J Math Imaging Vis (2009) 34: 61–88
DOI 10.1007/s10851-008-0129-7

Symmetric Non-rigid Registration: A Geometric Theory and Some Numerical Techniques

Hemant D. Tagare · David Groisser · Oskar Skrinjar

Published online: 3 February 2009

Abstract This paper proposes $\mathcal{L}_2$- and information-theory-based (IT) non-rigid registration algorithms that are *exactly symmetric*. Such algorithms pair the same points of two images after the images are swapped. Many commonly-used $\mathcal{L}_2$ and IT non-rigid registration algorithms are only approximately symmetric. The asymmetry is due to the objective function as well as due to the numerical techniques used in discretizing and minimizing the objective function. This paper analyzes and provides techniques to eliminate both sources of asymmetry.

This paper has five parts. The first part shows that objective function asymmetry is due to the use of standard differential volume forms on the domain of the images. The second part proposes alternate volume forms that completely eliminate objective function asymmetry. These forms, called *graph-based* volume forms, are naturally defined on the graph of the registration diffeomorphism f, rather than on the domain of f. When pulled back to the domain of f they involve the Jacobian J_f and therefore appear "non-standard". In the third part of the paper, graph-based volume forms are analyzed in terms of four key objective-function properties: symmetry, positive-definiteness, invariance, and lack of bias. Graph-based volume forms whose associated $\mathcal{L}_2$ objective functions have the first three properties are completely classified. There is an infinite-dimensional space of such graph-based forms. But within this space, up to scalar multiple, there is a unique volume form whose associated $\mathcal{L}_2$ objective function is unbiased. This volume form, which when pulled back to the domain of f is $(1+\det(J_f))$ times the standard volume form on Euclidean space, is exactly the differential-geometrically *natural* volume form on the graph of f. The fourth part of the paper shows how the same volume form also makes the IT objective functions symmetric, positive semi-definite, invariant, and unbiased. The fifth part of the paper introduces a method for removing asymmetry in numerical computations and presents results of numerical experiments. The new objective functions and numerical method are tested on a coronal slice of a 3-D MRI brain image. Numerical experiments show that, even in the presence of noise, the new volume form and numerical techniques reduces asymmetry practically down to machine precision without compromising registration accuracy.

Keywords Non-rigid registration · Symmetric registration · Inverse consistent registration · Image registration

H.D. Tagare (✉)
Dept. of Diagnostic Radiology, Dept. of Biomedical Engineering, Yale University, New Haven, CT 06520, USA
e-mail: hemant.tagare@yale.edu

D. Groisser
Dept. of Mathematics, University of Florida, Gainesville, FL, USA

O. Skrinjar
Georgia Institute of Technology, Atlanta, GA, USA

1 Introduction

Registration algorithms that produce the inverse warp when their input images are swapped are *symmetric* or *inverse-consistent*. Symmetric registration is important in many applications. For example, suppose that a registration algorithm warps one functional MRI image to fit another, after which their activation maps are compared voxel-wise. If after swapping the two images the registration algorithm does

not produce the inverse warp, then the result of the voxel-wise comparison may well depend (spuriously) on which image is being warped.

Asymmetric registration algorithms can be made symmetric by symmetrizing their objective functions. In this paper, we adopt a differential-geometric point of view and show how to symmetrize popular $\mathcal{L}_2$ and information-theoretic (IT) registration objective functions while preserving the following properties of the objective functions: (1) positive (semi-) definiteness, (2) invariance under volume-preserving transformations, and (3) lack of bias (these terms are defined in Sect. 3).

We begin by showing that asymmetry is ultimately due to the use of standard Euclidean volume forms. To symmetrize the objective functions, we replace the standard forms with new forms which are defined on the graph of the warping function. As it happens, there are infinitely many graph-based forms that make the $\mathcal{L}_2$ objective function symmetric, positive-definite, and invariant under volume preserving transformations. In the first theorem of this paper, we give a complete classification of such forms. In the second theorem we show that, up to multiplication by a positive scalar, there is a unique graph-based form that additionally makes the symmetrized $\mathcal{L}_2$ objective function unbiased. When pulled back to the domain of f, this form is simply $(1 + \det J_f)$ times the standard Euclidean form. Remarkably, it turns out that same volume form also symmetrizes IT objective functions.

A pleasing property of this theory is that the symmetrized objective functions retain the same nature as the original objective functions. Thus, the symmetrized $\mathcal{L}_2$ objective function remains an $\mathcal{L}_2$ objective function, and the symmetrized IT objective function remains an IT objective function.

Objective functions are not the only source of asymmetry in registration algorithms. The numerical technique used to discretize and minimize the objective function can also introduce asymmetry. To overcome numerical asymmetry, in Sect. 11 we present a symmetric numerical scheme. Numerical experimentation shows that the theory and the numerical techniques work well in practice—the graph-based volume forms reduce asymmetry in $\mathcal{L}_2$ and IT objective functions down to near machine precision.

As mentioned above, we symmetrize the objective functions while preserving three other properties. Why do we insist on these additional properties? The answer is quite simple. Symmetrizing an objective function changes it, and it is important to guarantee that the change does not destroy other useful properties of the objective function. The three additional properties assure us that the modified objective function remains minimizable (positive definiteness), that it does not lose its ability to track certain simple transformations (invariance under volume preserving transformations), and the new graph-based volume forms do not prefer one warp over another when registering constant images (lack of bias). Without these properties, the symmetrized objective function would not be completely trustworthy—it might either give no answer, or give an answer that changes if the object in the image translates or rotates, or give a biased answer. In our opinion, symmetry should be viewed as a part of a larger set of desirable properties of the objective function.

1.1 Previous Work

The literature on image registration is vast. Excellent reviews and surveys are available in books [1–3] and published papers [4–6].

Previous work on non-rigid registration can be classified into two classes. In the first class are methods which directly estimate the warping function and its inverse. In the second class are methods which estimate a pair of functions whose composition yields the warping function. We briefly survey the methods here, but postpone a more mathematical discussion of them to Sect. 7.1.

1.1.1 Direct Methods

Symmetric registration with direct methods was studied by Christensen and colleagues [7–9] who called it *inverse-consistent registration* or simply *consistent registration*. Christensen et al. impart symmetry by taking the asymmetric objective function with a warp in the forward direction and adding to it the same objective function with a warp in the backward direction. The forward and backward warps are constrained to be approximate inverses of one another. The constraint is enforced in a "soft" manner by adding penalty terms to the objective function. We will refer to this approach as the *penalized-asymmetry* approach.

The penalized-asymmetry approach has some limitations. First, it is only approximately symmetric. Further, the penalized-asymmetry approach can be viewed as a constrained minimization approach with a large number of (theoretically infinite) pointwise equality constraints on the forward and backward warps. Because the constraints are imposed by introducing penalty terms in the objective function, the minimum found by the algorithm is a compromise between minimizing the unpenalized objective function and satisfying the constraints. Finally, adding the forward and backward objective functions may not preserve the meaning of the objective functions. For example, the sum of forward and backward mutual informations is not another mutual information.

