

Mid-Space-Independent Deformable Image Registration

Iman Aganj^{a,*}, Juan Eugenio Iglesias^b, Martin Reuter^{a,c,d}, Mert Rory Sabuncu^{a,c,e}, Bruce Fischl^{a,c,f}

{iman, mreuter, msabuncu, fischl}@nmr.mgh.harvard.edu; e.iglesias@ucl.ac.uk

^a Athinoula A. Martinos Center for Biomedical Imaging, Radiology Department, Massachusetts General Hospital, Harvard Medical School, 149, 13th St., Suite 2301, Charlestown, MA 02129, USA

^b Translational Imaging Group, University College London, Malet Place Engineering Building, London WC1E 6BT, UK

^c Computer Science and Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 32 Vassar St., Cambridge, MA 02139, USA

^d German Center for Neurodegenerative Diseases (DZNE), Siegmund-Freud-Straße 27, 53127 Bonn, Germany

^e School of Electrical and Computer Engineering and Meinig School of Biomedical Engineering, Cornell University, 300 Rhodes Hall, Ithaca, NY 14853, USA

^f Harvard-MIT Division of Health Sciences and Technology, 77 Massachusetts Ave., Room E25-519, Cambridge, MA 02139, USA

* Corresponding author:

Iman Aganj

Email: iman@nmr.mgh.harvard.edu

Telephone: +1 (617) 724-5652

Web: <http://nmr.mgh.harvard.edu/~iman>

Address: 149, 13th St., Suite 2301, Charlestown, MA 02129, USA

NeuroImage, vol. 152, pp. 158-170, 2017.

<http://doi.org/10.1016/j.neuroimage.2017.02.055>

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0>

ABSTRACT

Aligning images in a mid-space is a common approach to ensuring that deformable image registration is symmetric – that it does not depend on the arbitrary ordering of the input images. The results are, however, generally dependent on the mathematical definition of the mid-space. In particular, the set of possible solutions is typically restricted by the constraints that are enforced on the transformations to prevent the mid-space from drifting too far from the native image spaces. The use of an implicit atlas has been proposed as an approach to mid-space image registration. In this work, we show that when the atlas is aligned to each image in the native image space, the data term of implicit-atlas-based deformable registration is inherently independent of the mid-space. In addition, we show that the regularization term can be reformulated independently of the mid-space as well. We derive a new symmetric cost function that only depends on the transformation morphing the images to each other, rather than to the atlas. This eliminates the need for anti-drift constraints, thereby expanding the space of allowable deformations. We provide an implementation scheme for the proposed framework, and validate it through diffeomorphic registration experiments on brain magnetic resonance images.

Keywords: Mid-space-independent (MSI) registration, deformable image registration, implicit atlas, symmetry, inverse-consistency.

1. Introduction

The computation of a set of dense spatial correspondences among images – a.k.a. image registration – is a central step in most population and longitudinal imaging studies. Linear transformation is often not sufficient to account for cross-subject anatomical variation or temporal changes in an individual anatomy, thereby making deformable image registration (Sotiras et al., 2013) a necessary part of most analysis pipelines. The importance of registration accuracy in neuroimaging is evident from the literature; for instance, inaccurate alignment has been shown to lead to incorrect diagnosis (Reuter et al., 2014) and ineffective radiotherapy (Castadot et al., 2008) of tumors, and the inability to detect early effects of Alzheimer’s disease (Cuingnet et al., 2011; Fischl et al., 2009).

In deformable registration, the choice of the reference space in which the images are compared affects the outcome, making the resulting deformation field dependent on this choice. Choosing the native space of one of the input images (say, the first image) as the reference breaks the symmetry of *pairwise* registration, meaning that reversing the order of the input images will produce different spatial correspondences. Such an inverse-inconsistency has been shown to be related to biased errors introduced into the estimation of Alzheimer’s disease effects (Fox et al., 2011; Hua et al., 2011; Thompson and Holland, 2011; Yushkevich et al., 2010), daily dose computation (Yang et al., 2008) and auto re-contouring (Ye and Chen, 2009) in radiation therapy, the quantification of lesion evolution in multiple sclerosis (Cachier and Rey, 2000; Rey et al., 2002), and the measurement of longitudinal changes (Reuter et al., 2012). Local volume changes in the deformation field and discretization artifacts are two major contributors to registration asymmetry. Pairwise registration has been proposed to be symmetrized by minimizing the average of two cost functions, each using one input image as the reference space (Cachier and Rey, 2000; Christensen and Johnson, 2001; Tagare et al., 2009; Trouné and Younes, 2000), which unfortunately results in the non-uniform integration of the image mismatch measure in the native spaces of the input images (Aganj et al., 2015b).

In a different approach to achieve symmetry in pairwise registration, both images are deformed and compared in an abstract reference space chosen to be “in between” the native spaces of the images, known as the *mid-space* (Ashburner and Ridgway, 2013; Avants and Gee, 2004; Beg and Khan, 2007; Chen and Ye, 2010; Joshi et al., 2004; Lorenzen et al., 2004; Lorenzen et al., 2006; Lorenzi et al., 2013; Noblet et al., 2008, 2012; Škrinjar et al., 2008; Yang et al., 2008; Ye and Chen, 2009). Since both images are treated equally, mid-space registration is invariant with respect to the ordering of the images. Such approaches essentially minimize their cost functions with respect to *two* transformations that take the two input images to the mid-space. However, without additional constraints, this increases the degrees of freedom of the problem twofold, compared to the end result of pairwise registration that is the *one* transformation taking one input image to the other. Furthermore, if the images are compared in the mid-space, the optimization algorithm is given the liberty to update the mid-space so as to decrease the cost function without necessarily changing the resulting image-to-image transformation. For example, the algorithm can shrink the regions with mismatching image intensities to make the deformed images look more similar in the mid-space, without necessarily making them more similar in their native spaces. To alleviate these issues, additional constraints are used to prevent the mid-space from drifting away from the native spaces of the two images. These *anti-drift* constraints, which are different from those regularizing the transformations, define the mid-space. They typically either restrict the space of possible pair of transformations (resulting in fewer degrees of freedom), or penalize those values of the two transformations that move the mid-space away from the native spaces. The most common such constraints, proposed in the mid-space registration and atlas

construction literature, are restrictions on the two transformations to have opposite displacement fields (Aljabar et al., 2008; Bhatia et al., 2004; Bouix et al., 2010; Fonov et al., 2011; Guimond et al., 2000; Miller et al., 1997; Noblet et al., 2012; Studholme and Cardenas, 2004; Yang et al., 2008) or velocity fields (Ashburner and Ridgway, 2013; Grenander and Miller, 1998; Lorenzi et al., 2013). In large deformation models, geodesic averaging of the deformations has also been proposed, which preserves the desired properties of the transformations (Avants and Gee, 2004; Joshi et al., 2004; Lorenzen et al., 2006). The anti-drift constraints, however, can have the side effect of restricting the final image-to-image transformation, thereby causing the exclusion of some legitimate results (see Section 2.1 for examples). Furthermore, the choice of these constraints may affect the results by biasing the registration algorithm towards favoring a particular set of transformations.

Unbiased *atlas* construction techniques can constitute mid-space registration, as the images are deformed to the atlas space (Ashburner and Ridgway, 2013; Hart et al., 2009; Joshi et al., 2004). In an atlas construction approach to image registration, the desired output is the deformation field, but not the auxiliary atlas. Consequently, one can analytically solve for the atlas in the cost function, leading to an implicit-atlas cost function that is minimized with respect to the image-to-atlas transformations. To that end, it was initially proposed to compare the deformed images to the atlas in the mid-space (Geng et al., 2009; Joshi et al., 2004). A better-justified generative model, however, progresses from the atlas to the images and compares the deformed atlas to the images in the native image spaces (Allasonnière et al., 2007; Ma et al., 2008; Sabuncu et al., 2009). Taking advantage of this native-space atlas construction resolves the issue of susceptibility to shrinkage-type problems, leading to a proper implicit-atlas cost function for mid-space registration (Ashburner and Ridgway, 2013). Nevertheless, the registration still remains a function of *two* transformations taking the images to a mathematically defined mid-space.

In this work, we derive the key fact that implicit-atlas registration has a data term that is inherently independent of the mid-space, and only depends on the overall image-to-image transformation. This implies that the individual image-to-atlas transformations are redundant and unnecessary to keep, and that anti-drift constraints are indeed not needed. We also show how to analytically solve the common Tikhonov regularization terms with respect to one of the image-to-atlas transformations. These lead us to a new cost function that, in contrast to the existing mid-space approaches, can be minimized directly with respect to the image-to-image transformation, with no anti-drift constraints. The proposed cost function is general and can be used with any deformation field parameterization, such as the displacement and the velocity fields.

This article extends our preliminary conference version (Aganj et al., 2015a). In particular, we propose a new regularization term in addition to the data term (Section 2), evaluate our method more comprehensively on 3D brain magnetic resonance images (Section 3), propose the extension of our framework to group-wise deformable registration (Appendix A), and provide further details on the derivations and the implementation of the method (Appendix B & Appendix C).

2. Methods

2.1. Mid-space based registration

We begin with a brief overview of mid-space based pairwise deformable registration. An extension of our framework to group-wise registration is suggested in Appendix A. Let $I_1, I_2: \Omega \rightarrow \mathbb{R}$ be the two d -

dimensional input images to be registered, where $\Omega \subseteq \mathbb{R}^d$.¹ The goal of pairwise deformable registration is to compute the regular transformation $T: \Omega \rightarrow \Omega$ that makes overlapping regions of I_1 and $I_2 \circ T$ locally correspond to each other; a task that is often accomplished by minimizing a cost function with respect to T . In one popular approach, the image-to-image transformation T is parameterized as $T = T_2 \circ T_1^{-1}$, where T_1 and T_2 deform I_1 and I_2 to a *mid-space* (see Section 1 for references). The deformed images may be compared in the mid-space by minimizing a data term such as the common sum of squared differences (SSD) term, $\int_{\Omega} ((I_1 \circ T_1)(y) - (I_2 \circ T_2)(y))^2 dy$ (Geng et al., 2009; Joshi et al., 2004).² The mid-space data term is by definition invariant with respect to the ordering of the images; i.e., swapping I_1 and I_2 will swap T_1 and T_2 in the produced set of transformations.

