n=1}^N w_n \right)^2}. \tag{28}$$

One other definition is

$$a_n = w_n \left(f'_n e^{-b'_n} - \mu \right). \tag{29}$$

The derivatives used for updating the rigid-body and diffeomorphic transformations are computed by substituting the above expressions. To compute the first derivatives, equations (28) and (29) would be substituted into equations (9) and (22). For the

approximate second derivatives, equations (27) and (28) would be substituted into equations (11) and (24).

The main difficulty arises from combining the inhomogeneity estimation with the registration. Others have used an elegant metamorphosis approach (Trouvé and Younes, 2005) for integrating intensity variations with deformations, but this is not appropriate here. In the current model, intensity variations are assumed to be a function of the MR scanner rather than an intrinsic property of the brain itself. Therefore, the inhomogeneity fields are estimated in the space of the original images. Our registration scheme involves

---

Algorithm 4 | Combined model estimation.

```
for  n = 1...N do
    Initialize qₙ, bₙ and vₙ.
end for
repeat
    for  n = 1...N do
        Compute φᵥₙ and Dφᵥₙ from vₙ
  (see Ashburner and Friston (2011)).
    end for
    Re-compute μ and g
  [equations (26) and (28)].
    for  n = 1...N do
        Compute ∂E/∂q (equation 9,
  using equation 29).
        Compute ∂²E/∂q² (equation 11,
  using equation 27).
        Gauss-Newton update of qₙ
  (equation 8).
    end for
    q̄ ← 1/N Σⁿ₌₁ᴺ qₙ.
    for  n = 1...N do
        qₙ ← qₙ − q̄.
    end for
    Re-compute μ (equation 26).
    for  n = 1...N do
        Compute dE(bₙ;h) (equation 30, using
  equation 29).
        Compute d²E(bₙ;h₁,h₂) (equation 31,
  using equation 27).
        Gauss-Newton update of bₙ.
    end for
    Re-compute μ and g (equations 26 and 28).
    for  n = 1...N do
        Compute dE(vₙ;h) (equation 22,
  using equation 27).
        Compute d²E(vₙ;h₁,h₂) (equation 24,
  using equation 27).
        Gauss-Newton update of vₙ.
    end for
    ū ← 1/N Σⁿ₌₁ᴺ (L†ᵥₙLᵥₙ) vₙ.
    for  n = 1...N do
        vₙ ← vₙ − Kᵥₙ ū.
    end for
until convergence
```

---

estimating mappings from the mean template to the individual scans. We can re-sample the original scans to bring them in alignment with the template, but sampling the template to align it with the individual scans would require additionally computing the inverse deformation, adding an additional level of computational complexity. Our solution is to first compute derivatives of the data term in the space of the template, making use of the Jacobians of the deformations in a substitution of variables.

Equations (14) and (15) may be re-written to incorporate equations (27) and (29).

$$dE(b_n; h) = \int_{\mathbf{x} \in \Omega_{f_n}} - \left| \mathbf{D}\boldsymbol{\varphi}_n^{-1}(\mathbf{x}) \right| (a_n \mu) \circ \boldsymbol{\varphi}_n^{-1}(\mathbf{x}) \, h(\mathbf{x}) \, d\mathbf{x}$$
$$+ \left\langle L_b^\dagger L_b b_n, h \right\rangle \tag{30}$$

$$d^2E(b_n; h_1, h_2) \simeq \int_{\mathbf{x} \in \Omega_{f_n}} \left| \mathbf{D}\boldsymbol{\varphi}_n^{-1}(\mathbf{x}) \right| (w_n \mu^2) \circ \boldsymbol{\varphi}_n^{-1}(\mathbf{x})$$
$$h_1(\mathbf{x}) \, h_2(\mathbf{x}) \, d\mathbf{x} + \left\langle L_b^\dagger L_b h_1, h_2 \right\rangle \tag{31}$$

This enables the derivatives to be first computed in the space of the template, and subsequently pushed forward to the space of the original images. The inhomogeneity fields are then re-estimated by an iteration of Gauss-Newton, although they are not mean-corrected.

The overall optimization scheme is shown in Algorithm 4.

2.6. IMPLEMENTATION DETAILS

Our implementation is written in a mixture of MATLAB and C code (mex files for the computationally expensive parts). For additional speed (and accuracy), the overall procedure is run over multiple spatial scales, beginning at the lowest resolution. At each scale, a solution is computed, which is prolonged to the next scale where it serves as a starting estimate for the next set of iterations at a higher resolution.

The implementation has large memory requirements, which are likely to exceed the addressable memory of 32-bit computers. To save some memory, many of the computations are done using single precision floating point. Briefly, the main memory consumption comes from:

- Image data (f_n). If each image contains J voxels, the memory for all images will be $4JN$ bytes. For example, a $256 \times 256 \times 256$ image (where $J = 16777216$) requires 64 MB to represent it as single precision floating point.
- Inhomogeneity fields (b_n) require $4JN$ bytes.
- Template (μ) and its spatial gradients (**g**) require $4J_\mu + 12J_\mu$ bytes, where J_μ is the number of voxels in the template.
- Velocity fields ($\mathbf{v}_n$) require $12J_\mu N$ bytes.
- Deformation fields ($\boldsymbol{\phi}_{\mathbf{v}_n}$) require $12J_\mu N$ bytes.
- Jacobian fields ($\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}$) require $36J_\mu N$ bytes.

The geodesic shooting step consumes a lot of additional memory. This includes $36J_\mu$ bytes for the Fourier transform of the Green's function (**K**), $12J_\mu$ bytes for the initial momentum ($\mathbf{u}_n$), plus $96J_\mu$ bytes for composing diffeomorphisms and their Jacobians. The maximum requirement at any point is just over $144J_\mu + N(8J + 60J_\mu)$ bytes.

The algorithm requires a few user-defined settings, which will now be outlined.
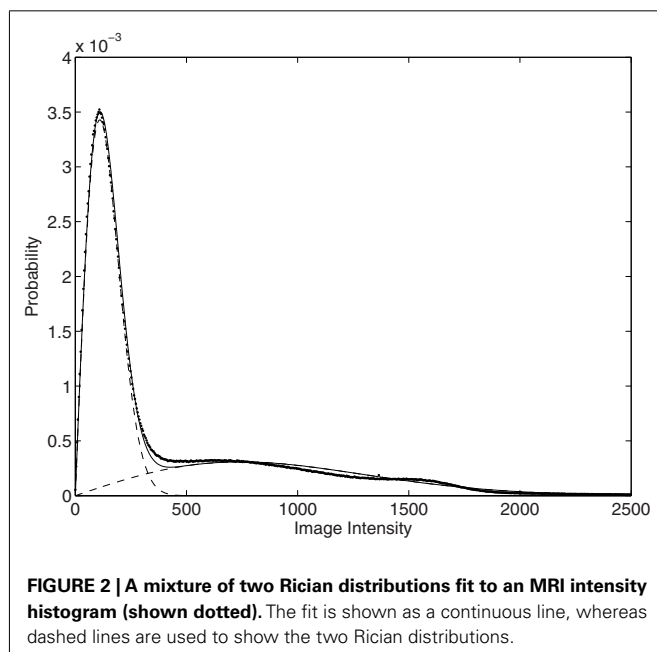
2.6.1. Noise estimates

The model assumes that images can be closely aligned, such that most of the residual variance is scanner noise. One of the settings needed by the algorithm is an estimate of the noise variance of each scan ($\frac{1}{\lambda_\eta}$). This may be assigned by the user, although our implementation defaults to using values estimated from the MR scans themselves.

Because MR images usually encode the magnitude of complex data, they have Rician noise. An estimate of the variance of this noise can be made by fitting a mixture of two Rician distributions to an intensity histogram from an image (see **Figure 2**). A reasonable scanner noise estimate may then be obtained from the smaller of the two estimated variances. The fitting procedure is similar to the expectation maximization for fitting a mixture of Gaussians, although the SNR fixed point formula (Koay and Basser, 2006) is used to compute the Rician parameters from the sample means and standard deviations.