In a slightly different approach, Ashburner et al. [10] use symmetric priors, with the final registration being approximately symmetric. Rogelj and Kovacic [13] propose a method which uses symmetrically designed forces to deform both images. This method requires maintaining forward and

backward transformations, but it does not use their inverses. This method is also approximately symmetric.

Cachier and Rey [12] propose using the $(1+\det J_f)$ term, but no geometric justification is provided for it, nor is its uniqueness established.

1.1.2 Composition Methods

As mentioned above, composition methods do not directly estimate the warping function or its inverse, but instead estimate functions whose compositions give the warp and its inverse. As we discuss in Sect. 7.1, how these methods relate to direct methods, and whether they have a natural geometry is not entirely clear. These methods are reported by Joshi at al. [14, 15, 19], Gee et. al [16], and Beg and Khan [11].

1.1.3 Other Work

Related to the problem of non-rigid registration is the problem of non-rigid correspondences between image structures. In [17, 18] a theory of symmetric non-rigid correspondences was developed by two of the authors of this paper. The idea of formulating the problem on the graph of a function dates to these publications.

1.1.4 Relation to Previous Work

Most of the previous work is algorithmic in nature, whereas the main contribution of our work is theoretical. Our work provides geometric insight into the origin of asymmetry and a way to convert asymmetric registration into symmetric registration. There are several key ideas that emerge from this analysis which, as far as we know, have not been mentioned before. For example, the fundamental role of the volume differential form in registration has not been alluded to or clarified. Further, the existence, classification, and uniqueness of symmetrizing volume forms is new. Finally, our theory shows how IT registration can be symmetrized in the same framework as $\mathcal{L}_2$ registration. To our knowledge, this has not been reported before; the algorithms mentioned above consider $\mathcal{L}_2$ registration only.

1.1.5 Miscellaneous Comments

We are often asked about practical applications of symmetric registration. In practice, symmetric registration has been used in atlas creation [14, 20], synthesizing average shapes of organs [21, 22], and processing of lung [23, 24] and ear images [25].

Finally, a comment about the term "symmetric registration". In mathematics, there is a standard term for the situation where the value of a function with two arguments stays the same when the arguments are swapped. Such a function is said to be *symmetric* with respect to its arguments, and the property is referred to as *symmetry*. Common examples of this are the symmetry of the metric $(d(x, y) = d(y, x))$, and the symmetry of an equivalence relation $((y \sim x)$ if $(x \sim y))$. Because we want the registration objective function value to stay the same when the input images are swapped (and the warp is replaced with the inverse warp), "symmetric registration" seems more appropriate to us than "inverse-consistent registration".

1.2 Organization of the Paper

We begin the paper in Sect. 2 with some background information and notation. Section 3 gives mathematical definitions of positive-definiteness, symmetry, bias, and invariance under volume-preserving transformations. Section 4 investigates the asymmetry of the $\mathcal{L}_2$ objective function.

Section 5 introduces the new space in which symmetrizing, non-standard forms arise in a natural way. There are many symmetrizing forms in this space, and Theorem 1 in Sect. 6 gives a complete classification of these forms. We then proceed to find the unique form that gives us the additional properties that we want. The main result is Theorem 2 of Sect. 7, which identifies this form. The resulting symmetric $\mathcal{L}_2$ objective function is also defined in this section.

In Sect. 8 we turn our attention to IT objective functions and show that they are asymmetric as well, and for exactly the same reason as $\mathcal{L}_2$. In Sect. 9, the IT objective functions are symmetrized using the volume form identified in Theorem 3 of Sect. 7.

Section 10 discusses regularization. Section 11 addresses the issues of numerical symmetry and proposes a symmetric numerical scheme. Section 12 contains simulations that evaluate the asymmetry and show that the new symmetric objective functions reduce asymmetry practically down to machine precision. The experiments also show that there is no loss of accuracy in registration when asymmetric objective functions are replaced with symmetric ones. In fact, a slight increase in accuracy is observed.

2 Images, Diffeomorphisms and Objective Functions

We begin with some definitions. In order to simplify the exposition, we assume certain degrees of differentiability for images and diffeomorphisms, but some of our results hold with weaker differentiability than the C^1 images and C^2 diffeomorphisms that our definitions restrict us to.

2.1 Images and their Ordered Pairs

The *domain* of an image is Ω, the closed unit cube of $\mathbb{R}^n$. In practice, we are interested in $n = 2$ or $n = 3$, but our theory

holds for any finite n. An *image* $I : \Omega \to \mathbb{R}$ is a continuously differentiable function from Ω to the real line. A *constant image* on Ω is a constant function from Ω to $\mathbb{R}$. Constant images are important in the study of bias.

Let $\Omega_1 = \Omega_2 = \Omega$ be two copies of Ω. An *ordered pair of images* (I_1, I_2) is a pair of images (functions) with the first defined on Ω_1 and the second on Ω_2. The ordered pair (I_2, I_1) is the ordered pair (I_1, I_2) swapped. In (I_2, I_1), the image I_2 is defined on Ω_1 and I_1 on Ω_2.

2.2 Diffeomorphisms

During registration, images are warped by diffeomorphisms from Ω to Ω. We assume that the diffeomorphisms are C^2. We denote the Jacobian matrix of a diffeomorphism $f : \Omega \to \Omega$ by J_f. A diffeomorphism $f : \Omega \to \Omega$ is *orientation-preserving* if $\det J_f(x) > 0$ for all $x \in \Omega$. The set of orientation preserving diffeomorphisms forms a group under composition. We denote this group by Diff^+.

If a diffeomorphism $f \in \mathrm{Diff}^+$ has $\det J_f(x) = 1$ for all $x \in \Omega$, then it is a *volume- and orientation-preserving* diffeomorphism. The set of all volume and orientation-preserving diffeomorphisms forms a subgroup of Diff^+, denoted by SDiff^+.

Since we are only interested in Diff^+ and SDiff^+ in the following, we drop the qualification "orientation-preserving." We simply refer to elements of Diff^+ as a diffeomorphisms and elements of SDiff^+ as volume-preserving diffeomorphisms.

There is a subtle but important point in the use of Diff^+. As we have defined it, the diffeomorphisms are bijective on Ω. Strictly speaking, this excludes translations, rotations, and scaling from Diff^+ because they move part of the image outside Ω. This is not as restrictive as it seems. In practice, the two images are usually translated, rotated, and scaled before non-rigid registration is applied. After translation etc., a window Ω is chosen on the region of overlap of the two images, and non-rigid registration is applied as a bijection on Ω. We are restricting ourselves to this situation, hence it is sufficient to define Diff^+ as we have done.