With no additional constraints, the dimensionality of the mid-space registration problem (solving for T_1 and T_2) is twice as large as the standard asymmetric problem (solving for T). Another drawback of this approach is that the mid-space can drift arbitrarily far away from the native spaces of the images due to large changes in T_1 and T_2 , for instance through combination with a transformation S , as $T_1 \circ S$ and $T_2 \circ S$, which decreases the mid-space cost function without changing the final $T = (T_2 \circ S) \circ (T_1 \circ S)^{-1} = T_2 \circ S \circ S^{-1} \circ T_1^{-1} = T_2 \circ T_1^{-1}$. An example of this phenomenon is the situation where the optimization algorithm updates T_1 and T_2 in order to shrink the regions where the deformed images do not match, resulting in a decrease in the mid-space cost function, without necessarily changing the end result, T . To avoid this issue, additional constraints are often employed to keep the mid-space “close” to the native image spaces. Such constraints reduce the degrees of freedom and to some extent prevent the mid-space drift, however, at the expense of limiting our ability to model all possible transformations T . Examples of such anti-drift constraints and their limitations are as follows:

- The transformations may be constrained to have zero-sum displacement fields, i.e. $u_1(y) + u_2(y) = 0$ where $u_i(y) := T_i(y) - y$, resulting in the constraint $T_1(y) + T_2(y) = 2y$ (or similarly the alternative constraint $T_1^{-1}(x) + T_2^{-1}(x) = 2x$). Now let us consider the 2D case where the true transformation T is a 180° rotation about the origin, i.e. $T(x) = -x$. By definition, $T_2(y) = (T \circ T_1)(y)$, which leads to $T_1(y) + T_2(y) = 0$. This, however, cannot be simultaneously satisfied along with the constraint $T_1(y) + T_2(y) = 2y$ by any T_1 and T_2 .
- The transformations may be forced to be the inverses of each other, i.e. $T_2 = T_1^{-1}$, either explicitly, or through opposite-sign velocity fields. Given that $T = T_2 \circ T_1^{-1}$, such a constraint leads to $T = T_2 \circ T_2$, meaning that T_2 is the functional square root (half iterate) of T . However, Kuczma et al. (1990) have given examples of a 2D diffeomorphic function that does not have any functional square root (the authors’ Example 11.10.1), and a 1D diffeomorphic function that does not have a diffeomorphic functional square root (the authors’ Example 11.4.1). Accordingly, by using a transformation T_2 to parameterize the solution as $T = T_2 \circ T_2$, the space of all diffeomorphic transformations T will not be covered (even if T_2 is not restricted to be diffeomorphic).

¹ Multi-spectral images, $I_1, I_2: \Omega \rightarrow \mathbb{R}^p, p > 1$, can also be similarly incorporated in this framework.

² Throughout this article, we use the vectors x , y , and z to denote the space of I_1 , the mid-space, and the space of I_2 , respectively.

Since there is no unique way to define the mid-space, the registration results, T , can depend on the choice of the anti-drift constraints imposed on T_1 and T_2 . In fact, as seen above, some legitimate values of T can never be reached due to such constraints.

2.2. Implicit-atlas based registration

2.2.1. Implicit-atlas data term

Another approach to mid-space based pairwise registration is through constructing an unbiased *atlas* from the images (Ashburner and Ridgway, 2013; Hart et al., 2009; Joshi et al., 2004). In atlas construction, the observed images are assumed to be instances generated from a template image (atlas) with some geometric and intensity variation. Therefore, the problem reduces to finding an atlas, $A: \Omega \rightarrow \mathbb{R}$, and regular transformations T_1 and T_2 , where T_1^{-1} and T_2^{-1} take the atlas from the mid-space to the native spaces of the images, in such a way that the deformed versions of the atlas resemble the observed images. The SSD data term to be minimized in such an optimization is sum of the (asymmetric) subject-atlas distance terms, as follows:

$$\tilde{D}(I_1, I_2, A, T_1, T_2) := \int_{\Omega} \left(I_1(x) - (A \circ T_1^{-1})(x) \right)^2 dx + \int_{\Omega} \left(I_2(z) - (A \circ T_2^{-1})(z) \right)^2 dz. \quad (1)$$

Comparing the images to the atlas in the physically meaningful native image spaces complies with the generative model assumption that the image is generated as a deformed version of the atlas (not vice versa), and that the Gaussian noise is added to the deformed atlas (Allasonnière et al., 2007). By applying the change of variables $x = T_1(y)$ and $z = T_2(y)$ in each integral, the above data term can be written as:

$$\tilde{D}(I_1, I_2, A, T_1, T_2) = \int_{\Omega} \left[\left((I_1 \circ T_1)(y) - A(y) \right)^2 J_1(y) + \left((I_2 \circ T_2)(y) - A(y) \right)^2 J_2(y) \right] dy, \quad (2)$$

where $J_1(y) := \det \partial T_1(y)$ and $J_2(y) := \det \partial T_2(y)$ are the Jacobian determinants of the transformations. The $\hat{A}(y)$ minimizing the cost function is derived as (Ma et al., 2008):³

$$\hat{A}(y) = \frac{J_1(y)(I_1 \circ T_1)(y) + J_2(y)(I_2 \circ T_2)(y)}{J_1(y) + J_2(y)}. \quad (3)$$

In an explicit-atlas scheme, the transformations and the atlas can be iteratively computed from Eq. (1) and Eq. (3) (Hart et al., 2009).⁴ However, the atlas can indeed be eliminated from Eq. (2) by substituting $A(y)$ with $\hat{A}(y)$, leading to an implicit-atlas data term (Ashburner and Ridgway, 2013):

$$\tilde{D}(I_1, I_2, \hat{A}, T_1, T_2) = \int_{\Omega} \frac{\left((I_1 \circ T_1)(y) - (I_2 \circ T_2)(y) \right)^2}{J_1^{-1}(y) + J_2^{-1}(y)} dy. \quad (4)$$

³ \hat{A} is the minimizer of not only the data term, but the entire cost function, because the regularization term (Section 2.2.2) does not depend on the atlas.

⁴ Note that not every solution to which this atlas-based approach converges is necessarily an image-matching solution. For example, if the atlas (weighted average of the deformed images, Eq. (3)) happens to become a constant image during the optimization ($A(y) = c$), which is clearly not the desired image-matching solution, then the gradient of Eq. (1) with respect to the transformations will be zero and the optimization will converge to this spurious optimum.

As we will see in Section 2.3, keeping both T_1 and T_2 is redundant, which leads to further simplification of the data term by making it independent of the mid-space.

2.2.2. Regularization term

For the regularization term, we use the common Tikhonov integral that penalizes the deviation of the Jacobians of the transformations from the identity matrix. We want the following energy term to remain minimal:

$$\tilde{R}(T_1, T_2) := \int_{\Omega} \left\| \partial(T_1^{-1}(x)) - \mathbb{I} \right\|_F^2 dx + \int_{\Omega} \left\| \partial(T_2^{-1}(z)) - \mathbb{I} \right\|_F^2 dz, \quad (5)$$

where \mathbb{I} stands for the $d \times d$ identity matrix, and $\|\cdot\|_F$ is the Frobenius norm. Note that we regularize $T_1^{-1}(x)$ and $T_2^{-1}(z)$ and compute the integrals in the physically-meaningful native image spaces, as opposed to regularizing $T_1(y)$ and $T_2(y)$ in the abstract mid-space. In Section 2.3.2, we will solve the minimization of \tilde{R} with respect to one of the transformations, and rewrite it independent of the mid-space.

2.3. Independence from the mid-space

2.3.1. Mid-space-independent data term

Here we demonstrate that the implicit-atlas based data term is *inherently* independent of the mid-space. In other words, we can find the optimal transformations T between the two images, without solving for T_1 and T_2 that define the mid-space.

Equation (4) is a cost function of both T_1 and T_2 . Even so, further simplification reveals that the two transformations are redundant. The change of variables $y = T_1^{-1}(x)$, with $dy = dx/(J_1 \circ T_1^{-1})(x)$, reveals:

$$\tilde{D}(I_1, I_2, \hat{A}, T_1, T_2) = \int_{\Omega} \frac{\left(I_1(x) - (I_2 \circ T_2 \circ T_1^{-1})(x) \right)^2}{1 + \frac{(J_1 \circ T_1^{-1})(x)}{(J_2 \circ T_1^{-1})(x)}} dx. \quad (6)$$

Recall that the output transformation T is computed as $T(x) = (T_2 \circ T_1^{-1})(x)$, the Jacobian determinant of which is $J(x) := \det \partial T(x) = (J_2 \circ T_1^{-1})(x)/(J_1 \circ T_1^{-1})(x)$. It now becomes clear that the above integral is, remarkably, only a function of T , and that the data term can be written independently of the individual T_1 and T_2 , as:

$$D(I_1, I_2, T) := \int_{\Omega} \left(I_1(x) - (I_2 \circ T)(x) \right)^2 \frac{J(x)}{1 + J(x)} dx. \quad (7)$$

With the data term being independent of T_1 and T_2 , the mid-space disappears, eliminating the problem of the mid-space drift as well. We no longer need to enforce any anti-drift constraints, hence naturally avoiding limiting the space of possible transformations T by the particular choice of such constraints, and in fact widening the space of allowable transformations. This is while reducing the degrees of freedom of the optimization to half of those of the unconstrained problem of solving for both T_1 and T_2 .

Equation (7) might seem like a non-uniform integral in the native space of I_1 . However, this data term originates from Eq. (1), which is the sum of two integrals taken uniformly in the native spaces of images,

and non-uniformly only in the mid-space. Therefore, the pitfalls associated with non-uniform integration in native spaces (Aganj et al., 2015b) are not expected to arise.⁵ We have shown that Eq. (1) can be written independently of not only the atlas, as in Eq. (4), but also the mid-space itself, as in Eq (7).

2.3.2. Mid-space-independent regularization term

Contrary to the data term, the regularization in Eq. (5) depends on both T_1 and T_2 , and therefore on the mid-space. This is expected, since taking the mid-space far away from the image spaces would cause distortion in both transformations, thereby increasing $\tilde{R}(T_1, T_2)$.

Using $T_2 = T \circ T_1$, we first rewrite $\tilde{R}(T_1, T_2)$ in terms of T_1 and T :

$$\begin{aligned}\tilde{R}(T_1, T \circ T_1) &= \int_{\Omega} \left\| \partial(T_1^{-1}(x)) - \mathbb{I} \right\|_F^2 dx + \int_{\Omega} \left\| \partial((T_1^{-1} \circ T^{-1})(z)) - \mathbb{I} \right\|_F^2 dz \\ &= \int_{\Omega} \left\| \partial(T_1^{-1}(x)) - \mathbb{I} \right\|_F^2 dx + \int_{\Omega} \left\| (\partial(T_1^{-1}) \circ T^{-1})(z) (\partial T^{-1} \circ T^{-1})(z) - \mathbb{I} \right\|_F^2 dz.\end{aligned}\tag{8}$$

The change of variables $z = T(x)$ leads to:

$$\tilde{R}(T_1, T \circ T_1) = \int_{\Omega} \left(\left\| \partial(T_1^{-1}(x)) - \mathbb{I} \right\|_F^2 + \left\| \partial(T_1^{-1}(x)) \partial T^{-1}(x) - \mathbb{I} \right\|_F^2 J(x) \right) dx.\tag{9}$$