Noise drawn from a Rician distribution deviates most from Gaussian in regions of low signal intensity, where it is closer to the Rayleigh distribution (Rician with zero signal). Although the mean-squares difference noise model (which assumes residuals are drawn from a Gaussian distribution) differs from the Rician assumptions, it is probably close enough within the regions of the image in which we are mostly interested. Even in the worst case, the difference between noise drawn from two Rayleigh distributions is close to Gaussian.

2.6.2. Differential operator for inhomogeneity

Without regularization, all the differences among the images would be explained by the estimated inhomogeneity fields. Therefore, their estimation needs to involve regularization, which will depend on the nature of the artifacts present. If there is no inhomogeneity, then very strong regularization may be used, whereas less would be used in the presence of large artifacts. Our

FIGURE 2 | A mixture of two Rician distributions fit to an MRI intensity histogram (shown dotted). The fit is shown as a continuous line, whereas dashed lines are used to show the two Rician distributions.

implementation regularizes by minimizing the bending energy of the fields ($\|L_b b\|^2 = \omega_0 \int_{\mathbf{x}\in\Omega} \|\nabla^2 b(\mathbf{x})\|^2 \, d\mathbf{x}$, where ω_0 is a hyper-parameter controlling the strength of the regularization), which gives generally smooth estimates. Neumann boundary conditions are imposed on b, such that the gradient is zero at the edge of the field of view.

2.6.3. Differential operator for diffeomorphisms

The solution of any image registration problem is heavily dependent on the choice of differential operator used to regularize it. Although there are principled (Bayesian) ways to optimize the operator (Simpson et al., 2011), such a strategy would be beyond the scope of the current work. Instead, a relatively *ad hoc* choice about the form of operator was made, although there were still some principled aspects involved.

The differential operator involved no penalty against absolute displacements, which has implications for the Green's function ($\mathbf{K_v}$) used by the geodesic shooting procedure. Such forms of regularization operator can not be inverted exactly to obtain a unique $\mathbf{K_v}$. Our implementation of geodesic shooting used Fast Fourier Transform (FFT) methods to obtain the Green's function. Because the DC coefficient of the FFT of the differential operator is zero, its reciprocal is infinity. We therefore set the DC coefficient of the Green's function to zero, and let the rigid-body registration account for global translations.

Many registration approaches use a Green's function that is simply a Gaussian. We chose to avoid this, as such functions privilege certain spatial scales above others (and also penalize absolute displacements). Some have instead used a $\mathbf{K_v}$ consisting of a mixture of Gaussians to account for multiple spatial scales (Risser et al., 2011). Our approach involves using a combination of the linear-elasticity and bending energy (or thin-plate) models.

$$\|\mathbf{L_v}\mathbf{v}\|^2 = \int_{\mathbf{x}\in\Omega} \left(\frac{\omega_1}{4} \left\| \mathbf{Dv}(\mathbf{x}) + (\mathbf{Dv}(\mathbf{x}))^T \right\|_F^2 + \omega_2 \mathrm{tr}(\mathbf{Dv}(\mathbf{x}))^2 \right.$$
$$\left. + \omega_3 \left\| \nabla^2 \mathbf{v}(\mathbf{x}) \right\|^2 \right) d\mathbf{x} \quad (32)$$

Three hyper-parameters are involved:

- ω_1 controls the amount of stretching and shearing (but not rotation).
- ω_2 controls the divergence, which in turn determines the amount of volumetric expansion and contraction.
- ω_3 controls the bending energy. This ensures that the resulting velocity fields have smooth spatial derivatives.

The Neumann boundary condition could not be used for the velocity fields, so these are assumed to be circular (the same as for a Fourier transform).

2.6.4. Acquisition timing adjustments

When there are only two scans to align, it is natural to encode the template at the point half-way between them. However, when the number of scans is greater than two, the choice of what time-point the template corresponds to is more arbitrary. If the regularization is the same for aligning all scans with the template, then the natural point is the average time of all scans. Generally, we

expect larger deformations for situations where the time interval between the template and scan is greater. Therefore, the regularization is adjusted for each scan so that it is (approximately) inversely proportional to the absolute time difference between the template and scan acquisition time. We also assume that the time for the template corresponds to the median of the acquisition times. Accounting for this requires the warping regularization to be adjusted, such that the penalty is in terms of an energy measure per unit of time. If the interval between template and scan is t_n units (e.g., years), the penalty for the deformation is defined as $\frac{1}{2}\left\|\mathbf{L}_{\mathbf{v}_n}\mathbf{v}_n\right\|^2 = \frac{1}{2|t_n|}\left\|\mathbf{L}_{\mathbf{v}}\mathbf{v}_n\right\|^2$, where $\mathbf{L}_{\mathbf{v}}$ encodes the penalty for one unit of time difference.

2.6.5. Integrations

The equations within the Methods section describe a continuous setting, whereas our actual implementation replaces the integrations over space with summations sampling the voxel centers. This sort of approximation is widely used for image registration, although it probably accounts for much of the findings in Yushkevich et al. (2010).

The integrations over time, which are used by the geodesic shooting procedure, also need to be discretized. These are currently done using an Euler integration scheme, which uses three steps per unit of time difference, plus an additional two steps to account for some of the larger distortions usually found in the soft tissue outside the skull. Faster registration could be achieved using fewer time steps, although this may lead to decreased accuracy.

3. RESULTS

It is difficult to evaluate models when there is no ground truth available. Although in theory, Bayesian model comparisons could be performed, these would require computations that are not currently feasible for very large models. Therefore, the evaluations are mostly anecdotal.

3.1. EFFECTS OF REGULARIZATION

The effects of regularizing diffeomorphisms using different choices of penalty is demonstrated using simulated 2D data. This involves the two images shown in **Figure 3**, which have dimensions of 256×128 pixels, and intensities ranging from 0 to 1. The areas of the circles and ellipses in the simulation were all approximately the same. These were registered together using the diffeomorphic framework – but without intensity inhomogeneity correction – using a variety of different forms of regularization. Results are shown in **Figures 4** and **5**. The first thing to notice is that the warped images all look relatively similar to each other, but the deformations (and their Jacobian determinants) differ markedly. The choice of regularization will play a significant role in any study where the aim is to localize volumetric differences.

1. The first form of regularization ($\omega_1 = 0.001$, $\omega_2 = 0.001$, and $\omega_3 = 0.1$, with an additional penalty on the square of absolute displacements of 0.0001) is intended to demonstrate the effect of using a Green's function that is approximately Gaussian. The behavior of such a kernel is such that the Jacobians of the resulting deformations are more extreme. For real longitudinal data in Alzheimer's disease, where ventricles expand over time, it
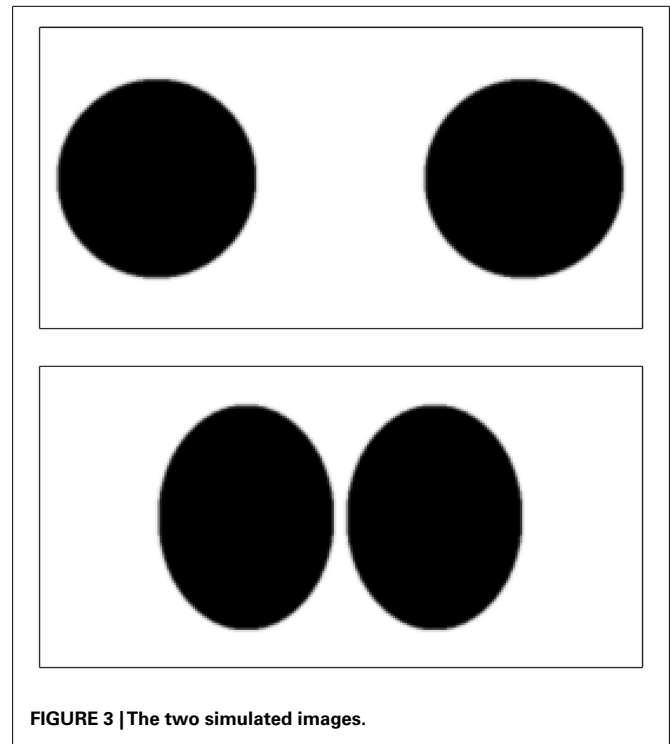
FIGURE 3 | The two simulated images.

will give the impression that brain atrophy is localized to the regions close to the ventricles.