If $f, g \in \mathrm{Diff}^+$, then $f \circ g \in \mathrm{Diff}^+$ and we write the determinant of the Jacobian of $f \circ g$ as $\det J_f \circ g \ \times \ \det J_g$ or simply as $\det J_f \circ g \ \det J_g$. This may seem a little odd at first, but is the mathematically correct notation—it means that the determinant is a product of two functions, the first of which is $\det J_f \circ g$ (the value of $\det J_f \circ g$ at a point $x \in \mathrm{Domain}(g)$ is the value of $\det J_f$ evaluated at $g(x) \in \mathrm{Domain}(f)$) times $\det J_g$.

2.3 Objective Functions

For a pair of images (I_1, I_2), a registration objective function $\mathcal{J}((I_1, I_2), f)$ measures how well the two images match after warping by the diffeomorphism $f \in \mathrm{Diff}^+$. Lower values of $\mathcal{J}$ indicate a better match. Registration algorithms seek an f that minimizes a given $\mathcal{J}$.

Registration objective functions are usually a weighted sum of an image-dependent term and a regularization term. The image-dependent term drives the warp in a direction that registers the image. The regularization term, which depends only on the warp (it is independent of the image), biases the solution towards smoother warps. In this paper, we are primarily concerned with the image-dependent term in the objective function, and unless we explicitly mention it, from now on we will take the term "objective function" to mean this term only. We do have suggestions for regularization terms that can be used along with the symmetrized objective functions. They are presented in Sect. 10.

3 Properties of Objective Functions

We would like registration objective functions to have the following four properties. The first property assures us that the objective function is well defined and that its minimum is meaningful:

1. Positive-definiteness: The objective function $\mathcal{J}((I_1, I_2), f)$ is *positive-definite* if for all ordered images (I_1, I_2) and all $f \in \mathrm{Diff}^+$

$$\mathcal{J}((I_1, I_2), f) \geq 0,$$

with equality if and only if $I_1 = I_2 \circ f$.

The next property assures us that there is no built-in bias in the registration algorithm. By this we mean that if the two input images are constant images (they have no features) then the registration algorithm should not prefer any diffeomorphism over another:

2. Lack of Bias: An objective function is *unbiased* if $\mathcal{J}((I_1, I_2), f_1) = \mathcal{J}((I_1, I_2), f_2)$ for all pairs (I_1, I_2) of constant images and all $f_1, f_2 \in \mathrm{Diff}^+$.

The third property is symmetry:

3. Symmetry: The objective function $\mathcal{J}((I_1, I_2), f)$ is *symmetric* if

$$\mathcal{J}((I_1, I_2), f) = \mathcal{J}((I_2, I_1), f^{-1}),$$

for all ordered images (I_1, I_2) and all $f \in \mathrm{Diff}^+$.

The final property we want is that the objective function track changes in the images. That is, if we begin with a pair of input images and systematically distort them, then we would like the registration algorithm to track the distortions. This idea is made precise using the following commutative

diagram:

$$\begin{array}{ccc} \mathbb{R} & & \mathbb{R} \\ \uparrow I_1 & & \uparrow I_2 \\ \Omega_1 & \overrightarrow{f} & \Omega_2 \\ \uparrow g_1 & & \uparrow g_2 \\ \Omega_1 & \overrightarrow{g_2^{-1} \circ f \circ g_1} & \Omega_2 \end{array} \tag{1}$$

The top part of the diagram shows images I_1 and I_2 on Ω_1 and Ω_2. The bottom part of the diagram has images $I_1 \circ g_1$ and $I_2 \circ g_2$ on Ω_1 and Ω_2 respectively. These images are simply the images from the top part distorted by g_1 and g_2 respectively. Any diffeomorphism f in the top part of the diagram has a corresponding diffeomorphism ${g_2}^{-1} \circ f \circ g_1$ in the bottom part which continues to pair the same set of points in the two images.

One way of assuring that the objective function does not lose sensitivity to changes in the images is to require the values $\mathcal{J}((I_1, I_2), f)$ and $\mathcal{J}((I_1 \circ g_1, I_2 \circ g_2), {g_2}^{-1} \circ f \circ g_1)$ to be equal for all f and for as large a set of g_1 and g_2 as possible. The largest possible set of g_1 and g_2 that we have been able to accommodate is $g_1, g_2 \in \text{SDiff}^+$. This gives us the definition:

4. Invariance under SDiff^+ (invariance under volume-preserving transformations): An objective function is invariant under SDiff^+ if

$$\mathcal{J}((I_1, I_2), f) = \mathcal{J}((I_1 \circ g_1, I_2 \circ g_2), {g_2}^{-1} \circ f \circ g_1)$$

for all (I_1, I_2), all $g_1, g_2 \in \text{SDiff}^+$, and all $f \in \text{Diff}^+$.

Our goal is to modify $\mathcal{L}_2$ and IT registration objective functions so that they satisfy all of the above properties. We begin by analyzing the $\mathcal{L}_2$ objective function.

4 The $\mathcal{L}_2$ Objective Function and Its Asymmetry

The $\mathcal{L}_2$ objective function is customarily defined as

$$\sqrt{\int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega},$$

where ω is the standard volume differential form, $\omega = dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$ (we use the formalism of differential forms in our analysis and the reader is reminded of some basic facts about them in the footnote[1]).

Squaring the objective function does not change the minimizing f, and it is common to use

$$\mathcal{J}((I_1, I_2), f) = \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega. \tag{2}$$

instead. We will refer to it as the $\mathcal{L}_2$ objective function.

The $\mathcal{L}_2$ objective function is asymmetric. To see this, consider a diffeomorphism $f^{-1} : \Omega_1 \to \Omega_1$ shown below from the lower copy of Ω_1 to the upper copy of Ω_1:

$$\begin{array}{ccc} \mathbb{R} & & \mathbb{R} \\ \uparrow I_1 & & \uparrow I_2 \\ \Omega_1 & \overrightarrow{f} & \Omega_2 \\ \uparrow f^{-1} & & \\ \Omega_1 & \overrightarrow{f^{-1}} & \Omega_2 \\ \downarrow I_2 & & \downarrow I_1 \\ \mathbb{R} & & \mathbb{R} \end{array} \tag{3}$$

Then, starting with the $\mathcal{L}_2$ objective function defined on the upper copy of Ω_1 and pulling it to the lower copy of Ω_1 via f^{-1} gives

$$\begin{aligned} \mathcal{J}((I_1, I_2), f) &= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega \\ &= \int_{\Omega_1} (I_1 \circ f^{-1} - I_2)^2 \, \left({f^{-1}}^* \omega \right), \end{aligned} \tag{4}$$

where, ${f^{-1}}^* \omega$ is the pull back of ω through f^{-1} evaluated as ${f^{-1}}^* \omega = \det J_{f^{-1}} \omega$.