Using $\|X\|_F^2 = \text{tr}(X^T X)$ and then completing the square, one can verify that the above is equal to:

$$\begin{aligned}\tilde{R}(T_1, T \circ T_1) &= \int_{\Omega} \text{tr} \left[(\partial T(x) - \mathbb{I})^T \left(\mathbb{I} + \frac{1}{J(x)} \partial T(x) \partial T(x)^T \right)^{-1} (\partial T(x) - \mathbb{I}) \right] dx \\ &+ \int_{\Omega} \text{tr} \left[\left(\partial(T_1^{-1}(x)) - W(x) \right) \left(\mathbb{I} + J(x) \partial T^{-1}(x) \partial T^{-1}(x)^T \right) \left(\partial(T_1^{-1}(x)) - W(x) \right)^T \right] dx,\end{aligned}\tag{10}$$

where $W(x) := (J(x)\mathbb{I} + \partial T(x)^T)(J(x)\mathbb{I} + \partial T(x)\partial T(x)^T)^{-1}\partial T(x)$. Diffeomorphism ($J(x) > 0$) implies that the middle factor in each of the two traces in Eq. (10) is positive definite, and consequently the integrands of the two integrals are nonnegative. Since both the data term, $D(I_1, I_2, T)$, and the first integral in Eq. (10) are independent of T_1 , the cost function (i.e. weighted sum of data and regularization terms) reaches its minimum with respect to T_1 when the second integral in Eq. (10) equals zero, which happens for a \hat{T}_1 satisfying $\partial(\hat{T}_1^{-1}) = W$. We will not need the value of \hat{T}_1 though, because T alone will provide us with the complete solution to pairwise registration. Thus, we proceed by assigning $T_1 = \hat{T}_1$, thereby reducing the regularization term to its first integral:

⁵ Depending on the amount of volume change, the data term has the following local behavior:

$$(I_1(x) - (I_2 \circ T)(x))^2 \frac{J(x)}{1 + J(x)} dx \approx \begin{cases} (I_1(x) - (I_2 \circ T)(x))^2 dx, & J(x) \gg 1 \\ (I_1(x) - (I_2 \circ T)(x))^2 \frac{1 + J(x)}{4} dx, & J(x) \approx 1, \\ ((I_1 \circ T^{-1})(z) - I_2(z))^2 dz, & J(x) \ll 1 \end{cases}$$

where $z = T(x)$. This implies that, in regions with extreme volume change, the data term locally switches the reference space to the native space of the expanding image.

$$\begin{aligned}
R(T) &:= \min_{T_1} \tilde{R}(T_1, T \circ T_1) \\
&= \int_{\Omega} \text{tr} \left[(\partial T(x) - \mathbb{I})^T \left(\mathbb{I} + \frac{1}{J(x)} \partial T(x) \partial T(x)^T \right)^{-1} (\partial T(x) - \mathbb{I}) \right] dx.
\end{aligned} \tag{11}$$

The middle factor in the trace is what distinguishes the proposed $R(T)$ from the standard asymmetric Tikhonov regularization, $\int_{\Omega} \|\partial T(x) - \mathbb{I}\|_F^2 dx$.

2.3.3. Mid-space-independent cost function

Image registration can now be independent of the mid-space by minimizing a weighted sum of the proposed data and regularization terms, as:

$$\hat{T} = \underset{T}{\operatorname{argmin}} D(I_1, I_2, T) + \lambda R(T), \tag{12}$$

where λ is a manually determined positive constant. In the continuous case, this cost function is symmetric, as one can verify that $D(I_1, I_2, T) = D(I_2, I_1, T^{-1})$ and $R(T) = R(T^{-1})$. In Section 2.4, we will discuss how to alleviate any possible bias due to discretization artifacts.

2.4. Implementation

We minimize the cost function in Eq. (12) via gradient descent with line search. Similar to the diffeomorphic demons framework (Vercauteren et al., 2009), we minimize in alternative steps the data term and the regularization term, using the compositive and additive schemes, respectively.⁶

For the data term, we update T by composing it with a transformation S , which is chosen to be the exponential of the update field to preserve the diffeomorphism. To that end, at each iteration we fix T and compute the variation of $D(I_1, I_2, T \circ S)$ with respect to S , with S currently assumed to be the identity transformation ($\text{Id}(x) = x$). Following the chain rule for the derivative, we have:

$$\left. \frac{\delta D(I_1, I_2, T \circ S)}{\delta S} \right|_{S=\text{Id}} = \frac{\partial}{\partial S} (T \circ S) \left. \frac{\delta D(I_1, I_2, T \circ S)}{\delta (T \circ S)} \right|_{S=\text{Id}} = \partial T^T \frac{\delta D(I_1, I_2, T)}{\delta T}. \tag{13}$$

The notation (x) has been dropped for brevity. We use the Lemma in Appendix B to compute the variation of the data term as follows:

$$\left. \frac{\delta D(I_1, I_2, T \circ S)}{\delta S} \right|_{S=\text{Id}} = - \frac{2(I_1 - I_2 \circ T)J}{(1+J)^2} (\nabla I_1 + J \nabla (I_2 \circ T)) + (I_1 - I_2 \circ T)^2 \nabla \left[\left(\frac{J}{1+J} \right)^2 \right]. \tag{14}$$

The last term of Eq. (14) involves the Hessian of the deformation field, which can introduce large discretization error that possibly outweighs the accuracy that this term brings about. In our experiments, ignoring this second-derivative term improved the optimization of the cost function.

As for the regularization term, we derive its variation in Appendix C as:

⁶ We also tried a simultaneous compositive scheme for both terms, which however did not improve the optimization.

$$\frac{\delta R(T)}{\delta T} = -\operatorname{div} \left[2JP(\mathbb{I} - P^T \partial T) + \|P^T \partial T\|_F^2 C \right], \quad (15)$$

where $P := (J\mathbb{I} + \partial T \partial T^T)^{-1}(\partial T - \mathbb{I})$, the divergence is defined as $\operatorname{div} X := (\nabla^T X^T)^T$, and $C := J \partial T^{-1T}$ is the cofactor matrix of ∂T . Keeping the transformation diffeomorphic is especially important for the proposed method, given that the Jacobian appears in both the data and regularization terms. Therefore, we also include the following symmetric logarithmic barrier term to keep the Jacobian determinant away from zero:

$$L(T) := \int_{\Omega} (\log J(x))^2 (1 + J(x)) dx, \quad (16)$$

where $\gamma L(T)$ is added to $R(T)$ (and to the regularization terms of the other methods in Section 3). Note that $L(T) = L(T^{-1})$. In our experiments, the best optimization performance was obtained with $\gamma = 10^{-6}$, which is the value used in the reported results. We calculate the variation of $L(T)$ using the Lemma in Appendix B, as:

$$\frac{\delta L(T)}{\delta T} = -C \nabla \left[2 \left(1 + \frac{1}{J} \right) \log J + (\log J)^2 \right]. \quad (17)$$

We start the line search by initializing $S(x) \leftarrow x - \Delta \frac{\delta D}{\delta S} \Big|_{S=\operatorname{Id}}$ for a small step size Δ . Then, at step k of the line search (starting from $k = 1$), we compute the intermediate transformation $T^k \leftarrow T \circ S$ and regularize it with $T^k \leftarrow T^k - 2^{k-1} \Delta \frac{\delta R}{\delta T} \Big|_{T=T^k}$.⁷ Next we compute the value of the cost function for the transformation T^k , and compare it with those of the previous steps of the line search. If the current value of cost function has not been the minimum for M consecutive line-search steps, then we stop the line search and update T to be the intermediate transformation T^{k*} that minimized the cost function. Otherwise, inspired by the diffeomorphic demons (Vercauteren et al., 2009), we update $S \leftarrow S \circ S$ and move on to step $k + 1$ of the line search. We heuristically chose $M = 4$ and $M = 2$ for our 2D and 3D experiments, respectively. Note that the intermediate transformations T^k do not need all to be stored in the memory during the line search; only T^{k*} that minimizes the cost function needs to be kept in the memory and updated at each line-search step.

As already mentioned, the proposed continuous cost function in Eq. (12) is symmetric, meaning that it is unaffected by concurrently changing the order of the input images and inverting the transformation. We take a unidirectional cost-function optimization approach, such as in (Leow et al., 2007), which resamples I_2 , but not I_1 . Therefore, discretization artifacts may produce a symmetry-breaking bias. Yushkevich et al. (2010) have shown, however, that in hippocampal atrophy estimation most of such bias could be attributed to *linear* registration, rather than the subsequent deformable registration, and that symmetrizing the linear registration (Reuter et al., 2010) can remove the bias. Nonetheless, further steps can be taken to address the discretization bias in *deformable* registration, for instance, by avoiding the resampling step after interpolation, thereby eliminating the resampling artifacts and their resultant registration asymmetry. To

⁷ We noticed that the regularization improves if it is done μ times consecutively (we used $\mu = 10$), each time as $T_k \leftarrow T_k - \frac{2^{k-1} \Delta}{\mu} \frac{\delta R}{\delta T} \Big|_{T=T_k}$.

that end, the continuously interpolated integrals of the cost function can be either calculated analytically or estimated by quasi-random sampling (Aganj et al., 2013). An alternative strategy to reduce asymmetry due to discretization would be to optimize the cost function in a bidirectional fashion, i.e. with respect to both T and T^{-1} (Sabuncu et al., 2009). Although this is (in the continuous domain) still equivalent to optimizing the original unidirectional cost function, it will balance the resampling error on the two image spaces at the discretization step.

3. Results and discussion

We tested the proposed deformable registration method on brain T1-weighted magnetic resonance images, which had been pre-processed (intensity-normalized, skull-stripped, linearly co-registered, and resampled to 1-mm³ isotropic voxel) in FreeSurfer (Fischl, 2012). We used both 2D and 3D images in our experiments (in sections 3.1 and 3.2, respectively) to demonstrate the applicability of our method independently of the dimensionality of the images. To evaluate our mid-space-independent (MSI) registration data term, we compared it to two other data terms; first, the standard asymmetric SSD:

$$D_{\text{asymmetric}}(I_1, I_2, T) := \int_{\Omega} (I_1(x) - (I_2 \circ T)(x))^2 dx, \quad (18)$$

with the following compositive gradient:

$$\left. \frac{\delta}{\delta S} D_{\text{asymmetric}}(I_1, I_2, T \circ S) \right|_{S=\text{Id}} = -2(I_1 - I_2 \circ T) \nabla(I_2 \circ T); \quad (19)$$

second, with the popular symmetrization data term that minimizes the average of the asymmetric forward and backward data terms (Cachier and Rey, 2000; Christensen and Johnson, 2001; Tagare et al., 2009; Trouvé and Younes, 2000):

$$D_{\text{symmetrization}}(I_1, I_2, T) := \int_{\Omega} (I_1(x) - (I_2 \circ T)(x))^2 \frac{1 + J(x)}{2} dx, \quad (20)$$

with the following compositive gradient, computed using the Lemma in Appendix B:

$$\left. \frac{\delta}{\delta S} D_{\text{symmetrization}}(I_1, I_2, T \circ S) \right|_{S=\text{Id}} = -(I_1 - I_2 \circ T)(J \nabla I_1 + \nabla(I_2 \circ T)). \quad (21)$$

To assess the performance of the data terms independently of the regularization, we first compared the three (asymmetric, symmetrization, and MSI) data terms while employing the standard asymmetric regularization term $\int_{\Omega} \|\partial T(x) - \mathbb{I}\|_F^2 dx$ (with the variation $-2\nabla^2 T$) for all three of them. Subsequently, to evaluate the proposed regularization, we used the MSI data and regularization terms together in registration.⁸