2. The second form ($\omega_1 = 0.001$, $\omega_2 = 0.001$, and $\omega_3 = 2.0$) is dominated by a penalty against the bending energy of the deformations. For expanding ventricles, it may give the impression that brain tissue around the ventricles also expands. In general though, this form of regularization has a number of nice properties, which include scale invariance.

3. The third form is dominated by the penalty against length changes. Unfortunately, within a continuous framework, the Green's function for this one is sharply peaked, such that the value at the center is infinity.

4. The fourth form predominantly penalizes the divergence of the velocity fields, which tends to push the Jacobian determinants toward a value of one. In the bottom right of **Figures 4** and **5**, we see that the estimated volumetric differences are very small. The pure form of this regularization also has a Green's function with a singularity.

The optimal form of regularization is likely to involve a combination of the above. Although liable to have a large impact on the accuracy of image registration algorithms, the neuroimaging literature contains little on the subject.

3.2. REAL LONGITUDINAL MRI

The algorithm was evaluated using data downloaded from Part 1 of the *OASIS Longitudinal MRI Data in Non-demented and Demented Older Adults*[6] dataset (Marcus et al., 2010). This contained longitudinal scans from 82 subjects (from OAS2_0001 to

---

[6] Freely available from http://www.oasis-brains.org/

FIGURE 4 | Warped simulated images. First column: upper image warped to match lower image (from **Figure 3**). Second column: deformation fields. Third column: logarithms of Jacobian determinants (color-scales are the same for all examples, and in the range of −3 to 3). First row: results from $\omega_1 = 0.001$, $\omega_2 = 0.001$, and $\omega_3 = 0.1$, with an additional penalty on the square of absolute displacements of 0.0001. Second row: results from $\omega_1 = 0.001$, $\omega_2 = 0.001$, and $\omega_3 = 2.0$. Third row: results from $\omega_1 = 0.05$, $\omega_2 = 0.0001$, and $\omega_3 = 0.0001$. Fourth row: results from $\omega_1 = 0.001$, $\omega_2 = 0.5$, and $\omega_3 = 0.001$.

OAS2_0099), each with data from between two and five time-points. Data from each time-point consisted of between two and five MRI scans, permitting improved signal to noise ratios via averaging. Further demographic information about the subjects may be obtained from the OASIS web site.

The first step of the processing was to create averages of the scans from each time-point. Because subjects may move slightly, these averages were computed after a group-wise rigid-body alignment. This was achieved using the estimated template from the group-wise registration – but with non-linear deformations disabled.

The evaluations were based on averages over a number of scans, where the noise is no longer Rician. To simulate typical user behavior (default settings), they were done using noise estimates obtained by fitting Rician distributions to non-Rician noise.

3.2.1. Pairwise symmetry

The first test was to assess whether the procedure is actually inverse consistent if run pairwise. This simply involved aligning the first and second time-points of a pair of longitudinal images, and assessing whether the results were compatible with those from aligning the second and first. They were found to be exactly consistent.

3.2.2. Anecdotal example 1

The first illustration uses data from a 75-year-old (at first scan) right handed male with mild cognitive impairment (OAS2_0002, MMSE = 22, CDR = 0.5). Although there were three images for this subject, we just ran the algorithm using the first and last, which were collected 1869 days apart. The primary aim was to show the decrease in residual difference, after both inhomogeneity correction and registration. This provides an indication of how well the registration works, but does not give the full story (Rohlfing, 2012) because some very implausible deformations may also greatly reduce the residuals. Jacobian determinant maps are also shown, which tell us about the plausibility of the volumetric changes involved. These results are shown in **Figure 6**, with more detail around the right hippocampus shown in **Figure 7**.

The first thing to note is that the registration greatly reduced the residual difference. After registration, there is little remaining

FIGURE 5 | Warped simulated images. First column: lower image warped to match upper image. Second column: deformation fields. Third column: logarithms of Jacobian determinants (color-scales are the same for all examples, and in the range of −3 to 3). First row: results from $\omega_1 = 0.001$, $\omega_2 = 0.001$, and $\omega_3 = 0.1$, with an additional penalty on the square of absolute displacements of 0.0001. Second row: results from $\omega_1 = 0.001$, $\omega_2 = 0.001$, and $\omega_3 = 2.0$. Third row: results from $\omega_1 = 0.05$, $\omega_2 = 0.0001$, and $\omega_3 = 0.0001$. Fourth row: results from $\omega_1 = 0.001$, $\omega_2 = 0.5$, and $\omega_3 = 0.001$.

structure to be seen in the residuals – particularly within the brain. We note also that the estimated Jacobian determinants seem to be plausible. A comparison between the residuals depicted in the bottom center and bottom left of the Figures shows the effect of the inhomogeneity correction, which also reduces the residual difference for these data.

The brain is enclosed within the skull, so there is relatively little external influence on its shape. This may be contrasted with the soft tissue outside the skull, which shows extensive shape changes due to the subject's head positioning within the scanner. The Jacobian images (**Figure 6**) also show some artifacts along the right hand edge. These result from the image data wrapping around (see close to the right hand edge on the coronal view of the second time-point image).

3.2.3. Anecdotal example 2
The second illustration uses data from a 66-year-old male with dementia (OAS2_0048, MMSE = 19, CDR = 1). There were five scans for this subject, collected over a period of 1233 days. Rigidly aligned versions of the images are shown along the top of **Figure 8**,

with the corresponding maps of expansion (divergence of initial velocity) shown below. Note that the expansion map of the middle time-point is almost zero, as that point served as the reference time for the group-wise alignment. Careful examination of the divergence maps also reveals what appear to be artifactual volume changes for the more prominent blood vessels. This effect was found in many of the subjects.

3.2.4. Principal components
Group-wise longitudinal registration was run for all 82 subjects' data. The region within the cranium of each subject was identified by running the "new segment" algorithm of SPM8 (Ashburner and Friston, 2005) on their aligned mean images (μ), and summing the estimated gray matter, white matter and CSF maps together. Divergence values from inside the cranium of each subject were collected, from which $N \times N$ Gram matrices were computed and normalized by the number of voxels. The Gram matrices were decomposed via an eigen-decomposition and the largest eigenvalues identified. The corresponding eigenvectors, scaled by the square root of the eigenvalues, are plotted in **Figure 9**.