Because $\mathcal{J}((I_2, I_1), f^{-1}) = \int_{\Omega_1} (I_2 - I_1 \circ f^{-1})^2 \, \omega$, we see from (4) that $\mathcal{J}((I_1, I_2), f) = \mathcal{J}((I_2, I_1), f^{-1})$ for all images if and only if $\det J_{f^{-1}} \, \omega = \omega$, for all $u \in \Omega_1$ i.e. if and only if $\det J_{f^{-1}} = 1$ for all $u \in \Omega_1$. Since this does not hold for all $f \in \text{Diff}^+$, the $\mathcal{L}_2$ objective function is asymmetric. This argument shows clearly that it is the $\det J_{f^{-1}}$ term that causes the asymmetry. Thus the asymmetry arises directly from the use of the standard volume form ω.

[1] A *differential form of degree k*, or *k-form*, on a manifold is a field of totally antisymmetric, contravariant k-tensors. We denote by α_x the value of a differential form α at the point x, and call α_x a *k-form at the point x*. At each point x of an n-dimensional manifold, the set of n-forms at x is a one-dimensional vector space. If the manifold is oriented, then the notions of "positive" and "negative" elements of this one-dimensional space are well-defined. A *volume form* on an oriented n-dimensional manifold is an n-form that is positive at each point.

If the manifold is an open subset U of R^n, then the *standard volume form* on U is $\omega := dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$; we will reserve the symbol ω exclusively for this form. The (standard) volume of any open subset O of U is $\int_O \omega$. Note that every volume form on U is of the form $h\omega$ for some function $h : U \to R^+$. All these facts extend to n-dimensional manifolds with boundary or manifolds with corners, such as Ω, in R^n.

The pull-back of an n-form α by a differentiable map f is $f^*\alpha = (\det J_f)\alpha$, where J_f is the Jacobian of f.

4.1 Symmetrizing the $\mathcal{L}_2$ Objective Function

To symmetrize the $\mathcal{L}_2$ objective function, we could replace the standard volume form ω with an alternate form of the type $\sigma(J_f)\omega$ (the function σ to be determined). Such a form can symmetrize the $\mathcal{L}_2$ objective function provided the $\sigma(J_f)$ term can compensate the $\det J_{f^{-1}}$ term that causes asymmetry.

Below, we introduce a geometric framework for registration in which the volume forms $\sigma(J_f)\omega$ arise naturally and in which $\int \sigma(J_f)\omega$ has a consistent geometric interpretation. In this framework $\int \sigma(J_f)\omega$ is interpreted as the volume of a set contained in the graph of the function f, with the graph $\mathcal{G}_f = \{(u, f(u)) \mid u \in \Omega_1\}$ viewed as a differentiable manifold in the product space $\Omega_1 \times \Omega_2$. This geometric point of view is fundamental to our approach to symmetrization.

Using the above framework we show that up to a constant multiple there is a single form, namely

$$(1 + \det J_f)\, \omega, \tag{5}$$

that gives the properties we ask for. Using this form, the $\mathcal{L}_2$ objective function is modified to

$$\mathcal{J}((I_1, I_2), f) = \int_{\Omega_1} (I_1 - I_2 \circ f)^2 (1 + \det J_f)\, \omega. \tag{6}$$

The modified $\mathcal{L}_2$ objective function is positive-definite, symmetric, invariant under volume-preserving transformations and unbiased. Further, as we show in Sect. 9, this form also makes the IT objective functions positive-definite, symmetric, invariant under volume preserving transformations and unbiased.

As will become clear in the exposition below, the volume form of (5) has a very natural graph-based interpretation: it is the sum of pullbacks to $\mathcal{G}_f$ of the standard volume forms on Ω_1 and Ω_2.

5 Product Spaces and Graphs

5.1 Product Spaces and Projections

The product space $\Omega_1 \times \Omega_2$ is the set of all ordered pairs $\{(u, v) \mid u \in \Omega_1, v \in \Omega_2\}$ and has two natural projection functions $\pi_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ and $\pi_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ defined by $\pi_1((u, v)) = u$ and $\pi_2((u, v)) = v$. The projection functions are continuously differentiable.

5.2 Graphs

The graph of any differentiable function $f : \Omega_1 \to \Omega_2$ is the set $\mathcal{G}_f = \{(u, f(u)) \mid u \in \Omega_1\}$ and is a $\dim(\Omega_1)$-dimensional, differentiable, and orientable submanifold of $\Omega_1 \times \Omega_2$. If $f \in \mathrm{Diff}^+$, then an orientation can be easily chosen on $\mathcal{G}_f$ such that the projection functions, restricted to $\mathcal{G}_f$, become orientation-preserving diffeomorphisms from $\mathcal{G}_f$ to Ω_1 and Ω_2. We denote the projection functions restricted to $\mathcal{G}_f$ by $p_1 = \pi_1 |_{\mathcal{G}_f}$ and $p_2 = \pi_2 |_{\mathcal{G}_f}$ respectively.

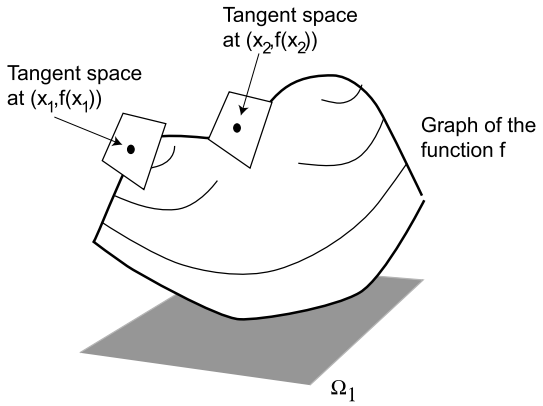


Fig. 1 Parallel tangent planes

The functions p_1, p_2 are related to f by $p_2 \circ {p_1}^{-1} = f$ and $p_1 \circ {p_2}^{-1} = f^{-1}$. We use these relations below.

5.3 Volume Forms on the Graph

If ω is the standard volume form on Ω_1, then $p_1^*\omega$ (the pullback of ω to $\mathcal{G}_f$) is a volume form on $\mathcal{G}_f$. Multiplying $p_1^*\omega$ with any function on $\mathcal{G}_f$ gives another volume form on $\mathcal{G}_f$. All volume forms on $\mathcal{G}_f$ can be obtained this way.

Recall that we want to consider volume forms that depend on J_f. Geometrically speaking, these volume forms depend on the inclination (or the "attitude") of the tangent space of $\mathcal{G}_f$ at $(x, f(x))$. In Fig. 1 the graph $\mathcal{G}_f$ is illustrated as a surface over Ω_1. In the figure, the tangent spaces to $\mathcal{G}_f$ at $(x_1, f(x_1))$ and at $(x_2, f(x_2))$ are parallel, and we want the volume forms at $(x_1, f(x_1))$ and $(x_2, f(x_2))$ to be an "equal" multiple of p_1^*w. To say this more precisely, partition $\mathcal{G}_f$ into equivalence classes where points belong to the same class if and only if they have parallel tangent spaces. We want to consider volume forms that have the same value for all points in one equivalence class. But the value can differ across equivalence classes.