⁸ For the symmetrization method, one could also use a symmetrized regularization term such as:

$$R_{\text{symmetrization}}(T) := \frac{1}{2} \left(\int_{\Omega} \|\partial T(x) - \mathbb{I}\|_F^2 dx + \int_{\Omega} \|\partial(T^{-1}(z)) - \mathbb{I}\|_F^2 dz \right),$$

which, by following similar steps as in Appendix C, can be seen to have the following variation:

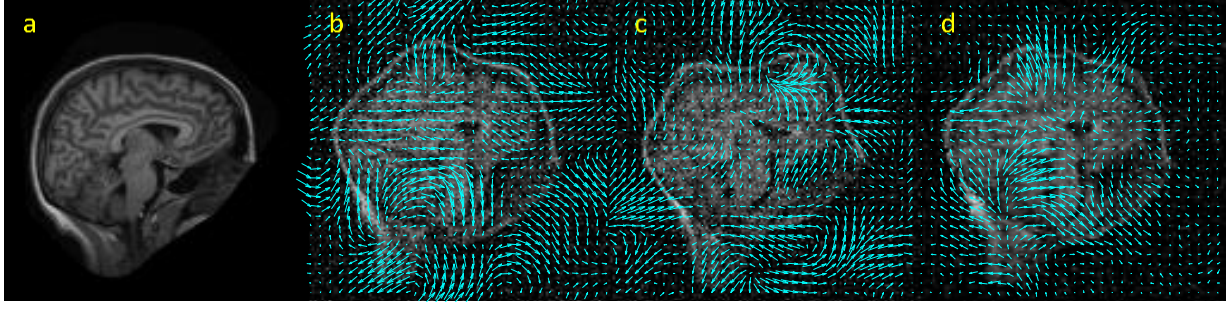


Figure 1. The original image (a) was deformed by two synthetic random transformations, and Gaussian noise was added (b,c). The deformed images were then registered with each other (d).

We implemented the deformable image registration (for all the mentioned cost functions) in Matlab using the displacement-field representation of the transformation.⁹ All input images had intensity values between 0 and 1, and we used the double-precision floating-point format to represent the images and transformations. Registration was done subsequently in three levels of resolution: quarter-, half-, and full-resolution. In the quarter- and half-resolution levels, we used cubic interpolation to down-sample the input images, and then to up-sample the output transformation to be used as the initial value for the next level of resolution.

The optimal values for the registration parameters generally vary among different cost functions. Thus, for a fair comparison, we considered the results of each cost function with the parameter values that lead to its optimal performance. For each pair of input images and for each method, we repeatedly ran the registration with a wide range of values for the regularization parameter λ and the step size Δ , and then chose the experiment with the best result (i.e. the one producing the optimal value for the criterion that is explained in each of the following subsections). In the following, we report the cross-subject mean of the optimal parameter values λ^* and Δ^* in each set of experiments (choosing extreme values for λ or Δ produced suboptimal results or caused the optimization to not converge).

3.1. Retrieval of synthetic 2D deformations

In the first set of experiments, we compared the methods on the mid-sagittal planes of 20 brain images taken from the publicly available Open Access Series of Imaging Studies (OASIS) database (Marcus et al., 2007), using synthetic deformations that provide us with the entire ground-truth transformation. From each volume, we extracted the sagittal slice, I , located four voxels to the right of the mid-sagittal plane, and resampled it to the size 128×128 (Figure 1a). For each subject, we synthesized two random deformation fields T_1^{synth} and T_2^{synth} , spatially low-pass filtered them, and applied them to I . Next, we added Gaussian noise (η_1 and η_2) with standard deviation of 0.1 to the deformed images to obtain two synthetic images, $I_1 = I \circ T_1^{\text{synth}} + \eta_1$ and $I_2 = I \circ T_2^{\text{synth}} + \eta_2$ (Figure 1b,c), which we then registered with each other using different methods (Figure 1d).¹⁰ The registration experiments consisted of 1000 gradient descent iterations

$$\frac{\delta}{\delta T} R_{\text{symmetrization}}(T) = -\text{div} \left[\partial T + \partial T^{-1T} (\mathbb{I} - \partial T^{-1}) C + \frac{1}{2} \|\partial T^{-1} - \mathbb{I}\|_F^2 C \right].$$

We tried this regularization, which, however, proved to be very unstable in practice. This is possibly because of the Jacobian matrix inversion, ∂T^{-1} , which is a numerically unstable operation due to discretization artifacts.

⁹ Our Matlab implementation is publicly available at: www.nitrc.org/projects/msi-register

¹⁰ The preprocessing scripts to generate the input data are available upon request.

at each of the three resolution levels. For each of the 20 subjects and each of the 4 methods, we repeated the registration with the regularization parameter λ taking a range of 50 different values (from 0.00001 to 1), and the step size Δ taking a range of 30 different values (from 0.01 to 256).

Ideally, we would expect to retrieve a transformation T satisfying $T_1^{\text{synth}} \approx T_2^{\text{synth}} \circ T$. Accordingly, we computed the error $\epsilon := \int_{\Omega} \left\| T_1^{\text{synth}}(x) - (T_2^{\text{synth}} \circ T)(x) \right\|_2^2 dx$, and for each subject and method, chose the best result (the one with the smallest ϵ) among the experiments with different values of λ and Δ . The cross-subject averages of the optimal ϵ^* , λ^* , and Δ^* are presented in Table 1, and a summary of the results of pairwise comparisons between different methods is presented in Table 2. As can be seen, all the methods significantly reduced the error compared to the pre-registration (i.e. $T(x) = x$) error. The proposed MSI data term with standard asymmetric regularization resulted in the smallest cross-subject average error. Specifically, the MSI data term significantly outperformed the symmetrization data term, but performed only slightly better than the asymmetric data term.¹¹ Regarding the proposed regularization, methods using standard asymmetric regularization resulted in significantly smaller errors than the one using MSI regularization did (see Section 3.2 for a discussion).

The runtimes averaged across the optimal experiments are shown in Table 1 (see Section 3.2 for hardware specifications). GPU parallelization did not offer any advantage in our 2D experiments.

¹¹ Please note that the definition of ϵ is inherently asymmetric, as the difference between T_1^{synth} and $T_2^{\text{synth}} \circ T$ is computed in the native space of image 1. Since the asymmetric data term, too, minimizes the error in the native space of image 1, ϵ may be biased towards favoring asymmetric approaches. To compute the error on the native space of image 2, as in $\int_{\Omega} \left\| (T_1^{\text{synth}} \circ T^{-1})(z) - T_2^{\text{synth}}(z) \right\|_2^2 dz$, we would have had to invert T , which is not straightforward with the displacement-field representation of the transformation, and introduces large numerical errors. Nevertheless, in Section 3.2, we will compare the methods differently and on both native spaces.

Table 1: Minimum error ϵ^* , optimal regularization parameter λ^* , optimal step size Δ^* , and the runtime of optimal experiment have been averaged for each method across subjects and presented along with their standard errors of the mean.

Mean across optimal experiments:	Before registration	Data term: Asymmetric Regularization: Asymmetric	Data term: Symmetrization Regularization: Asymmetric	Data term: MSI Regularization: Asymmetric	Data term: MSI Regularization: MSI
ϵ^*	5.602 ± 0.082	4.425 ± 0.075	4.470 ± 0.075	4.421 ± 0.070	4.701 ± 0.078
λ^*	-	0.0027 ± 0.0003	0.006 ± 0.001	0.004 ± 0.001	0.021 ± 0.002
Δ^*	-	29 ± 5	11 ± 3	49 ± 7	0.3 ± 0.1
Runtime (CPU)	-	8.8 ± 0.6 m	8.1 ± 0.2 m	7.9 ± 0.1 m	22.4 ± 0.5 m

Table 2. Pairwise comparison of the results in Section 3.1. In row i and column j , the percentage of the subjects for which method i outperformed method j (resulted in lower ϵ^*) is shown, along with the p -values for two-tailed paired Student's t - (p_t) and sign rank (p_s) tests.

Mean across optimal experiments:	Before registration	Data term: Asymmetric Regularization: Asymmetric	Data term: Symmetrization Regularization: Asymmetric	Data term: MSI Regularization: Asymmetric	Data term: MSI Regularization: MSI
Before registration	-	0% $p_t = 7 \times 10^{-17}$ $p_s = 9 \times 10^{-5}$	0% $p_t = 6 \times 10^{-16}$ $p_s = 9 \times 10^{-5}$	0% $p_t = 1 \times 10^{-15}$ $p_s = 9 \times 10^{-5}$	0% $p_t = 7 \times 10^{-15}$ $p_s = 9 \times 10^{-5}$
Data term: Asymmetric Regularization: Asymmetric	100% $p_t = 7 \times 10^{-17}$ $p_s = 9 \times 10^{-5}$	-	70% $p_t = 0.02$ $p_s = 0.03$	45% $p_t = 0.8$ $p_s = 0.8$	100% $p_t = 6 \times 10^{-8}$ $p_s = 9 \times 10^{-5}$
Data term: Symmetrization Regularization: Asymmetric	100% $p_t = 6 \times 10^{-16}$ $p_s = 9 \times 10^{-5}$	30% $p_t = 0.02$ $p_s = 0.03$	-	25% $p_t = 0.02$ $p_s = 0.01$	100% $p_t = 7 \times 10^{-7}$ $p_s = 9 \times 10^{-5}$
Data term: MSI Regularization: Asymmetric	100% $p_t = 1 \times 10^{-15}$ $p_s = 9 \times 10^{-5}$	55% $p_t = 0.8$ $p_s = 0.8$	75% $p_t = 0.02$ $p_s = 0.01$	-	100% $p_t = 5 \times 10^{-8}$ $p_s = 9 \times 10^{-5}$
Data term: MSI Regularization: MSI	100% $p_t = 7 \times 10^{-15}$ $p_s = 9 \times 10^{-5}$	0% $p_t = 6 \times 10^{-8}$ $p_s = 9 \times 10^{-5}$	0% $p_t = 7 \times 10^{-7}$ $p_s = 9 \times 10^{-5}$	0% $p_t = 5 \times 10^{-8}$ $p_s = 9 \times 10^{-5}$	-

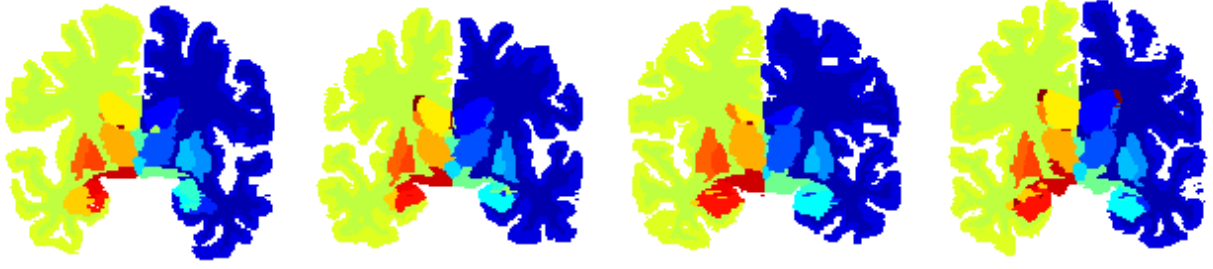


Figure 2. Coronal slices from 3D labels of four of the subjects used in Section 3.2.