FIGURE 6 | Illustration of the results obtained from matching a pair of images of a subject with mild cognitive impairment, which were collected 1895 days apart (OAS2_0002). The three images of the residual difference shown along the bottom are all windowed the same. Black indicates a value of −500 or less, whereas white indicates values of 500 or above.
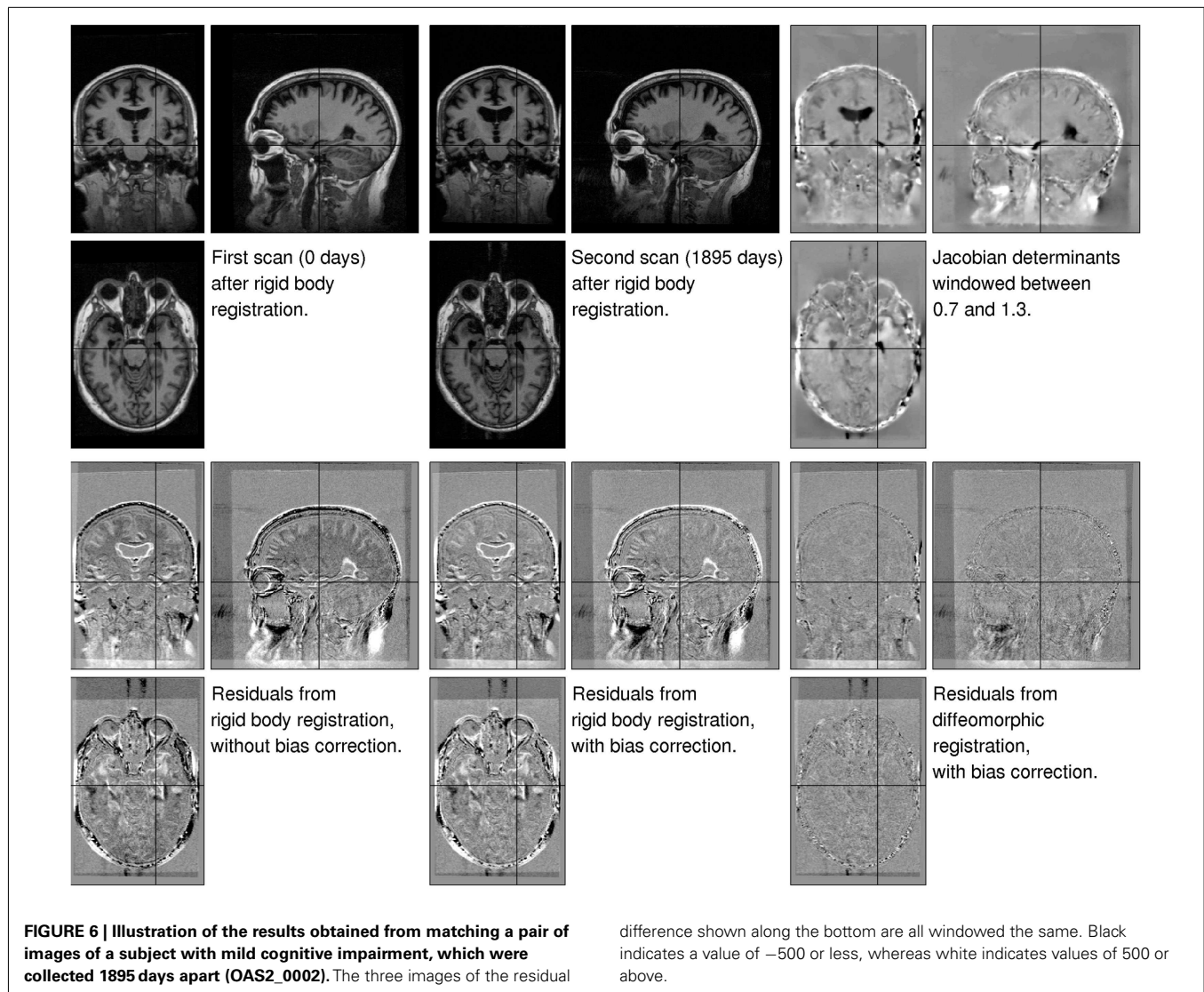
In general, the plots for multiple time-points appeared fairly linear, although we do notice a steeper gradient between the first two points for some of the subjects. Although it appears to be some form of artifact, we do not yet have a good explanation for it.

3.3. MEAN IMAGES

Images of mean expansion rate were computed for each subject, by fitting voxel-wise linear models through the divergence maps.

A mapping between each subject's gray matter, white matter, and CSF and the population mean of these tissues was estimated using Dartel (Ashburner, 2007; Ashburner and Friston, 2009). In addition, an affine mapping between the population mean and MNI space was also estimated. The compositions of these mappings were then used to warp each subject's mean (μ) and their expansion rate images to MNI space[7].

---

[7]We note that this is the approach typically adopted within the SPM software, and that parallel transport (Younes et al., 2008) should instead have been used to warp the velocity fields to MNI space, prior to computing their divergences.

A simple average (not weighted by Jacobian determinants) of all the warped mean images was computed. Similarly, simple averages of the warped expansion rate maps for the control subjects and subjects with dementia were also computed. These averages are shown in **Figure 10**, and clearly show the pattern of atrophy typically found in aging and dementia. Outside the brain, age associated skin thickness decreases can also be seen (Shuster et al., 2006).

4. DISCUSSION

Research in biology is about much more than collecting p-values. Ultimately, we want to understand the mechanisms behind brain growth, development, aging, and various disease processes. Mechanistic models require internal consistency and plausible underlying assumptions. When internal consistency is not achieved, it is an indication that something is wrong. We demonstrated that for pairwise registration, our approach gives consistent solutions – irrespective of the order of the images.