The inclination of the tangent space at $(x, f(x))$ is completely determined by the Jacobian $J_f(x)$. Let $\mathbb{R}^{n\times n}$ be the space of $n \times n$ matrices. Denoting $(0, \infty)$ by $\mathbb{R}^+$, we have $M^+ = \det^{-1}(\mathbb{R}^+)$ as the set of $n \times n$ matrices with positive determinant. For each $x \in \Omega_1$, we may view "Jacobian" as a map given by $(f, x) \mapsto J_f(x)$. For each $x \in \Omega_1$ the map $f \mapsto J_f(x)$ is onto. Suppose that $\sigma : M^+ \to \mathbb{R}$ is a real-valued function on M^+. Then, σ induces a function on $\mathcal{G}_f$ given by $h = \sigma \circ J_f$, and we can construct the volume form $\alpha_x = h(x)p_1^*\omega_x = \sigma(J_f(x))p_1^*\omega_x$ whose value is the same for all points $(x, f(x))$ in the same equivalence class.

Definition Every positive differentiable function $\sigma : M^+ \to \mathbb{R}$ defines a differentiable *tangent-dependent volume form* at $(x, f(x))$ according to

$$\alpha_{(x,f(x))} = \sigma(J_f(x))\, p_1^* \omega_x. \tag{7}$$

5.4 The $\mathcal{L}_2$ Objective Function on the Graph of a Diffeomorphism

Given the pair (I_1, I_2), define $I_1 \ominus I_2 : \Omega_1 \times \Omega_2 \to \mathbb{R}$ by $I_1 \ominus I_2 = I_1 \circ \pi_1 - I_2 \circ \pi_2$. If $I_1 \ominus I_2$ is restricted to $\mathcal{G}_f$ then the integral of the square of $I_1 \circ p_1 - I_2 \circ p_2$ on $\mathcal{G}_f$ gives a graph-based $\mathcal{L}_2$ objective function.

Definition Let $\sigma : M^+ \to \mathbb{R}$ be as above. The *graph-based $\mathcal{L}_2$ ($\mathcal{GL}_2$) objective function determined by σ* is

$$\mathcal{J}((I_1, I_2), f) = \int_{\mathcal{G}_f} (I_1 \ominus I_2)^2 \alpha \tag{8}$$

where α is a tangent-dependent volume on $\mathcal{G}_f$ given by (7).

It is instructive to change the domain of the integral from $\mathcal{G}_f$ to Ω_1.

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), f) \\
&\quad = \int_{\mathcal{G}_f} (I_1 \ominus I_2)^2 \alpha \\
&\quad = \int_{\mathcal{G}_f} (I_1 \circ p_1(w) - I_2 \circ p_2(w))^2 \sigma(J_f \circ p_1(w))\, p_1^* \omega_{p_1(w)} \\
&\quad = \int_{\Omega_1} (I_1(x) - I_2 \circ p_2 \circ {p_1}^{-1}(x))^2 \sigma(J_f(x)) \omega_x \\
&\quad = \int_{\Omega_1} (I_1 - I_2 \circ f)^2(x) \sigma(J_f(x)) \omega_x,
\end{aligned} \tag{9}$$

where $x = p_1(w)$.

This calculation shows that the $\mathcal{GL}_2$ is simply the $\mathcal{L}_2$ objective function with the standard volume form ω replaced by the volume form $\sigma(J_f)\omega$. As discussed in Sect. 4.1 this is exactly what we want.

6 The Volume Form

Having defined a class of volume forms and an objective function based on them, the next task is to identify a volume form that makes the $\mathcal{GL}_2$ objective function positive-definite, unbiased, symmetric and invariant under SDiff^+. We do this in two steps:

1. First, we translate the requirements of positive-definiteness, symmetry and invariance under SDiff^+ to requirements on the function σ. Using this, we identify all volume forms that give positive-definiteness, symmetry and invariance under SDiff^+.
2. From these volume forms we pick the forms for which the objective function has no bias. We shall see that, up to a scalar multiple, there is a unique such form.

6.1 Positive-Definiteness

Proposition 1 *The $\mathcal{GL}_2$ objective function determined by σ is positive-definite if and only if $\sigma(J_f) > 0$ for all $J_f \in M^+$.*

Proof If $\sigma(J_f(x)) > 0$ then (9) shows that $\mathcal{GL}_2$ is positive-definite.

The proof of the converse is by contradiction. Suppose there is an $A \in M^+$ for which $\sigma(A) \leq 0$ and which gives a positive-definite $\mathcal{GL}_2$. Then construct a diffeomorphism $f : \Omega_1 \to \Omega_2$ such that $J_f(x) = A$ for all points x of an open set $O \subset \Omega_1$. Let $O' \subset O$ be open, and take $\Delta : \Omega_1 \to \mathbb{R}$ to be a continuously differentiable function which is zero on the complement of O, strictly positive on O' and non-negative on $O - O'$ (the existence of such a function follows from a partition of unity argument). Finally, choose I_1 to be any image on Ω_1 and $I_2 = (I_1 + \Delta) \circ f^{-1}$. Then, writing $x = p_1(w)$

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), f) \\
&\quad = \int_{\mathcal{G}_f} (I_1 \ominus I_2)^2 \alpha \\
&\quad = \int_{\Omega_1} (I_1(x) - (I_1 + \Delta) \circ f^{-1} \circ f(x))^2 \sigma(J_f(x)) \omega_x \\
&\quad = \int_O \Delta^2 \sigma(A) \omega.
\end{aligned}$$

If $\sigma(A) = 0$, then $\mathcal{J}((I_1, I_2), f) = 0$. But $I_1 \neq I_2 \circ f$, which contradicts the definition of non-negativity.

If $\sigma(A) < 0$, then $\mathcal{J}((I_1, I_2), f) < 0$, which also contradicts the definition of non-negativity. This proves the result. □

6.2 Symmetry and Invariance Under SDiff^+

To analyze symmetry and invariance under SDiff^+ we need the following simple result:

Proposition 2 *If α and β are two non-zero continuous volume forms on Ω_1 then,*

$$\int_{\Omega_1} h\alpha = \int_{\Omega_1} h\beta, \tag{10}$$

for all continuous (or continuously differentiable) $h : \Omega_1 \to \mathbb{R}$ if and only if $\alpha = \beta$.

Proof The proof is routine and depends on continuity. It is omitted to conserve space. □

We use this result to analyze symmetry.