3.2. Registration of labeled 3D images

We performed a second set of experiments using a dataset of 3D brain T1-weighted magnetic resonance images (Fischl et al., 2002). The dataset included 39 (14 male and 25 female) subjects, of which 28 were healthy and 11 had been diagnosed with dementia. The mean and standard deviation of the age was 55 and 23 years, respectively. This dataset additionally included manual-label volumes for 37 neuroanatomical structures (Figure 2), which had been used in the creation of the volumetric atlas of FreeSurfer (Fischl, 2012). To reduce the blank areas in the background, we cropped the images to the size $158 \times 182 \times 190$. Out of the $39 \times 38 = 1482$ possible combinations for ordered pairs, we randomly selected 40 pairs of images and registered each pair using the aforementioned methods. The registration experiments consisted of 100 gradient descent iterations at each of the three resolution levels. For each of the 40 pairs of subjects and each of the 4 methods, we repeated the registration with the regularization parameter λ taking a range of 30 different values (from 0.00001 to 1), and the step size Δ taking a range of 15 different values (from 0.01 to 500).

For validation and comparison, we considered 12 key subcortical regions: left/right amygdala, caudate, hippocampus, pallidum, putamen, and thalamus.¹² As a similarity score, we computed in the union of these regions the portion of the voxels with matching labels between the two images. The scores are defined respectively in the native spaces of image 1 and image 2 as:

$$\begin{aligned}\psi_1 &:= \frac{1}{|\Omega_1|} \sum_{x \in \Omega_1} \delta_{l_1(x), (l_2 \circ T)(x)}, \\ \psi_2 &:= \frac{1}{|\Omega_2|} \sum_{z \in \Omega_2} \delta_{(l_1 \circ T^{-1})(z), l_2(z)},\end{aligned}\tag{22}$$

where Ω_i is the set of all voxels inside the union of the 12 subcortical regions in image i , with $|\Omega_i|$ being its size; $\delta_{i,j} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$ is the Kronecker delta; and l_1 and l_2 are the label images that take 13 different integer values: 1 to 12 indicating each of the 12 subcortical regions, and 0 for a voxel outside of those 12 regions. Consequently, $\delta_{l_1(x), l_2(z)}$ is 1 if l_1 has the same label at voxel x as l_2 does at voxel z , and is 0 otherwise. For ψ_1 , we deformed the label volume of image 2 with the computed transformation T , and for

¹² Given that cortical regions are best aligned using *surface-based* registration (Fischl, 2012), we consider only subcortical regions here for the evaluation of our compared *volumetric* registration methods.

ψ_2 , we pushed the label values from image 1 forward to the space of image 2 according to T . We used the nearest-neighbor label interpolation in both cases. For each of the 4 methods and each of the 40 pairs of subjects, we chose the maximal ψ_1^* and the maximal ψ_2^* , achieved among the experiments with different values of λ and Δ . The cross-subject average of ψ_1^* in the native space of image 1, along with the averages of optimal λ^* and Δ^* are presented in Table 3, and a summary of the pairwise comparisons between different methods is presented in Table 4. Results of the label-matching analysis in the native space of image 2 (i.e., using ψ_2^*) are presented in Table 5 and Table 6.

All the methods significantly improved label matching compared to the pre-registration case. When we compared the labels in the native space of image 1, asymmetric registration produced the highest label-matching score, which was significantly higher than those by both the symmetrization method and (according to sign rank test but not t -test) the proposed MSI method. This is expected, given that the asymmetric method minimizes a cost function that is defined in the space of image 1. Conversely, when we compared the labels in the native space of image 2, the MSI data term with asymmetric regularization resulted in the highest label-matching score among the methods, which was significantly higher than that of the asymmetric data term. In both native spaces, the proposed MSI data term with asymmetric regularization resulted in higher label-matching score than the symmetrization data term did, which was significant in the native space of image 1.

Using the MSI regularization significantly reduced the label-matching score compared to using the standard asymmetric regularization. The sub-optimal performance of this regularization term may be due to numerical instabilities that are caused by the matrix inversion $(J\mathbb{I} + \partial T \partial T^T)^{-1}$, which is necessary to compute the variation of the regularization term in Eq. (15). Despite the fact that the exponential update (for the data term) and the logarithmic barrier keep the transformation mostly diffeomorphic, there still remain some voxels with non-positive Jacobian determinant in the transformation that can make the mentioned matrix inversion unstable. This is not an issue for the asymmetric regularization as the variation of the asymmetric regularization term, $-2\nabla^2 T$, is linear with respect to the transformation and not susceptible to non-diffeomorphic regions.

Next, we assessed the sensitivity of the label-matching scores to the ordering of the input images. Using the same optimal parameters, we repeated the experiments but this time with swapped input images ($I_1 \leftrightarrow I_2$), and computed ψ_1^{bw} and ψ_2^{bw} similarly to above for the new backward experiments. For symmetric registration, we expect the score of the forward registration in the reference space (ψ_1) to be similar to that of the backward registration in the moving image space (ψ_2^{bw}), and also the score of the forward registration in the moving image space (ψ_2) to be similar to that of the backward registration in the reference space (ψ_1^{bw}). Accordingly, we computed the following measure of label-matching asymmetry, whose mean is reported for different methods in Table 3 and Table 5:

$$\sigma := (|\psi_1 - \psi_2^{\text{bw}}| + |\psi_2 - \psi_1^{\text{bw}}|)/2. \quad (23)$$

As expected, the asymmetric data term resulted in a larger σ than the two symmetric data terms did (with asymmetric regularization). Discretization artifacts – during both registration and score computation – are an important factor contributing to the positive values of σ for the symmetric cost functions (see Section 2.4). The values of σ are considerably smaller for optimal transformations that are chosen based on label-matching in the space of image 2 (Table 5), where labels in the moving image space are supposedly better

defined due to smoother inverse transformations. Asymmetric regularization is another factor contributing to an elevated σ . Using the MSI (instead of asymmetric) regularization helped to reduce σ for the MSI data term when transformations had been optimized based on ψ_1 (but not when based on ψ_2).

To assess the smoothness of the optimal transformations, we also computed the percentage of voxels where the Jacobian determinant of the optimal transformation was non-positive ($J(x) \leq 0$), shown in Table 3 and Table 5. This value is lowest for the symmetrization data term. Furthermore, for all the methods, this value is significantly lower when optimal transformations are chosen based on label matching in the native space of image 2. This may be because a lower number of voxels with non-positive Jacobian determinant implies a smoother inverse transform, hence a better label definition in the space of image 2.

Although we distributed the bulk of our experiments on various hardware platforms, we used consistent hardware for the abovementioned inverse optimal experiments while measuring the runtimes, which are shown in Table 3 and Table 5. When the asymmetric regularization was used, the runtimes were roughly in the same range for the three data terms, while the experiments with optimal label matching in the space of image 2 were faster than those with optimal matching in the space of image 1. We observed that the differences between the runtimes were mainly because of the variation in the total number of line-search iterations performed in the experiments, to which the runtime is proportional. When the MSI regularization was used, registration took about 4 times longer, mostly due to the matrix-inversion operations in Eq. (11) and Eq. (15). For the CPU experiments, we used processors based on the Intel Xeon E5-2667 v3 Haswell 3.2 GHz (3.6 GHz with turbo) with DDR 4 memory. We ran two experiments simultaneously on each 8-core 56GB-RAM virtual machine. (Note that Matlab may internally multi-thread some of its matrix operations.) Additionally, we repeated each optimal experiment using an NVIDIA Tesla K80 graphics processing unit (GPU) and 1 CPU core, which made the 3D registration on average 16 times faster.

Table 3: Maximum label-matching score ψ_1^* in the native space of *image 1*, optimal regularization parameter λ^* , optimal step size Δ^* , label-matching asymmetry σ , the portion of voxels with $J \leq 0$, and the runtime of the optimal experiment have been averaged for each method across subjects and presented along with their standard errors of the mean.

Mean across optimal experiments:	Before registration	Data term: Asymmetric Regularization: Asymmetric	Data term: Symmetrization Regularization: Asymmetric	Data term: MSI Regularization: Asymmetric	Data term: MSI Regularization: MSI
ψ_1^*	0.638 ± 0.012	0.849 ± 0.004	0.835 ± 0.004	0.847 ± 0.003	0.784 ± 0.006
λ^*	-	0.003 ± 0.001	0.006 ± 0.001	0.003 ± 0.001	0.0006 ± 0.0002
Δ^*	-	143 ± 13	59 ± 9	151 ± 15	3.1 ± 0.5
σ	-	0.099 ± 0.008	0.023 ± 0.003	0.083 ± 0.007	0.050 ± 0.003
$J \leq 0$	-	$(2.2 \pm 0.2) \%$	$(0.08 \pm 0.02) \%$	$(1.7 \pm 0.3) \%$	$(4.2 \pm 0.3) \%$
Runtime (CPU)	-	8.4 ± 0.3 hr	7.9 ± 0.7 hr	7.7 ± 0.3 hr	26.9 ± 0.4 hr
Runtime (GPU)	-	30 ± 1 m	29 ± 3 m	30 ± 2 m	115 ± 2 m

Table 4. Pairwise comparison of the label-matching results in the native space of *image 1*. In row i and column j , the percentage of the subjects for which method i outperformed method j (resulted in higher ψ_1^*) is shown, along with the p -values for two-tailed paired Student's t - (p_t) and sign rank (p_s) tests.

Mean across optimal experiments:	Before registration	Data term: Asymmetric Regularization: Asymmetric	Data term: Symmetrization Regularization: Asymmetric	Data term: MSI Regularization: Asymmetric	Data term: MSI Regularization: MSI
Before registration	-	0% $p_t = 3 \times 10^{-20}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 7 \times 10^{-20}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 2 \times 10^{-20}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 7 \times 10^{-17}$ $p_s = 4 \times 10^{-8}$
Data term: Asymmetric Regularization: Asymmetric	100% $p_t = 3 \times 10^{-20}$ $p_s = 4 \times 10^{-8}$	-	92.5% $p_t = 2 \times 10^{-8}$ $p_s = 1 \times 10^{-7}$	72.5% $p_t = 0.13$ $p_s = 0.01$	100% $p_t = 6 \times 10^{-15}$ $p_s = 4 \times 10^{-8}$
Data term: Symmetrization Regularization: Asymmetric	100% $p_t = 7 \times 10^{-20}$ $p_s = 4 \times 10^{-8}$	7.5% $p_t = 2 \times 10^{-8}$ $p_s = 1 \times 10^{-7}$	-	10% $p_t = 3 \times 10^{-7}$ $p_s = 1 \times 10^{-7}$	100% $p_t = 3 \times 10^{-12}$ $p_s = 4 \times 10^{-8}$
Data term: MSI Regularization: Asymmetric	100% $p_t = 2 \times 10^{-20}$ $p_s = 4 \times 10^{-8}$	27.5% $p_t = 0.13$ $p_s = 0.01$	90% $p_t = 3 \times 10^{-7}$ $p_s = 1 \times 10^{-7}$	-	100% $p_t = 2 \times 10^{-14}$ $p_s = 4 \times 10^{-8}$
Data term: MSI Regularization: MSI	100% $p_t = 7 \times 10^{-17}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 6 \times 10^{-15}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 3 \times 10^{-12}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 2 \times 10^{-14}$ $p_s = 4 \times 10^{-8}$	-

Table 5: Maximum label-matching score ψ_2^* in the native space of *image 2*, optimal regularization parameter λ^* , optimal step size Δ^* , label-matching asymmetry σ , the portion of voxels with $J \leq 0$, and the runtime of the optimal experiment have been averaged for each method across subjects and presented along with their standard errors of the mean.