Mechanistic models should be based ideally on well accepted underlying assumptions, which for image registration are that
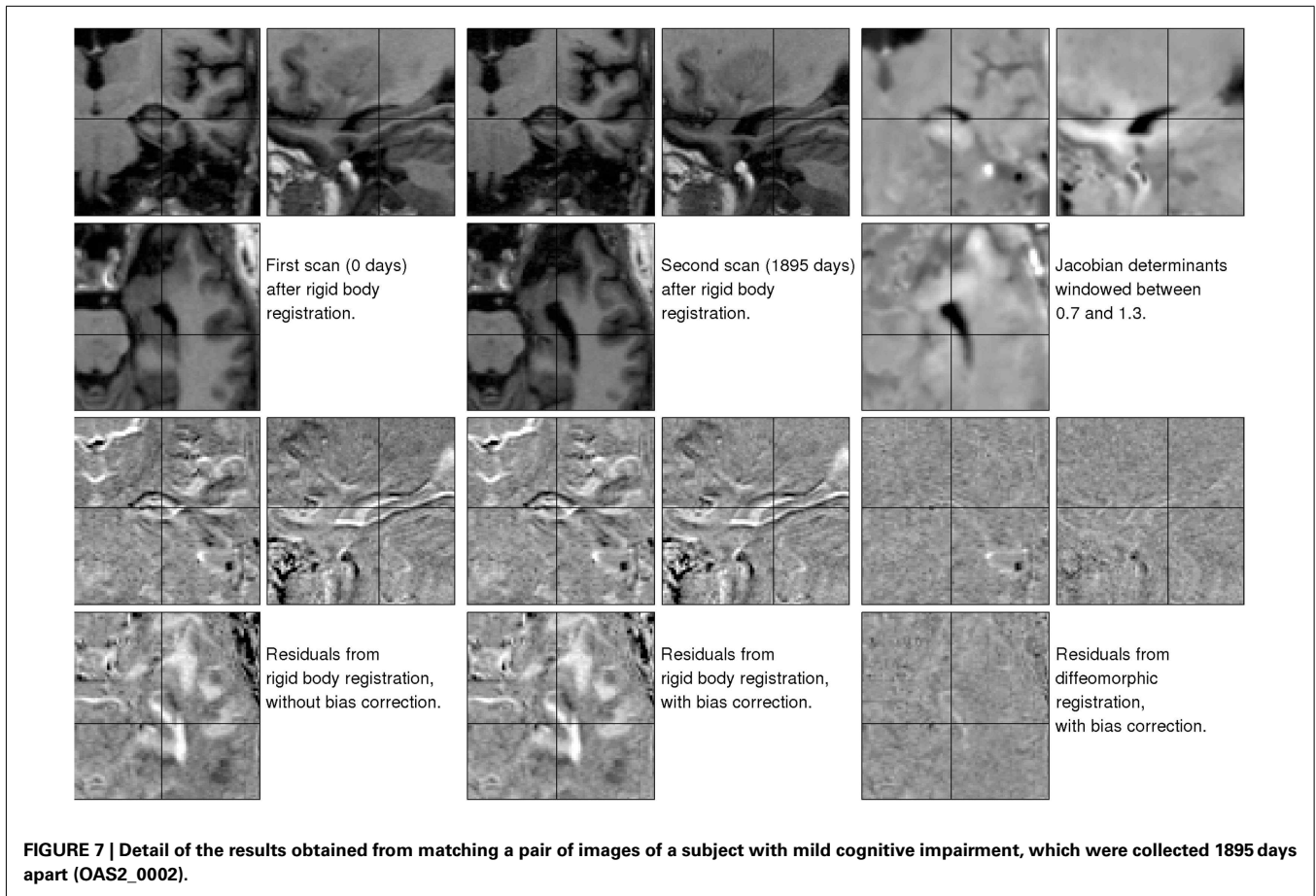
FIGURE 7 | Detail of the results obtained from matching a pair of images of a subject with mild cognitive impairment, which were collected 1895 days apart (OAS2_0002).
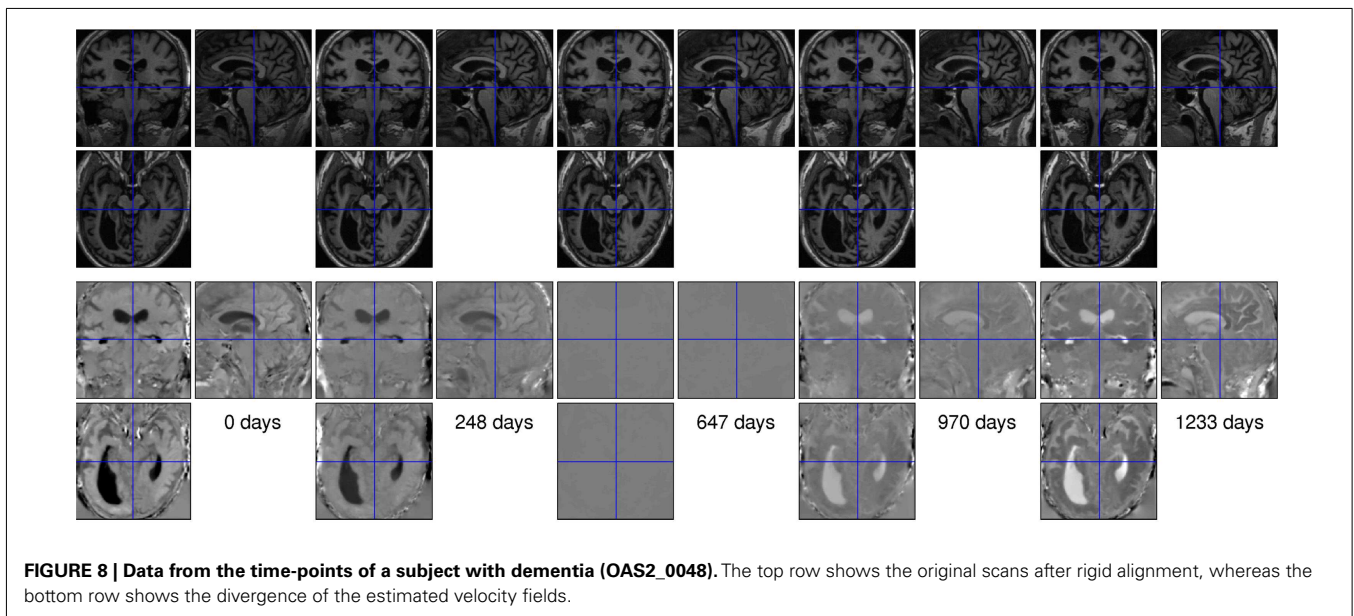
FIGURE 8 | Data from the time-points of a subject with dementia (OAS2_0048). The top row shows the original scans after rigid alignment, whereas the bottom row shows the divergence of the estimated velocity fields.

lengths, areas, and volumes should never fall below zero. Relative volumes are computed from deformations via the Jacobian determinants, so a necessary condition for a valid growth model is that these Jacobian determinants must be positive. Achieving this requires a diffeomorphic deformation framework.

FIGURE 9 | Plots of main eigenvector from each subject's divergence maps within the cranium. The columns show plots from those control subjects who were scanned only twice, plots from control subjects who were scanned more than twice, plots from subjects with dementia who were scanned twice, and plots from

subjects with dementia who were scanned more than twice. Dotted lines show the best linear fit. Note that the plots are sorted according to their average slope, which was done for easier visualization. Some of the eigenvectors were also rescaled by −1, such that all the gradients are positive.

Some readers may object to the use of large deformation diffeomorphic registration[8] approaches for modeling the relatively small longitudinal changes seen in aging. In principle though, a related framework would also be applicable to modeling growth and development of the brain – or indeed any other organ – from fetus through to adult. Small-deformation approximations would fail for these more extreme changes, whereas a diffeomorphic approach could potentially model volumes and lengths that change by orders of magnitude.

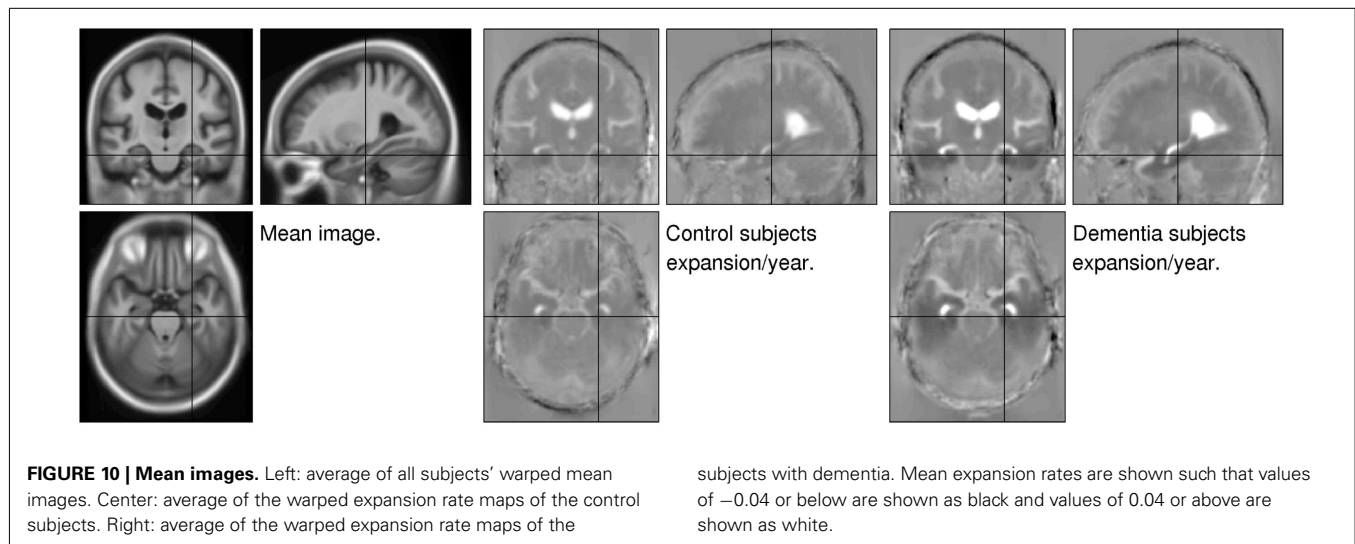Further modifications to the current approach would be required to account for intensity changes that are intrinsic to the brain. For example, the brains of young infants have changing appearances throughout myelogenesis. Similarly, white matter

hypo-intensities, stroke, etc., may be more properly explained by intensity changes rather than shape changes, although it is not always entirely clear what solution is more appropriate. Others are beginning to develop models for simultaneous shape and appearance changes (Trouvé and Younes, 2005; Hong et al., 2012), although there is still much more to be done. Alternatively, greater robustness may be achieved by using a matching term other than the L^2 norm, for which residual differences ideally follow a Gaussian distribution. A distribution with fatter tails, such as a mixture of Gaussians, may be more appropriate for modeling outliers (Penny et al., 2007).

The effects of changing the form and magnitude of the regularization are still relatively unexplored. It is likely that the optimal amount will depend on the ultimate objectives of a study. More regularization will decrease the noise in the estimated deformations – at the expense of introducing bias. Less regularization will decrease the bias, at the expense of fitting noise. If the

---

[8]The alternative is to use small deformation assumptions, which suppose that deviations from the identity transform can constructed by addition of displacements, rather than composing a series of tiny deformations.