6.2.1 Symmetry

Given f, for $i = 1, 2$ let $\tilde{p}_i = \pi_i|_{\mathcal{G}_{f^{-1}}}$, and let $\tilde{\alpha}$ denote the $\mathcal{G}L_2$ volume form on $\mathcal{G}_{f^{-1}}$ determined by σ. Then given I_1, I_2 we can use the diffeomorphism $\tilde{p}_1^{-1} : \Omega_1 \to \mathcal{G}_f$ or $\tilde{p}_2^{-1} : \Omega_2 \to \mathcal{G}_f$ to pull back the integral defining $\mathcal{J}((I_2, I_1), f^{-1})$ to an integral over Ω_1 or Ω_2 respectively, i.e. to re-express the integral in terms of the variable $x \in \Omega_1$ (as in the derivation of (9)) or in terms of $y \in \Omega_2$. If we pull back to Ω_2, then, since $\tilde{p}_2^{-1} = \tilde{p}_1^{-1} \circ f$ (note that now f is viewed as a map $\Omega_2 \to \Omega_1$; it is f^{-1} that is viewed as a map $\Omega_1 \to \Omega_2$) we find

$$\begin{aligned}
&\mathcal{J}((I_2, I_1), f^{-1}) \\
&\quad = \int_{\Omega_2} f^* \left((\tilde{p}_1^{-1})^* ((I_2 \ominus I_1)^2 \tilde{\alpha} \right) \\
&\quad = \int_{\Omega_2} f^* \left((I_2 - I_1 \circ f^{-1})^2 \sigma(J_{f^{-1}}) \omega \right) \quad \text{by (6)} \\
&\quad = \int_{\Omega_2} (I_2 \circ f - I_1)^2 (\sigma(J_{f^{-1}} \circ f) f^* \omega \\
&\quad = \int_{\Omega_2} (I_1 - I_2 \circ f)^2 \sigma(J_{f^{-1}} \circ f) \det(J_f) \omega \\
&\quad = \int_{\Omega} (I_1 - I_2 \circ f)^2 \sigma((J_f)^{-1}) \det(J_f) \omega \qquad (11)
\end{aligned}$$

since $\Omega_2 = \Omega$ and $J_{f^{-1}} \circ f = (J_f)^{-1}$. But symmetry requires that $\mathcal{J}((I_2, I_1), f^{-1}) = \mathcal{J}((I_1, I_2), f)$. Since for any f, the images I_1 and I_2 can be chosen such that $I_2 - I_1 \circ f^{-1}$ is any continuously differentiable function, applying Proposition 2 to (11) and (9) implies that $\mathcal{J}$ is symmetric if and only if

$$\sigma(J_f(x)^{-1}) \det(J_f(x)) = \sigma(J_f(x))$$

for all $f \in \text{Diff}^+$ and all $x \in \Omega^1$. Since, varying f and x, the matrix $J_f(x)$ can take all possible values in M^+, we have the result:

Proposition 3 *The $\mathcal{G}L_2$ objective function determined by σ is symmetric if and only if*

$$\sigma(A^{-1}) \det(A) = \sigma(A) \qquad (12)$$

for all $A \in M^+$.

6.2.2 Invariance under SDiff$^+$

Referring back to the commutative diagram (1) on p. 65, we have for the top part:

$$\mathcal{J}((I_1, I_2), f) = \int_{\mathcal{G}_f} (I_1 \ominus I_2)^2 \alpha$$

$$= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \sigma(J_f) \omega.$$

Pulling this integral to the bottom copy of Ω_1 using $g_1 \in \text{SDiff}^+$ gives

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), f) \\
&\quad = \int_{\Omega_1} (I_1 \circ g_1 - I_2 \circ f \circ g_1)^2 \sigma(J_f) \det J_{g_1} \omega \\
&\quad = \int_{\Omega_1} (I_1 \circ g_1 - I_2 \circ f \circ g_1)^2 \sigma(J_f) \omega \\
&\quad = \int_{\Omega_1} (I_1 \circ g_1 - I_2 \circ g_2 \circ {g_2}^{-1} \circ f \circ g_1)^2 \sigma(J_f) \omega,
\end{aligned}$$

since $\det J_{g_1} = 1$, and $g_2 \circ {g_2}^{-1} = id$.

Next,

$$\begin{aligned}
&\mathcal{J}((I_1 \circ g_1, I_2 \circ g_2), {g_2}^{-1} \circ f \circ g_1) \\
&\quad = \int_{\mathcal{G}_{{g_2}^{-1} \circ f \circ g_1}} (I_1 \circ g_1 \ominus I_2 \circ g_2)^2 \alpha \\
&\quad = \int_{\Omega_1} (I_1 \circ g_1 - I_2 \circ g_2 \circ {g_2}^{-1} \circ f \circ g_1)^2 \sigma(J_{{g_2}^{-1} \circ f \circ g_1}) \omega.
\end{aligned}$$

Since we want $\mathcal{J}((I_1, I_2), f) = \mathcal{J}((I_1 \circ g_1, I_2 \circ g_2), {g_2}^{-1} \circ f \circ g_1)$ for all I_1, I_2, f, g_1, g_2, it must be that

$$\sigma(J_f)\omega = \sigma(J_{{g_2}^{-1} \circ f \circ g_1})\omega = \sigma(J_{{g_2}^{-1}} J_f J_{g_1})\omega,$$

or $\sigma(J_f) = \sigma(J_{{g_2}^{-1}} J_f J_{g_1})$.

Since J_f can take any value in M^+ and J_{g_2} and $J_{g_1^{-1}}$ can take any values in $SL(n)$, the above relation implies that

$$\sigma(A) = \sigma(BAC), \qquad (13)$$

must hold for all $A \in M^+$ and all $B, C \in SL(n)$. Thus σ must be constant over the set $L_A = \{BAC \mid B, C \in SL(n)\}$.

Lemma 1 *For any $A \in M^+$, the set $L_A = \{BAC \mid B, C \in SL(n)\}$ is the level set of the determinant function passing through A. This implies that invariance under SDiff$^+$ holds if and only if the function $\sigma : M^+ \to \mathbb{R}^+$ can be written as*

$$\sigma(A) = \zeta(\det A),$$

for some function $\zeta : \mathbb{R}^+ \to \mathbb{R}^+$.

Proof Let $\tilde{L}_A = \{X \mid X \in M^+, \ \det X = \det A\}$. We will show that $\tilde{L}_A = L_A$. If $D \in L_A$, then $D = BAC$, and $\det D = \det A$, since $\det B = \det C = 1$. Thus, $D \in \tilde{L}_A$, showing that $L_A \subset \tilde{L}_A$. Next, let $X \in \tilde{L}_A$. Then $X = (XA^{-1})A$, and the matrix $XA^{-1} \in SL(n)$ since $\det X = \det A$. This shows that $X \in L_A$, thus showing that $L_A = \tilde{L}_A$ and establishing the first part of the lemma. The second part is obvious because σ is constant over any L_A. □

Because $\det : M^+ \to \mathbb{R}^+$ is a continuously differentiable surjective function with a non-vanishing derivative, the function σ is continuously differentiable if and only if ζ is. Thus we have:

Proposition 4 *The $\mathcal{GL}_2$ objective function determined by σ is invariant under SDiff$^+$ if and only if $\sigma : M^+ \to \mathbb{R}$ can be expressed as*

$$\sigma(A) = \zeta(\det A), \tag{14}$$

for some continuously differentiable function $\zeta : \mathbb{R}^+ \to \mathbb{R}^+$.

In view of Proposition 4, we will henceforth refer to the $\mathcal{G}L_2$ objective function determined by σ as "the $\mathcal{G}L_2$ objective function determined by ζ."