Mean across optimal experiments:	Before registration	Data term: Asymmetric Regularization: Asymmetric	Data term: Symmetrization Regularization: Asymmetric	Data term: MSI Regularization: Asymmetric	Data term: MSI Regularization: MSI
ψ_2^*	0.6218 ± 0.0161	0.8143 ± 0.0065	0.8158 ± 0.0059	0.8159 ± 0.0063	0.7365 ± 0.0089
λ^*	-	0.004 ± 0.001	0.005 ± 0.001	0.005 ± 0.001	0.003 ± 0.001
Δ^*	-	35 ± 2	25 ± 2	34 ± 2	0.3 ± 0.1
σ	-	0.010 ± 0.003	0.005 ± 0.001	0.006 ± 0.001	0.033 ± 0.003
$J \leq 0$	-	$(0.033 \pm 0.008) \%$	$(0.011 \pm 0.003) \%$	$(0.018 \pm 0.005) \%$	$(1.8 \pm 0.2) \%$
Runtime (CPU)	-	6.19 ± 0.05 hr	6.32 ± 0.06 hr	6.17 ± 0.03 hr	26.6 ± 0.6 hr
Runtime (GPU)	-	22.0 ± 0.1 m	22.3 ± 0.2 m	21.82 ± 0.04 m	119 ± 2 m

Table 6. Pairwise comparison of the label-matching results in the native space of *image 2*. In row i and column j , the percentage of the subjects for which method i outperformed method j (resulted in higher ψ_2^*) is shown, along with the p -values for two-tailed paired Student's t - (p_t) and sign rank (p_s) tests.

Mean across optimal experiments:	Before registration	Data term: Asymmetric Regularization: Asymmetric	Data term: Symmetrization Regularization: Asymmetric	Data term: MSI Regularization: Asymmetric	Data term: MSI Regularization: MSI
Before registration	-	0% $p_t = 3 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 5 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 2 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 3 \times 10^{-15}$ $p_s = 4 \times 10^{-8}$
Data term: Asymmetric Regularization: Asymmetric	100% $p_t = 3 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	-	45% $p_t = 0.2$ $p_s = 0.4$	30% $p_t = 0.002$ $p_s = 0.004$	100% $p_t = 5 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$
Data term: Symmetrization Regularization: Asymmetric	100% $p_t = 5 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	55% $p_t = 0.2$ $p_s = 0.4$	-	30% $p_t = 0.9$ $p_s = 0.06$	100% $p_t = 5 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$
Data term: MSI Regularization: Asymmetric	100% $p_t = 2 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	67.5% $p_t = 0.002$ $p_s = 0.004$	70% $p_t = 0.9$ $p_s = 0.06$	-	100% $p_t = 1 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$
Data term: MSI Regularization: MSI	100% $p_t = 3 \times 10^{-15}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 5 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 5 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	0% $p_t = 1 \times 10^{-18}$ $p_s = 4 \times 10^{-8}$	-

Lastly, to see the range of label-matching scores that existing state-of-the-art registration toolboxes would produce on our data, we repeated the experiments using Advanced Normalization Tools (ANTs) Symmetric Normalization (SyN) (Avants et al., 2008) and FSL FMRIB’s Nonlinear Image Registration Tool (FNIRT) (Andersson et al., 2007). We used the same three levels of resolution and a maximum of 100 iterations per level, although registration often converged in less than 100 iterations. The cross-subject mean of (ψ_1, ψ_2) was $(0.74 \pm 0.02, 0.72 \pm 0.02)$ for ANTs SyN and $(0.80 \pm 0.01, 0.70 \pm 0.02)$ for FNIRT. Since these label-matching scores were computed using the toolboxes’ default or suggested parameter values¹³, they should not be directly compared to the values in Table 3 and Table 5, which have been optimized by an exhaustive search in the parameter space.

4. Conclusions

We have demonstrated for the first time that implicit-atlas image registration is inherently independent of the mid-space, which led to deriving a new data term for deformable image registration. We have also derived a new regularization term by analytically reformulating the regularization independently of the mid-space. The independence of the cost function from the image-to-atlas transformations alleviates the need for enforcing artificial anti-drift constraints that potentially restrict the space of possible deformations. We have evaluated our method through experiments on 2D and 3D brain magnetic resonance image datasets, and shown that the proposed data term often outperformed competing methods. Future research includes: stabilizing the MSI regularization via a proper projection of the Jacobian matrix onto the space of diffeomorphism, detailed implementation and thorough validation of MSI *group-wise* deformable registration, and incorporation of the MSI cost function in surface-based registration.

Acknowledgments

Support for this research was provided by the National Institutes of Health (NIH), specifically the National Institute of Diabetes and Digestive and Kidney Diseases (K01DK101631, R21DK108277), the National Institute for Biomedical Imaging and Bioengineering (P41EB015896, R01EB006758, K25EB013649, R21EB018907, R01EB019956), the National Institute on Aging (AG022381, 5R01AG008122-22, R01AG016495-11, R01AG016495), the National Center for Alternative Medicine (RC1AT005728-01), the National Institute for Neurological Disorders and Stroke (R01NS052585, R21NS072652, R01NS070963, R01NS083534, U01NS086625), the National Cancer Institute (K25CA181632), and the NIH Blueprint for Neuroscience Research (U01MH093765), part of the multi-institutional Human Connectome Project.

¹³ We used the following options for each toolbox:

```
For antsRegistration, release 1.9: --metric MeanSquares[<image1>,<image2>,1,0] --
transform SyN[0.25,3,0] --shrink-factors 4x2x1 --smoothing-sigmas 2x1x0 --
convergence 100x100x100
```

```
For fnirt, release 5.0.7: --subsamp=4,2,1 --infwhm=4,2,0 --reffwhm=4,2,0 --
miter=100,100,100 --estint=1,1,0 --applyrefmask=0,0,0 --applyinmask=0,0,0
```

Note that an accurate and fair assessment among mathematical cost functions requires all implementation factors (except the cost function) to be fixed in the experiments; otherwise, the disparities in different implementations (e.g., the choices of the representation of the transformation, the update scheme for the transformation, the optimization algorithm, etc.) could dominate the differences between the cost functions that we wish to compare. The cost-function comparison reported in the tables satisfies this requirement, as we used a consistent implementation for it.

Additional support was provided by the BrightFocus Foundation (A2016172S, A2012333), the Massachusetts Alzheimer’s Disease Research Center (P50AG005134), and the Ellison Medical Foundation (Autism & Dyslexia Project). JEI was supported by the Spanish Ministry of Economy and Competitiveness (TEC2014-51882-P), the European Union’s Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant 654911, project “THALAMODEL”), and the European Research Council (ERC Starting Grant no. 677697 “BUNGEE-TOOLS”). Computational resources were provided through NIH Shared Instrumentation Grants (S10RR023401, S10RR019307, S10RR023043, S10RR028832), a Microsoft Azure for Research Award, and an NVIDIA Corporation GPU Grant. The OASIS project, the data of which were used in Section 3.1, is also supported by the NIH (P50AG05681, P01AG03991, R01AG021910, P50MH071616, U24RR021382, R01MH56584).

BF has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. BF’s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

Appendix A. Extension to group-wise registration

In this appendix, we show how our MSI framework can be extended for the registration of N images $I_1, \dots, I_N: \Omega \rightarrow \mathbb{R}$, where $\Omega \subseteq \mathbb{R}^d$. For group-wise registration, we want to compute the set of regular transformations $\{T_{ij}: \Omega \rightarrow \Omega \mid i, j = 1, \dots, N\}$, where T_{ij} deforms I_i so as to make $I_i \circ T_{ij}$ and I_j most similar to each other. In a common approach to group-wise registration, the pairwise transformations $\{T_{ij}\}$ are parameterized as $T_{ij} = T_i \circ T_j^{-1}$, where T_1, \dots, T_N are transformations that take I_1, \dots, I_N to a mid-space (see Section 1 for references to the mid-space approaches).

Similar to the pairwise case of Eq. (1), we define the SSD data term as:

$$\tilde{D}(I_1, \dots, I_N, A; T_1, \dots, T_N) := \sum_{i=1}^N \int_{\Omega} \left(I_i(x) - (A \circ T_i^{-1})(x) \right)^2 dx. \quad (24)$$

After applying the change of variables $x = T_i(y)$ in each integral, the $\hat{A}(y)$ minimizing the cost function is derived as (Ma et al., 2008):

$$\hat{A}(y) = \frac{\sum_i J_i(y) (I_i \circ T_i)(y)}{\sum_i J_i(y)}, \quad (25)$$

where $J_i(y) := \det \partial T_i(y)$. The atlas can now be eliminated from Eq. (24) by substituting $A(y)$ with $\hat{A}(y)$, leading to an implicit-atlas data term:

$$\begin{aligned}
\tilde{D}(I_1, \dots, I_N, \hat{A}; T_1, \dots, T_N) &= \int_{\Omega} \sum_{i=1}^N \left((I_i \circ T_i)(y) - \hat{A}(y) \right)^2 J_i(y) dy \\
&= \int_{\Omega} \sum_{i=1}^N \left((I_i \circ T_i)(y) - \frac{\sum_k J_k(y) (I_k \circ T_k)(y)}{\sum_k J_k(y)} \right)^2 J_i(y) dy \\
&= \sum_{i=1}^N \int_{\Omega} (I_i^2 \circ T_i)(y) J_i(y) dy + \int_{\Omega} \left(\frac{\sum_k J_k(y) (I_k \circ T_k)(y)}{\sum_k J_k(y)} \right)^2 \sum_{i=1}^N J_i(y) dy \quad (26) \\
&\quad - 2 \int_{\Omega} \frac{\sum_k J_k(y) (I_k \circ T_k)(y)}{\sum_k J_k(y)} \sum_{i=1}^N J_i(y) (I_i \circ T_i)(y) dy \\
&= \int_{\Omega} \sum_{i=1}^N I_i^2(x) dx - \int_{\Omega} \frac{(\sum_i J_i(y) (I_i \circ T_i)(y))^2}{\sum_i J_i(y)} dy.
\end{aligned}$$