FIGURE 10 | Mean images. Left: average of all subjects' warped mean images. Center: average of the warped expansion rate maps of the control subjects. Right: average of the warped expansion rate maps of the subjects with dementia. Mean expansion rates are shown such that values of −0.04 or below are shown as black and values of 0.04 or above are shown as white.

aim is to compute volume changes of brain structures – or even whole brains – it may be better to use less regularization because the noise will be averaged out over multiple voxels. However, if the aim is to make use of values in individual voxels, the optimal bias-variance tradeoff will be achieved with heavier regularization.

We have presented a generative modeling framework for longitudinal MRI, which combines rigid alignment, diffeomorphic warping and differential intensity non-uniformity correction with respect to a within-subject template that evolves to be an average with regard to all three of these aspects. The approach is symmetric and transitive by construction. In the pairwise case, it is not only inverse consistent, but the path (on the manifold of diffeomorphisms) from one image to the other via the template is consistent with the direct geodesic path between the images. However, there is scope for further refinement of the model when dealing with images collected at more than two time-points. Such extensions would require ideas about variable rates of growth to be incorporated (Fishbaugh et al., 2011; Niethammer et al., 2011; Trouvé and Vialard, 2012) and will be investigated in future work.

REFERENCES

Al-Mohy, A., and Higham, N. (2009). Computing the Fréchet derivative of the matrix exponential, with an application to condition number estimation. *SIAM J. Matrix Anal. Appl.* 30, 1639–1657.

Arsigny, V. (2006). *Processing Data in Lie Groups: An Algebraic Approach. Application to Non-Linear Registration and Diffusion Tensor MRI.* Ph.D. thesis, École Polytechnique, Paris.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113.

Ashburner, J., Andersson, J., and Friston, K. (1999). High-dimensional image registration using symmetric priors. *Neuroimage* 9, 619–628.

Ashburner, J., and Friston, K. (2005). Unified segmentation. *Neuroimage* 26, 839–851.

Ashburner, J., and Friston, K. (2009). Computing average shaped tissue probability templates. *Neuroimage* 45, 333–341.

Ashburner, J., and Friston, K. (2011). Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *Neuroimage* 55, 954–967.

Avants, B., Epstein, C., Grossman, M., and Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41.

Beg, M., and Khan, A. (2007). Symmetric data attachment terms for large deformation image registration. *IEEE Trans. Med. Imaging* 26, 1179–1189.

Bro-Nielsen, M., and Gramkow, C. (1996). "Fast fluid registration of medical images," in *Visualization in Biomedical Computing,* eds K. H. Höhne and R. Kikinis (Berlin: Springer), 265–276.

Bruveris, M., Gay-Balmaz, F., Holm, D., and Ratiu, T. (2011). The momentum map representation of images. *J. Nonlinear Sci.* 21, 115–150.

Cachier, P., and Rey, D. (2000). "Symmetrization of the non-rigid registration problem using inversion-invariant energies: application to multiple sclerosis," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2000, Lecture Notes in Computer Science,* Vol. 1935,

eds S. L. Delp, A. M. DiGioia and B. Jaramaz (Berlin: Springer), 472–481.

Christensen, G., and Johnson, H. (2001). Consistent image registration. *IEEE Trans. Med. Imaging* 20, 568–582.

Fishbaugh, J., Durrleman, S., and Gerig, G. (2011). "Estimation of smooth growth trajectories with controlled acceleration from time series shape data," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, Lecture Notes in Computer Science,* Vol. 6892, eds G. Fichtinger, A. Martel, and T. Peters (Berlin: Springer), 401–408.

Fox, N., Ridgway, G., and Schott, J. (2011). Algorithms, atrophy and Alzheimer's disease: cautionary tales for clinical trials. *Neuroimage* 57, 15–18.

Freeborough, P., and Fox, N. (1997). The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. Med. Imaging* 16, 623–629.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* 2, 861–863.

Hart, G., Zach, C., and Niethammer, M. (2009). "An optimal control approach for deformable registration," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009* (New York: IEEE), 9–16.

Holland, D., McEvoy, L. K., Dale, A. M., and Alzheimer's Disease Neuroimaging Initiative. (2012). Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI. *Hum. Brain Mapp.* 33, 2586–2606.

Hong, Y., Joshi, S., Sanchez, M., Styner, M., and Niethammer, M. (2012). "Metamorphic geodesic regression," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, Lecture Notes in Computer Science*, Vol. 7512, eds N. Ayache, H. Delingette, P. Golland and K. Mori (Berlin: Springer), 197–205.

Hua, X., Gutman, B., Boyle, C., Rajagopalan, P., Leow, A., Yanovsky, I., et al. (2011). Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. *Neuroimage* 57, 5–14.

Koay, C., and Basser, P. (2006). Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. *J. Magn. Reson.* 179, 317–322.

Leung, K. K., Ridgway, G. R., Ourselin, S., Fox, N. C., and Alzheimer's Disease Neuroimaging Initiative. (2012). Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *Neuroimage* 59, 3995–4005.

Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., and Weiskopf, N. (2010). Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. *Magn. Reson. Med.* 64, 229–238.

Marcus, D., Fotenos, A., Csernansky, J., Morris, J., and Buckner, R. (2010).

Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J. Cogn. Neurosci.* 22, 2677–2684.

Miller, M., Trouvé, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.* 24, 209–228.

Moakher, M. (2002). Means and averaging in the group of rotations. *SIAM Rev. Soc. Ind. Appl. Math.* 24, 1–16.

Modat, M., Ridgway, G., Hawkes, D., Fox, N., and Ourselin, S. (2010). "Nonrigid registration with differential bias correction using normalised mutual information," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2010* (New York: IEEE), 356–359.

Modersitzki, J. and Wirtz, S. (2006). "Combining homogenization and registration," in *Biomedical Image Registration: Third International Workshop, WBIR 2006, LNCS*, eds J. P. W. Pluim, B. Likar and F. A. Gerritsen (Berlin: Springer), 257–263.

Moler, C., and Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev. Soc. Ind. Appl. Math.* 45, 3–49.

Niethammer, M., Hart, G., and Zach, C. (2009). "An optimal control approach for the registration of image time-series," in *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 Held Jointly with the 2009 28th Chinese Control Conference, CDC/CCC 2009*, eds J. Baillieul and L. Guo (New York: IEEE), 2427–2434.

Niethammer, M., Huang, Y., and Vialard, F. (2011). "Geodesic regression for image time-series," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011* (Springer), 655–662.

Pennec, X., and Arsigny, V. (2012). "Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups," in *Matrix Information Geometry*, eds F. Nielsen and R. Bhatia (Berlin: Springer).

Penny, W., Kilner, J., and Blankenburg, F. (2007). Robust Bayesian general linear models. *Neuroimage* 36, 661–671.

Reuter, M., and Fischl, B. (2011). Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage* 57, 19.

Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012a).

Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61, 1402–1418.

Reuter, M., Schmansky, N., Rosas, H., and Fischl, B. (2012b). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61, 1402–1418.

Risser, L., Vialard, F., Wolz, R., Murgasova, M., Holm, D., and Rueckert, D. (2011). Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. *IEEE Trans. Med. Imaging* 30, 1746–1759.

Rohlfing, T. (2012). Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE Trans. Med. Imaging* 31, 153–163.

Shuster, S., Black, M., and McVitie, E. (2006). The influence of age and sex on skin thickness, skin collagen and density. *Br. J. Dermatol.* 93, 639–643.

Simpson, I., Schnabel, J., Groves, A., Andersson, J., and Woolrich, M. (2011). Probabilistic inference of regularisation in non-rigid registration. *Neuroimage* 59, 2438–2451.

Skrinjar, O., Bistoquet, A., and Tagare, H. (2008). Symmetric and transitive registration of image sequences. *Int. J. Biomed. Imaging* 2008, 686875.