6.3 Existence of Positive-Definite, Symmetric, and Invariant Volume Forms

We now seek to establish the existence of volume forms that are simultaneously positive-definite, symmetric, and invariant under SDiff$^+$. The main result is Theorem 1 below, which demonstrates the existence of such forms and completely characterizes them.

Invariance under SDiff$^+$ requires σ to have the form $\sigma(A) = \zeta(\det A)$ for all $A \in M^+$. Substituting this form into the symmetry condition $\sigma(A^{-1})\det(A) = \sigma(A)$ and moving $\det A$ to the right hand side gives

$$\zeta(\det A^{-1}) = \det A^{-1}\ \zeta(\det A),$$

for all $A \in M^+$. Since $\det A$ can take any value in $\mathbb{R}^+$, this condition can be written as

$$\zeta(u) = u\ \zeta\left(\frac{1}{u}\right), \tag{15}$$

for all $u \in \mathbb{R}^+$.

We can construct all functions ζ satisfying (15) in the following way. Choose any function $\zeta_+ : [1, \infty) \to \mathbb{R}^+$ and define $\zeta : \mathbb{R}^+ \to \mathbb{R}^+$ as

$$\zeta(u) = \begin{cases} \zeta_+(u) & \text{if } u \geq 1 \\ u\ \zeta_+(\frac{1}{u}) & \text{otherwise.} \end{cases} \tag{16}$$

To see that this recipe is complete, note that if ζ satisfies (15), the required ζ_+ is equal to ζ on $[1, \infty)$. Further, any ζ obtained from (16) satisfies (15).

In addition to satisfying (15), we would like ζ to be continuously differentiable. Continuous differentiability of ζ is possible at $u > 1$ and $u < 1$ if and only if ζ_+ is continuously differentiable. At $u = 1$, continuous differentiability is obtained by equating the right and left limits of the derivative of ζ.

We have:

$$\begin{aligned} \frac{d}{du}\zeta(u)\,|_{u=1^+} &= \zeta'_+(1), \\ \frac{d}{du}\zeta(u)\,|_{u=1^-} &= \lim_{u\to 1-}\left\{\zeta_+\left(\frac{1}{u}\right) - u\frac{1}{u^2}\zeta'_+\left(\frac{1}{u}\right)\right\} \\ &= \zeta_+(1) - \zeta'_+(1). \end{aligned}$$

Thus, ζ is continuously differentiable at $u = 1$ if and only if $\zeta'_+(1) = \frac{1}{2}\zeta_+(1)$.

Finally, $\sigma(A) = \zeta(\det A) > 0$ if and only if the function $\zeta > 0$, which is possible if and only if $\zeta_+ > 0$. Thus, we reach the result:

Theorem 1 *The $\mathcal{GL}_2$ objective function is positive-definite, symmetric and invariant under SDiff$^+$ if and only if the function $\sigma : M^+ \to \mathbb{R}$ can be written as*

$$\sigma(A) = \zeta(\det A), \tag{17}$$

where, ζ satisfies $\zeta(u) = u\zeta(1/u)$, and is given by

$$\zeta(u) = \begin{cases} \zeta_+(u) & \text{if } u \geq 1 \\ u\ \zeta_+(\frac{1}{u}) & \text{otherwise,} \end{cases} \tag{18}$$

for a positive and continuously differentiable function $\zeta_+ : [1, \infty) \to \mathbb{R}^+$ satisfying

$$\zeta'_+(1) = \frac{1}{2}\zeta_+(1). \tag{19}$$

Thus there are infinitely many $\mathcal{GL}_2$ objective functions that are positive-definite, symmetric and invariant under SDiff$^+$. The above theorem classifies all of them—each such objective function is uniquely determined by the function ζ_+.

Some recipes for constructing positive-definite, symmetric, and invariant $\mathcal{GL}_2$ objective functions are the following:

1. Take $\zeta_+(u) = 1 + u$. This satisfies the derivative constraint of (19). The resulting σ function is

   $$\sigma(A) = \begin{cases} 1 + \det A & \text{if } \det A \geq 1 \\ \det A\ (1 + \frac{1}{\det A}) = 1 + \det A & \text{otherwise.} \end{cases}$$

   That is, $\sigma(A) = 1 + \det A$ for all $A \in M^+$. This gives the $\mathcal{GL}_2$ objective function

   $$\begin{aligned} &\mathcal{J}((I_1, I_2), f) \\ &\quad = \int_{\mathcal{G}_f} (I_1 \ominus I_2)^2 \alpha \\ &\quad = \int_{\Omega_1} (I_1(x) - I_2(f(x)))^2\ (1 + \det J_f)\ \omega_x, \end{aligned}$$

   which is easily evaluated as being positive-definite, symmetric, and invariant under SDiff$^+$.

2. Generalizing the above example, take

$$\zeta_+(u) = \sqrt[n]{a_n u^n + \cdots + a_1 u + a_0}$$

to be the n-th root of any nth-order symmetric polynomial (i.e. $a_{n-i} = a_i$ for $i = 0, \ldots, \lfloor n/2 \rfloor - 1$) with positive coefficients. Repeating the above calculation gives $\sigma(A) = \sqrt[n]{a_n \det A^n + \cdots + a_1 \det A + a_0}$ and the $\mathcal{GL}_2$ objective function

$$\mathcal{J}((I_1, I_2), f) = \int_{\mathcal{G}_f} (I_1 \ominus I_2)^2 \alpha = \int_{\Omega_1} (I_1(x) - I_2(f(x)))^2 \times \sqrt[n]{a_n \det J_f^n + \cdots + a_1 \det J_f + a_0}\, \omega_x,$$

which is positive-definite, symmetric and invariant under SDiff^+.

7 Lack of Bias and Uniqueness of the Volume Form

Following Theorem 1, will now assume that the $\mathcal{GL}_2$ objective function is given by

$$\mathcal{J}((I_1, I_2), f) = \int_{\Omega_1} (I_1(x) - I_2 \circ f(x))^2 \zeta(\det J_f(x)) \omega_x,$$

where ζ satisfies $\zeta(u) = u\zeta(1/u)$.

When I_1, I_2 are constant images $\mathcal{J}((I_1, I_2), g) = (I_1 - I_2)^2 \int_{\Omega_1} \zeta(\det J_f(x))\omega_x$, so that the objective function is unbiased iff $\int_{\Omega_1} \zeta(\det J_{f_1}(x))\omega_x = \int_{\Omega_1} \zeta(\det J_{f_2}(x))\omega_x$ for all $f_1, f_2 \in \mathrm{Diff}^+$.