The first term of the expression derived above, which was obtained by the changes of variables $x = T_i(y)$, is constant with respect to the transformations, and so can be ignored in the optimization. Consequently, the problem is reduced to *maximizing* the following data term:

$$D(I_1, \dots, I_N; T_1, \dots, T_N) := \int_{\Omega} \frac{(\sum_i J_i(y) (I_i \circ T_i)(y))^2}{\sum_i J_i(y)} dy. \quad (27)$$

An important property of this data term is that composing T_1, \dots, T_N with a new transformation S does not change D :

$$\begin{aligned}
D(I_1, \dots, I_N; T_1 \circ S, \dots, T_N \circ S) &= \int_{\Omega} \frac{(\sum_i (J_i \circ S)(z) (I_i \circ T_i \circ S)(z))^2}{\sum_i (J_i \circ S)(z) J_S(z)} dz \\
&= \int_{\Omega} \frac{(\sum_i (J_i \circ S)(z) (I_i \circ T_i \circ S)(z))^2}{\sum_i (J_i \circ S)(z)} J_S(z) dz \quad (28) \\
&= \int_{\Omega} \frac{(\sum_i J_i(y) (I_i \circ T_i)(y))^2}{\sum_i J_i(y)} dy \\
&= D(I_1, \dots, I_N; T_1, \dots, T_N),
\end{aligned}$$

where $J_S(z) := \det \partial S(z)$, and $y := S(z)$ with $dy = J_S(z) dz$. The shrinkage-type issues that were mentioned in Section 2.1 are alleviated because of this property of the atlas-based registration, which is thanks to the fact that the atlas and the images are compared in the physically-meaningful native image spaces, as opposed to the abstract mid-space. This property also implies the inherent independence of the data term from the mid-space, which we exploit here. We first choose the native space of one of the images – say, the first image – as a hub, and define $Q_i := T_{i1} = T_i \circ T_1^{-1}$. Note that $Q_1 = \text{Id}$ is fixed as the identity transformation. We then rewrite the objective function with respect to T_1 and Q_2, \dots, Q_N . To obtain $\{T_{ij}\}$, it will be sufficient to solve for Q_2, \dots, Q_N (but not T_1), since $T_{ij} = T_i \circ T_j^{-1} = Q_i \circ T_1 \circ T_1^{-1} \circ Q_j^{-1} = Q_i \circ Q_j^{-1}$. With the new parameterization, the data term to be maximized becomes:

$$\begin{aligned}
D(I_1, \dots, I_N; T_1, \dots, T_N) &= D(I_1, \dots, I_N; T_1, Q_2 \circ T_1, \dots, Q_N \circ T_1) \\
&= D(I_1, \dots, I_N; \text{Id}, Q_2, \dots, Q_N), \quad (29)
\end{aligned}$$

where we used the composition property in Eq. (28). With the data term being independent of T_1, \dots, T_N , the mid-space disappears, thereby eliminating the problem of the mid-space drift and the need for anti-drift constraints as well.

As for the Tikhonov regularization term, we define similarly to Eq. (5):

$$\tilde{R}(T_1, \dots, T_N) := \sum_{i=1}^N \int_{\Omega} \left\| \partial \left(T_i^{-1}(x) \right) - \mathbb{I} \right\|_F^2 dx. \quad (30)$$

Choosing I_1 as a hub again, by following similar calculations as in Section 2.3.2, minimizing $\tilde{R}(T_1, \dots, T_N)$ can be seen to be equivalent to *maximizing* the following:

$$\int_{\Omega} \left(\left\| V(x) U^{-1}(x)^T \right\|_F^2 - \left\| \partial \left(T_1^{-1}(x) \right) U(x) - V(x) U^{-1}(x)^T \right\|_F^2 \right) dx, \quad (31)$$

where:

$$\begin{aligned} U(x) U(x)^T &:= \sum_{i=1}^N \partial Q_i^{-1}(x) C_i(x), \\ V(x) &:= \sum_{i=1}^N C_i(x). \end{aligned} \quad (32)$$

$C_i(x) := J_{Q_i}(x) \partial Q_i^{-1}(x)^T$ is the cofactor matrix of $\partial Q_i(x)$, with $J_{Q_i}(x) := \det \partial Q_i(x)$.¹⁴ Note that diffeomorphism ($J_i(x) > 0$) guarantees the existence of $U(x)$ that satisfies Eq. (32), since the right hand side, $\sum_i \partial Q_i^{-1}(x) C_i(x) = \sum_i J_{Q_i}(x) \partial Q_i^{-1}(x) \partial Q_i^{-1}(x)^T$, is a linear combination of positive semi-definite matrices with positive weighting, and therefore is itself positive semi-definite.

Since the data term, $D(I_1, \dots, I_N; \text{Id}, Q_2, \dots, Q_N)$, is independent of T_1 , the objective function reaches its maximum value with respect to T_1 when the second Frobenius norm in Eq. (31) equals zero, which happens for \hat{T}_1 that satisfies $\partial \left(\hat{T}_1^{-1}(x) \right) = V(x) (U(x) U(x)^T)^{-1}$. As mentioned before, we do not need the value of \hat{T}_1 , since the $N - 1$ transformations Q_2, \dots, Q_N provide us with the complete solution to group-wise registration. Thus, we proceed by assigning $T_1 = \hat{T}_1$, thereby reducing the regularization term to only its first Frobenius norm, expanded as:

$$R(Q_2, \dots, Q_N) := \int_{\Omega} \text{tr} \left[\left(\sum_{i=1}^N C_i(x) \right) \left(\sum_{i=1}^N \partial Q_i^{-1}(x) C_i(x) \right)^{-1} \left(\sum_{i=1}^N C_i(x) \right)^T \right] dx. \quad (33)$$

Therefore, registration will be performed by maximizing the following objective function:

$$\hat{Q}_2, \dots, \hat{Q}_N = \underset{Q_2, \dots, Q_N}{\operatorname{argmax}} D(I_1, \dots, I_N; \text{Id}, Q_2, \dots, Q_N) + \lambda R(Q_2, \dots, Q_N), \quad (34)$$

¹⁴ Recall that $Q_1(y) = y$, $\partial Q_1(y) = \mathbb{I}$, $C_1(y) = \mathbb{I}$, and $J_{Q_1}(y) = 1$.

where λ is the positive regularization parameter. In the continuous case, this objective function – originating from Eq. (24) and Eq. (30) – is unbiased towards the choice of the hub image (here I_1). See Section 2.4 for a discussion on discretization artifacts.

Note that using the MSI group-wise registration is not helpful for template construction, since the template exists in the mid-space that is eliminated.

Appendix B. Variation of an integral that includes the Jacobian determinant

We use the following Lemma to compute the derivative of an integral term with respect to a transformation.

Lemma. For $T: \Omega \rightarrow \Omega$ (with $\Omega \subseteq \mathbb{R}^d$), $F: \Omega^2 \times \mathbb{R}^+ \rightarrow \mathbb{R}$, and $J(x) := \det \partial T(x)$, let:

$$\varphi := \int_{\Omega} F(T(x), J(x), x) dx. \quad (35)$$

The variation of φ with respect to T is:

$$\frac{\delta \varphi}{\delta T} = \nabla_T F - (J \nabla_T + C \nabla_x) \frac{\partial F}{\partial J} - \frac{\partial^2 F}{\partial J^2} C \nabla J, \quad (36)$$

where C is the cofactor matrix of the Jacobian matrix ∂T , and ∇_T and ∇_x are vectors of gradient with respect to the elements of T and x , respectively. The notation (x) has been dropped for brevity.

Proof. We start by using the following variation formula:

$$\frac{\delta \varphi}{\delta T} = \nabla_T F - \operatorname{div} \nabla_{\partial T} F, \quad (37)$$

where $\nabla_{\partial T}$ is the matrix of the gradient with respect to the elements of ∂T , and $\operatorname{div} X := (\nabla^T X^T)^T$. Since in Eq. (35), ∂T only appears in the Jacobian determinant, J , the chain rule implies that:

$$\nabla_{\partial T} F = \frac{\partial F}{\partial J} \nabla_{\partial T} J. \quad (38)$$

Next we apply the product rule:

$$\operatorname{div} \nabla_{\partial T} F = (\nabla_{\partial T} J) \nabla \left(\frac{\partial F}{\partial J} \right) + \frac{\partial F}{\partial J} \operatorname{div} \nabla_{\partial T} J. \quad (39)$$

According to Eq. (37), $\operatorname{div} \nabla_{\partial T} J$ is the variation of $-\int_{\Omega} J(x) dx$ with respect to T . The change of variables $z := T(x)$, with $dz = J(x) dx$, however, reveals that this integral is constant, $-\int_{\Omega} J(x) dx = -\int_{\Omega} dz = -|\Omega|$, making its variation vanish: $\operatorname{div} \nabla_{\partial T} J = 0$.

Given that $\nabla_{\partial T} J = C$, the first term in the right-hand side of Eq. (39) can be expanded by the chain rule as:

$$(\nabla_{\partial T} J) \nabla \left(\frac{\partial F}{\partial J} \right) = C \nabla \left(\frac{\partial F}{\partial J} \right) = C \nabla_x \left(\frac{\partial F}{\partial J} \right) + C \partial T^T \nabla_T \left(\frac{\partial F}{\partial J} \right) + \frac{\partial^2 F}{\partial J^2} C \nabla J. \quad (40)$$

The matrix inversion rule, $\partial T^{-1} = \frac{1}{J} C^T$, leads to $C \partial T^T = J \mathbb{I}$. Therefore:

$$(\nabla_{\partial T} J) \nabla \left(\frac{\partial F}{\partial J} \right) = C \nabla_x \left(\frac{\partial F}{\partial J} \right) + J \nabla_T \left(\frac{\partial F}{\partial J} \right) + \frac{\partial^2 F}{\partial J^2} C \nabla J. \quad (41)$$

Combining Eqs. (37,39,41) proves the Lemma.

□

A particular use of this lemma is in compositive updates, where the following needs to be computed:

$$\partial T^T \frac{\delta \varphi}{\delta T} = \partial T^T \nabla_T \left(F - J \frac{\partial F}{\partial J} \right) - J \nabla_x \frac{\partial F}{\partial J} - \frac{\partial^2 F}{\partial J^2} J \nabla J. \quad (42)$$

In particular, for the weighted SSD, $F(T(x), J(x), x) = (I_1(x) - (I_2 \circ T)(x))^2 f(J(x))$, we have:

$$\partial T^T \frac{\delta \varphi}{\delta T} = -2(I_1 - I_2 \circ T) [(f(J) - g(J)) \nabla I_1 + g(J) \nabla (I_2 \circ T)] + (I_1 - I_2 \circ T)^2 \nabla(g(J)), \quad (43)$$

where $g(J) := f(J) - J f'(J)$. Letting $f(J) = \frac{J}{1+J}$ and $f(J) = \frac{1+J}{2}$ lead to Eq. (14) and Eq. (21), respectively.