Sled, J., Zijdenbos, A., and Evans, A. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97.

Smith, S., De Stefano, N., Jenkinson, M., and Matthews, P. (2001). Normalized accurate measurement of longitudinal brain change. *J. Comput. Assist. Tomogr.* 25, 466.

Studholme, C., Cardenas, V., Song, E., Ezekiel, F., Maudsley, A., and Weiner, M. (2004). Accurate template-based correction of brain MRI intensity distortion with application to dementia and aging. *IEEE Trans. Med. Imaging* 23, 99–110.

Tagare, H., Groisser, D., and Skrinjar, O. (2009). Symmetric non-rigid registration: a geometric theory and some numerical techniques. *J. Math. Imaging Vis.* 34, 61–88.

Thomas, A. G., Marrett, S., Saad, Z. S., Ruff, D. A., Martin, A., and Bandettini, P. A. (2009). Functional but not structural changes associated with learning: an exploration of longitudinal voxel-based morphometry (VBM). *Neuroimage* 48, 117–125.

Thompson, W., Holland, D., and Alzheimer's Disease Neuroimaging Initiative. (2011). Bias in tensor based morphometry stat-roi measures may result in unrealistic power estimates. *Neuroimage* 57, 1–4.

Trouvé, A., and Vialard, F. X. (2012). Shape splines and stochastic shape evolutions: a second order point of view. *Quart. Appl. Math* 70, 219–251.

Trouvé, A., and Younes, L. (2005). Metamorphoses through lie group action. *Found. Comput. Math.* 5, 173–198.

Woods, R. (2003). Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. *Neuroimage* 18, 769–788.

Younes, L. (2007). Jacobi fields in groups of diffeomorphisms and applications. *Q. Appl. Math.* 65, 113–134.

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

Yushkevich, P., Avants, B., Das, S., Pluta, J., Altinay, M., Craige, C., et al. (2010). Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 T MRI data. *Neuroimage* 50, 434–445.

APPENDIX

A. DERIVATIVES FOR RIGID-BODY REGISTRATION

To differentiate equation (5), we consider a single image and drop the n subscripts.

$$\frac{\lambda}{2} \int_{\mathbf{x} \in \Omega_f} \left(\frac{\partial \left(f - \mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) \right)^2}{\partial q_i} \right) d\mathbf{x} = \lambda \int_{\mathbf{x} \in \Omega_f} \left(\mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) - f \right) \left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} \cdot \left((\nabla \mu) \circ \boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) \right) d\mathbf{x}. \tag{A1}$$

Then a change of variables is incorporated to obtain the following:

$$\lambda \int_{\mathbf{x} \in \Omega'} |\mathbf{D}\boldsymbol{\xi}_{\mathbf{q}}| \left(\left(\left(\mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) - f \right) \left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} \cdot \left((\nabla \mu) \circ \boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) \right) \right) \circ \boldsymbol{\xi}_{\mathbf{q}} \right) d\mathbf{x} = \lambda \int_{\mathbf{x} \in \Omega'} |\mathbf{D}\boldsymbol{\xi}_{\mathbf{q}}| \left(\mu - f \left(\boldsymbol{\xi}_{\mathbf{q}} \right) \right) \left(\left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial \mathbf{q}_i} \circ \boldsymbol{\xi}_{\mathbf{q}} \right) \cdot (\nabla \mu) \right) d\mathbf{x}. \tag{A2}$$

Computing the above requires $\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} \circ \boldsymbol{\xi}_{\mathbf{q}}$. Inverting equation (4) gives:

$$\xi_{\mathbf{q}}^{-1} (\mathbf{x}) = \mathbf{I}_{3,4} \mathbf{M}_{\mu}^{-1} \mathbf{R}_{\mathbf{q}}^{-1} \mathbf{M} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \tag{A3}$$

Using the identity $\frac{\partial (\mathbf{R}^{-1})}{\partial q} = -\mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial q} \mathbf{R}^{-1}$ to differentiate this, results in:

$$\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} = -\mathbf{I}_{3,4} \mathbf{M}_{\mu}^{-1} \mathbf{R}_{\mathbf{q}}^{-1} \frac{\partial \mathbf{R}_{\mathbf{q}}}{\partial q_i} \mathbf{R}_{\mathbf{q}}^{-1} \mathbf{M} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \tag{A4}$$

This is combined with equation (4) to generate:

$$\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} \circ \boldsymbol{\xi}_{\mathbf{q}} = -\mathbf{I}_{3,4} \mathbf{M}_{\mu}^{-1} \mathbf{R}_{\mathbf{q}}^{-1} \frac{\partial \mathbf{R}_{\mathbf{q}}}{\partial q_i} \mathbf{M}_{\mu} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \tag{A5}$$

Gauss-Newton optimization requires a positive definite approximation to the matrix of second derivatives equation (11). The actual Hessian matrix is given by the following, but it is not guaranteed to be positive definite:

$$\frac{\partial^2}{\partial q_i \partial q_j} \frac{\lambda}{2} \int_{\mathbf{x} \in \Omega_f} \left(f - \mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) \right)^2 d\mathbf{x} = \lambda \int_{\mathbf{x} \in \Omega_f} \left(\frac{\partial \mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right)}{\partial q_i} \frac{\partial \mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right)}{\partial q_j} + \left(\mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) - f \right) \frac{\partial^2 u \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right)}{\partial q_i \partial q_j} \right) d\mathbf{x} \tag{A6}$$

On average, $\mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) - f$ should be zero at the solution, so the second term can be omitted to obtain a positive definite approximation:

$$\lambda \int_{\mathbf{x} \in \Omega_f} \frac{\partial \mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right)}{\partial q_i} \frac{\partial \mu \left(\boldsymbol{\xi}_{\mathbf{q}}^{-1} \right)}{\partial q_j} d\mathbf{x} = \lambda \int_{\mathbf{x} \in \Omega_f} \left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} \cdot \left((\nabla \mu) \circ \boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) \right) \left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_j} \cdot \left((\nabla \mu) \circ \boldsymbol{\xi}_{\mathbf{q}}^{-1} \right) \right) d\mathbf{x} \tag{A7}$$

Finally, a change of variables gives:

$$\lambda \int_{\mathbf{x} \in \Omega'} |\mathbf{D}\boldsymbol{\xi}_{\mathbf{q}}| \left(\left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_i} \circ \boldsymbol{\xi}_{\mathbf{q}} \right) \cdot (\nabla \mu) \right) \left(\left(\frac{\partial \boldsymbol{\xi}_{\mathbf{q}}^{-1}}{\partial q_j} \circ \boldsymbol{\xi}_{\mathbf{q}} \right) \cdot (\nabla \mu) \right) d\mathbf{x}. \tag{A8}$$

B. TEMPLATE GRADIENTS

The gradient in equation (23) is not the same as the spatial gradient of the template image (μ), which would be computed via

$$\nabla \mu = \frac{\sum_{n=1}^{N} |\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}| \nabla \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n} \right) + \sum_{n=1}^{N} \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n} \right) \nabla |\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}|}{\sum_{n=1}^{N} |\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}|} - \frac{\left(\sum_{n=1}^{N} |\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}| \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n} \right) \right) \left(\sum_{n=1}^{N} \nabla |\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}| \right)}{\left(\sum_{n=1}^{N} |\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}| \right)^2}. \tag{A9}$$

The above expression incorporates the gradient of the Jacobian determinants, which would have a detrimental effect on the registration (see Appendix C). The Gâteaux variation of the matching term, with respect to variations in the initial velocity, is

$$\frac{d}{d\tau} \frac{\lambda_n}{2} \int_{\mathbf{x} \in \Omega_\mu} \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x}) \right| \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x}) - \mu \left(\mathbf{x} - \tau \mathbf{h}(\mathbf{x}) \right) \right)^2 dx \Bigg|_{\tau=0}$$

$$= \frac{d}{d\tau} \frac{\lambda_n}{2} \int_{\mathbf{x} \in \Omega_\mu} \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x}) \right| \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x}) - \frac{\sum_{l=1}^{N} \lambda_l \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_l}(\mathbf{x}) \right| f_l \circ \boldsymbol{\phi}_{\mathbf{v}_l}(\mathbf{x} - \tau \mathbf{h}(\mathbf{x}))}{\sum_{l=1}^{N} \lambda_l \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_l}(\mathbf{x}) \right|} \right)^2 dx \Bigg|_{\tau=0}$$

$$= \lambda_n \int_{\mathbf{x} \in \Omega_\mu} \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n}(\mathbf{x}) \right| \left(f_n \circ \boldsymbol{\phi}_{v_n}(\mathbf{x}) - \mu(\mathbf{x}) \right) \mathbf{g}(\mathbf{x}) \cdot \mathbf{h}(\mathbf{x}) \, d\mathbf{x} \tag{A10}$$

where μ and $\mathbf{g}$ are computed as in equations (21) and (23). A similar scheme may be used to derive the positive definite approximation to the second Gâteaux variation.

For the software implementation, the gradients of the warped images are computed by sampling the image and its gradients according to the transformation, and multiplying the gradients by the transpose of the Jacobian tensor at each point.

$$\nabla \left(f_n \circ \boldsymbol{\phi}_{\mathbf{v}_n} \right) = \left(\mathbf{D}\boldsymbol{\phi}_{\mathbf{v}_n} \right)^T \left(\left(\nabla f_n \right) \circ \boldsymbol{\phi}_{\mathbf{v}_n} \right) \tag{A11}$$

C. PAIRWISE SYMMETRY

Pairwise symmetric registration is a special case of the group-wise formulation, and is of interest to many. We note that for pairwise registration (i.e., $N = 2$), where we define $\mathbf{v} = \mathbf{v}_1 = -\mathbf{v}_2$ (and the regularization term has been halved for convenience), the objective function reduces to

$$\mathcal{E} = \frac{1}{2} \int_{\mathbf{x} \in \Omega_\mu} \frac{\lambda_1 \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}}(\mathbf{x}) \right| \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}}(\mathbf{x}) \right|}{\lambda_1 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}}(\mathbf{x}) \right| + \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}}(\mathbf{x}) \right|} \left(f_1 \circ \boldsymbol{\phi}_{\mathbf{v}}(\mathbf{x}) - f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}}(\mathbf{x}) \right)^2 d\mathbf{x} + \frac{1}{2} \| \mathbf{L}_{\mathbf{v}} \mathbf{v} \|^2. \tag{A12}$$

The solution, where derivatives of the objective function with respect to variations in $\mathbf{v}$ are zero, satisfies the following:

$$\frac{\lambda_1 \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right| \left(\lambda_1 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| \nabla \left(f_1 \circ \boldsymbol{\phi}_{\mathbf{v}} \right) + \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right| \nabla \left(f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}} \right) \right) \left(f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}} - f_1 \circ \boldsymbol{\phi}_{\mathbf{v}} \right)}{\left(\lambda_1 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| + \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right| \right)^2} = \mathbf{L}_{\mathbf{v}}^\dagger \mathbf{L}_{\mathbf{v}} \mathbf{v}. \tag{A13}$$

For the geodesic shooting in approach to work, we need to consider the solution along the entire trajectory of the diffeomorphism. To do this, we introduce another diffeomorphic mapping, $\boldsymbol{\zeta}$, which is used to assess the effect of determining the initial velocities at some point other than the mid-point. For the geodesic shooting in equation (19) to work properly, the left-hand side of equation (45) (which we here refer to as momentum, $\mathbf{u}$) must become $|\mathbf{D}\boldsymbol{\zeta}(\mathbf{D}\boldsymbol{\zeta})^T(\mathbf{u} \circ \boldsymbol{\zeta})|$ if we replace $\boldsymbol{\phi}_{\mathbf{v}}$ and $\boldsymbol{\phi}_{-\mathbf{v}}$ with $\boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta}$ and $\boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta}$ respectively. For typesetting reasons, we decompose the left-hand side of equation (45) into two factors and consider what happens to each of them if we make those substitutions.

For the first factor [corresponding to $a_n(\mathbf{x})$ in equation 22], using the identity $|\mathbf{D}(\boldsymbol{\phi} \circ \boldsymbol{\zeta})| = |\mathbf{D}\boldsymbol{\zeta}|(|\mathbf{D}\boldsymbol{\phi}| \circ \boldsymbol{\zeta})$ we obtain

$$\frac{\lambda_1 \lambda_2 \left| \mathbf{D} \left(\boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right| \left| \mathbf{D} \left(\boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right| \left(f_1 \circ \boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta} - f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta} \right)}{\lambda_1 \left| \mathbf{D} \left(\boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right| + \lambda_2 \left| \mathbf{D} \left(\boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right|} = |\mathbf{D}\boldsymbol{\zeta}| \left(\frac{\lambda_1 \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right| \left(f_1 \circ \boldsymbol{\phi}_{\mathbf{v}} - f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}} \right)}{\lambda_1 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| + \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right|} \right) \circ \boldsymbol{\zeta}. \tag{A14}$$

For the second factor [corresponding to $\mathbf{g}(\mathbf{x})$ in equation 22], we in addition use the element-wise matrix multiplication $\nabla \left(f \circ \boldsymbol{\zeta} \right) = \left(\mathbf{D}\boldsymbol{\zeta} \right)^T \left(\nabla f \right) \circ \boldsymbol{\zeta}$ to obtain

$$\frac{\lambda_1 \left| \mathbf{D} \left(\boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right| \nabla \left(f_1 \circ \boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta} \right) + \lambda_2 \left| \mathbf{D} \left(\boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right| \nabla \left(f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta} \right)}{\lambda_1 \left| \mathbf{D} \left(\boldsymbol{\phi}_{\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right| + \lambda_2 \left| \mathbf{D} \left(\boldsymbol{\phi}_{-\mathbf{v}} \circ \boldsymbol{\zeta} \right) \right|} = \left(\mathbf{D}\boldsymbol{\zeta} \right)^T \left(\frac{\lambda_1 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| \nabla \left(f_1 \circ \boldsymbol{\phi}_{\mathbf{v}} \right) + \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right| \nabla \left(f_2 \circ \boldsymbol{\phi}_{-\mathbf{v}} \right)}{\lambda_1 \left| \mathbf{D}\boldsymbol{\phi}_{\mathbf{v}} \right| + \lambda_2 \left| \mathbf{D}\boldsymbol{\phi}_{-\mathbf{v}} \right|} \right) \circ \boldsymbol{\zeta}. \tag{A15}$$

By recombining the factors, we see that symmetry is achieved, while still satisfying the requirements of a geodesic shooting approach based on the *EPdiff* equation (Euler-Poincaré equation on diffeomorphisms).

This symmetric registration approach is very similar to that of Hart et al. (2009), Niethammer et al. (2009), although others have proposed different strategies for achieving inverse consistency. Previously Beg and Khan (2007) presented two adjustments to enforce inverse consistency in the large deformation diffeomorphic metric mapping (LDDMM) approach. The first of these was based on aligning an image pair to their simple half-way average. It is similar to the approach of Avants et al. (2008), as well as one of the approaches

proposed in Younes (2007). Unfortunately, it introduces a discontinuity in the evolution equations at the mid-point (Bruveris et al., 2011). The second approach of Beg and Khan (2007) does not introduce this discontinuity, but leads to more complicated evolution equations that do not strictly obey *EPdiff*. Similarly, the other strategy for achieving inverse consistency proposed in Younes (2007) also does not obey these equations.

We note that our matching term in equation (A12) does not satisfy all the desiderata set out by Tagare et al. (2009). In particular, we see that for aligning two constant images, this term will depend on the deformations. However, it is worth noting that the derivatives of the matching term with respect to velocity (see equation A13) are zero for constant images, so there should not be any tendency toward favoring any particular solution.