Appendix C. Variation of the regularization integral

We compute the variation of $R(T)$ (Eq. (11)), starting from Eq. (37) with $F = \text{tr}[\Gamma^T \Lambda^{-1} \Gamma]$, where $\Gamma := \partial T - \mathbb{I}$ and $\Lambda := \mathbb{I} + \frac{1}{J} \partial T \partial T^T$. Since F only depends on the derivative of T (but not on T itself), we have $\nabla_T F = 0$, and:

$$\frac{\delta R(T)}{\delta T} = -\text{div} \nabla_{\partial T} F. \quad (44)$$

We now compute $(\nabla_{\partial T} F)_{ij}$ (the element i, j of the matrix $\nabla_{\partial T} F$) by deriving F with respect to ∂T_{ij} . For brevity, we shall fix i and j and denote the derivative of a matrix X with respect to ∂T_{ij} simply as X' . Using the differentiation rules for the inverse, $(\Lambda^{-1})' = -\Lambda^{-1} \Lambda' \Lambda^{-1}$, and the product, one can verify that:

$$F' = \text{tr}[\Gamma^T \Lambda^{-1} \Gamma]' = \text{tr}[(2\Gamma'^T - \Gamma^T \Lambda^{-1} \Lambda') \Lambda^{-1} \Gamma]. \quad (45)$$

By definition, $\Gamma' = (\partial T - \mathbb{I})' = \Theta^{ij}$, where Θ^{ij} is the $d \times d$ matrix that has 1 as its element i, j and 0 elsewhere; i.e., $(\Theta^{ij})_{kl} = \delta_{ki} \delta_{lj}$, with δ the Kronecker delta. As for Λ' , we have:

$$\Lambda' = \left(\mathbb{I} + \frac{1}{J} \partial T \partial T^T \right)' = -\frac{C_{ij}}{J^2} \partial T \partial T^T + \frac{1}{J} \Theta^{ij} \partial T^T + \frac{1}{J} \partial T \Theta^{ji}, \quad (46)$$

where we used $J' = C_{ij}$, with C the cofactor matrix of ∂T . By substituting for Γ , Γ' , Λ , and Λ' in Eq. (45), while using the trace properties $\text{tr}[X^T X] = \|X\|_F^2$ and $\text{tr}[\Theta^{ij} X] = X_{ji}$, Eq. (44) leads to the gradient presented in Eq. (15).

References

- Aganj, I., Iglesias, J.E., Reuter, M., Sabuncu, M.R., Fischl, B., 2015a. Mid-space-independent symmetric data term for pairwise deformable image registration. Medical Image Computing and Computer-Assisted Intervention – MICCAI, Munich, Germany.
- Aganj, I., Reuter, M., Sabuncu, M.R., Fischl, B., 2015b. Avoiding symmetry-breaking spatial non-uniformity in deformable image registration via a quasi-volume-preserving constraint. *NeuroImage* 106, 238-251.
- Aganj, I., Yeo, B.T.T., Sabuncu, M.R., Fischl, B., 2013. On removing interpolation and resampling artifacts in rigid image registration. *Image Processing, IEEE Transactions on* 22, 816-827.
- Aljabar, P., Bhatia, K.K., Murgasova, M., Hajnal, J.V., Boardman, J.P., Srinivasan, L., Rutherford, M.A., Dyet, L.E., Edwards, A.D., Rueckert, D., 2008. Assessment of brain growth in early childhood using deformation-based morphometry. *NeuroImage* 39, 348-358.
- Allasonnière, S., Amit, Y., Trouvé, A., 2007. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69, 3-29.
- Andersson, J.L.R., Jenkinson, M., Smith, S., 2007. Non-linear registration aka Spatial normalisation. FMRIB Centre, Oxford.
- Ashburner, J., Ridgway, G.R., 2013. Symmetric diffeomorphic modelling of longitudinal structural MRI. *Frontiers in neuroscience* 6.
- Avants, B., Gee, J.C., 2004. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage* 23, Supplement 1, S139-S150.
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis* 12, 26-41.
- Beg, M.F., Khan, A., 2007. Symmetric data attachment terms for large deformation image registration. *Medical Imaging, IEEE Transactions on* 26, 1179-1189.
- Bhatia, K.K., Hajnal, J.V., Puri, B.K., Edwards, A.D., Rueckert, D., 2004. Consistent groupwise non-rigid registration for atlas construction. *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pp. 908-911 Vol. 901.
- Bouix, S., Rathi, Y., Sabuncu, M., 2010. Building an Average Population HARDI Atlas. MICCAI Workshop on CD-MRI.
- Cachier, P., Rey, D., 2000. Symmetrization of the non-rigid registration problem using inversion-invariant energies: Application to multiple sclerosis. In: Delp, S., DiGoia, A., Jaramaz, B. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. Springer Berlin / Heidelberg, pp. 697-708.

Castadot, P., Lee, J.A., Parraga, A., Geets, X., Macq, B., Grégoire, V., 2008. Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors. *Radiotherapy and Oncology* 89, 1-12.

Chen, Y., Ye, X., 2010. Inverse Consistent Deformable Image Registration. In: Alladi, K., Klauder, J.R., Rao, C.R. (Eds.), *The Legacy of Alladi Ramakrishnan in the Mathematical Sciences*. Springer New York, pp. 419-440.

Christensen, G.E., Johnson, H.J., 2001. Consistent image registration. *Medical Imaging, IEEE Transactions on* 20, 568-582.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage* 56, 766-781.

Fischl, B., 2012. FreeSurfer. *NeuroImage* 62, 774-781.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der, K.A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341-355.

Fischl, B., Stevens, A.A., Rajendran, N., Yeo, B.T.T., Greve, D.N., Van Leemput, K., Polimeni, J.R., Kakunoori, S., Buckner, R.L., Pacheco, J., Salat, D.H., Melcher, J., Frosch, M.P., Hyman, B.T., Grant, P.E., Rosen, B.R., van der Kouwe, A.J.W., Wiggins, G.C., Wald, L.L., Augustinack, J.C., 2009. Predicting the location of entorhinal cortex from MRI. *NeuroImage* 47, 8-17.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., 2011. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* 54, 313-327.

Fox, N.C., Ridgway, G.R., Schott, J.M., 2011. Algorithms, atrophy and Alzheimer's disease: Cautionary tales for clinical trials. *NeuroImage* 57, 15-18.

Geng, X., Christensen, G.E., Gu, H., Ross, T.J., Yang, Y., 2009. Implicit reference-based group-wise image registration and its application to structural and functional MRI. *NeuroImage* 47, 1341-1351.

Grenander, U., Miller, M.I., 1998. Computational anatomy: an emerging discipline. *Q. Appl. Math.* LVI, 617-694.

Guimond, A., Meunier, J., Thirion, J.-P., 2000. Average Brain Models: A Convergence Study. *Computer Vision and Image Understanding* 77, 192-210.

Hart, G.L., Zach, C., Niethammer, M., 2009. An optimal control approach for deformable registration. *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pp. 9-16.

Hua, X., Gutman, B., Boyle, C.P., Rajagopalan, P., Leow, A.D., Yanovsky, I., Kumar, A.R., Toga, A.W., Jack Jr, C.R., Schuff, N., Alexander, G.E., Chen, K., Reiman, E.M., Weiner, M.W., Thompson, P.M., 2011. Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. *NeuroImage* 57, 5-14.

- Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, Supplement 1, S151-S160.
- Kuczma, M., Choczewski, B., Ger, R., 1990. *Iterative Functional Equations*. Cambridge University Press.
- Leow, A.D., Yanovsky, I., Ming-Chang, C., Lee, A.D., Klunder, A.D., Lu, A., Becker, J.T., Davis, S.W., Toga, A.W., Thompson, P.M., 2007. Statistical Properties of Jacobian Maps and the Realization of Unbiased Large-Deformation Nonlinear Image Registration. *Medical Imaging, IEEE Transactions on* 26, 822-832.
- Lorenzen, P., Davis, B., Joshi, S., 2004. Model based symmetric information theoretic large deformation multi-modal image registration. *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pp. 720-723 Vol. 721.
- Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., Joshi, S., 2006. Multi-modal image set registration and atlas formation. *Medical image analysis* 10, 440-451.
- Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X., 2013. LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage* 81, 470-483.
- Ma, J., Miller, M.I., Trouné, A., Younes, L., 2008. Bayesian template estimation in computational anatomy. *NeuroImage* 42, 252-261.
- Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* 19, 1498-1507.
- Miller, M., Banerjee, A., Christensen, G., Joshi, S., Khaneja, N., Grenander, U., Matejic, L., 1997. Statistical methods in computational anatomy. *Statistical Methods in Medical Research* 6, 267-299.
- Noblet, V., Heinrich, C., Heitz, F., Arnsbach, J.-P., 2008. Symmetric nonrigid image registration: Application to average brain templates construction. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. Springer Berlin / Heidelberg, pp. 897-904.
- Noblet, V., Heinrich, C., Heitz, F., Arnsbach, J.-P., 2012. An efficient incremental strategy for constrained groupwise registration based on symmetric pairwise registration. *Pattern Recognition Letters* 33, 283-290.
- Reuter, M., Gerstner, E., Rapalino, O., Batchelor, T., Rosen, B., Fischl, B., 2014. Impact of MRI head placement on glioma response assessment. *Journal of Neuro-Oncology* 118, 123-129.
- Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: A robust approach. *NeuroImage* 53, 1181-1196.
- Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61, 1402-1418.
- Rey, D., Subsol, G., Delingette, H., Ayache, N., 2002. Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. *Medical image analysis* 6, 163-179.

- Sabuncu, M., Yeo, B., Van Leemput, K., Vercauteren, T., Golland, P., 2009. Asymmetric image-template registration. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*. Springer Berlin / Heidelberg, pp. 565-573.
- Škrinjar, O., Bistoquet, A., Tagare, H., 2008. Symmetric and Transitive Registration of Image Sequences. *International Journal of Biomedical Imaging*.
- Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable Medical Image Registration: A Survey. *Medical Imaging, IEEE Transactions on* 32, 1153-1190.
- Studholme, C., Cardenas, V., 2004. A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recognition Letters* 25, 1191-1202.
- Tagare, H., Groisser, D., Škrinjar, O., 2009. Symmetric non-rigid registration: A geometric theory and some numerical techniques. *Journal of Mathematical Imaging and Vision* 34, 61-88.
- Thompson, W.K., Holland, D., 2011. Bias in tensor based morphometry Stat-ROI measures may result in unrealistic power estimates. *NeuroImage* 57, 1-4.
- Trouvé, A., Younes, L., 2000. Diffeomorphic matching problems in one dimension: Designing and minimizing matching functionals. *Computer Vision - ECCV 2000*. Springer Berlin / Heidelberg, pp. 573-587.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* 45, S61-S72.
- Yang, D., Li, H., Low, D.A., Deasy, J.O., Naqa, I.E., 2008. A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Physics in medicine and biology* 53, 6143.
- Ye, X., Chen, Y., 2009. A New Algorithm for Inverse Consistent Image Registration. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Wang, J.-X., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnação, M., Silva, C., Coming, D. (Eds.), *Advances in Visual Computing*. Springer Berlin / Heidelberg, pp. 855-864.
- Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C., 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.