We analyze this condition by the calculus of variations. Let $f \in \mathrm{Diff}^+$, and let $\Xi_1 \subset \Omega_1$ be a region that is a proper subset of Ω_1 with boundary $\partial\Xi_1$ (Fig. 2) which is a hypersurface of co-dimension 1 (a curve for the two-dimensional case and a surface for the three-dimensional case). Let $\Xi_2 = f(\Xi_1)$; then Ξ_2 is a proper subset of Ω_2 with a boundary $\partial\Xi_2$. Let X be a smooth vector field on Ω_2 vanishing on the complement Ξ_2' of Ξ_2, and let $\{\Phi_t\}_{t\in\mathbb{R}}$ be the flow of X. Then for all t, the map Φ_t is the identity, and $\Phi_t(\Xi_2) = \Xi_2$, $\Phi_t(\Xi_2') = \Xi_2'$. Thus, $\Phi_t \circ f$ coincides with f on the complement Ξ_1.

Further, for all values of t, $\Phi_t \circ f \in \mathrm{Diff}^+$ (i.e. $\Phi_t \circ f$ is an orientation-preserving diffeomorphism from Ω_1 to Ω_2). Applying the lack of bias condition to f and $\Phi_t \circ f$ we get $\int_{\Omega_1} \zeta(\det J_f(x))\omega_x = \int_{\Omega_1} \zeta(\det J_{\Phi_t \circ f}(x))\omega_x$. Taking into account the fact that $\Phi_t \circ f$ coincides with f on the complement Ξ_2, the condition reduces to $\int_{\Xi_1} \zeta(\det J_f(x))\omega_x = \int_{\Xi_1} \zeta(\det J_{\Phi_t \circ f}(x))\omega_x$, for all values of t. Since the left hand side of this equation is independent of t and $\Phi_0 \circ f = f$, the equation is satisfied only if

$$\frac{d}{dt} \int_{\Xi_1} \zeta(\det J_{\Phi_t \circ f})\omega \mid_{t=0} = 0. \tag{20}$$

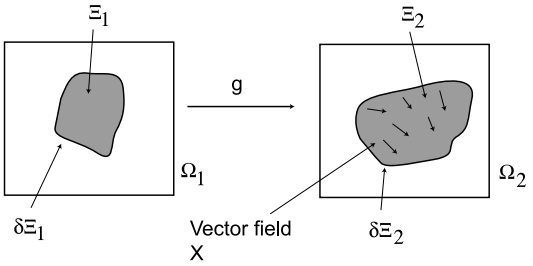


Fig. 2 Analysis of bias

Now,

$$\begin{aligned}\int_{\Xi_1} \zeta(\det J_{\Phi_t \circ f})\omega &= \int_{\Xi_1} \zeta(\det J_{\Phi_t \circ f}) \frac{(\Phi_t \circ f)^*\omega}{\det J_{\Phi_t \circ f}} \\ &= \int_{\Xi_1} \tilde{\zeta}(\det J_{\Phi_t \circ f})\, (\Phi_t \circ f)^*\omega \\ &= \int_{\Xi_1} (\tilde{\zeta} \circ \det J_{\Phi_t \circ f})\, (\Phi_t \circ f)^*\omega,\end{aligned}$$

where $\tilde{\zeta}(u) = \zeta(u)/u$. Further, by invariance of the integral under orientation-preserving diffeomorphisms

$$\begin{aligned}&\int_{\Xi_1} (\tilde{\zeta} \circ \det J_{\Phi_t \circ f})\, (\Phi_t \circ f)^*\omega \\ &= \int_{\Xi_2} (f^{-1})^* \left[(\tilde{\zeta} \circ \det J_{\Phi_t \circ f})\, (\Phi_t \circ f)^*\omega \right] \\ &= \int_{\Xi_2} \left[\tilde{\zeta} \circ \det J_{\Phi_t \circ f} \circ f^{-1} \right] (\Phi_t)^*\omega \\ &= \int_{\Xi_2} \tilde{\zeta} \circ \left[(\det J_{\Phi_t} \circ f).(\det J_f) \right] \circ f^{-1}\, (\Phi_t)^*\omega \\ &= \int_{\Xi_2} \tilde{\zeta} \circ \left[\det J_{\Phi_t} \cdot r_f \right] (\Phi_t)^*\omega,\end{aligned}$$

where $r_f = \det(J_f) \circ f^{-1}$. Hence, (20) can be written as

$$\frac{d}{dt} \int_{\Xi_2} \tilde{\zeta} \circ \left[\det J_{\Phi_t} . r_f \right] (\Phi_t)^*\omega \mid_{t=0} = 0. \tag{21}$$

In Appendix, using Stokes' theorem we show that

Lemma 2 *With notation as above,*

$$\begin{aligned}&\frac{d}{dt} \int_{\Xi_2} \tilde{\zeta} \circ \left[\det J_{\Phi_t} . r_f \right] (\Phi_t)^*\omega \mid_{t=0} \\ &= -\int_{\Xi_2} \langle \zeta''(r_f) dr_f, X \rangle \omega,\end{aligned} \tag{22}$$

where $\langle \zeta''(r_f) dr_f, X \rangle$ denotes the pointwise evaluation of the 1-form $\zeta''(r_f) dr_f$ on the vector field X.

This lemma leads to the following fundamental result.

Theorem 2 *The $\mathcal{GL}_2$ objective function determined by ζ is positive-definite, symmetric, invariant under SD^+ and unbiased if and only if*

$$\zeta(r) = a(1+r), \tag{23}$$

for some $a > 0$.

Proof Lack of bias requires (21) to hold. According to Lemma 2, this implies that for all regions Ξ_1 bounded by a simple closed curve, $(\zeta'' \circ r_f) dr_f = 0$ on Ξ_1. For $r \in \mathbb{R}^+$, there exists a region Ξ_1 containing any point $u \in \Omega_1$, and a diffeomorphism $f \in \text{Diff}^+$ such that $dr_f(u) \neq 0$ and $r_f(u) = r$. Hence $\zeta''(r) = 0$ for all $r > 0$, and therefore $\zeta(r) = ar + b$ for some $a, b \in \mathbb{R}$. Using (19) we conclude that $a = b$. Finally, positivity of ζ implies $a > 0$, proving the result. □

Writing $\zeta(r) = a(r+1)$, we see that the a factor just scales the $\mathcal{GL}_2$ objective function according to

$$\mathcal{J}((I_1, I_2), f) = a \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, (\det J_f + 1)\omega,$$

and setting $a = 1$ does not affect the minimizer of $\mathcal{J}$. Thus we are led to the conclusion that, up to a scalar multiple, there is a unique $\mathcal{GL}_2$ objective function that satisfies our axioms, and it is given by

$$\mathcal{J}((I_1, I_2), f) = \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, (\det J_f + 1)\omega. \tag{24}$$

Thus the tangent-dependent volume form that is used in this $\mathcal{GL}_2$ objective function is

$$\alpha_{(x, f(x))} = (1 + \det J_f(x)) p_1^* \omega_x. \tag{25}$$

7.1 Relation to Other Registration Objective Functions

We pause briefly to comment on the relation of Theorem 2 to previously published symmetric and approximately symmetric registration algorithms.

7.1.1 Direct Methods

The objective function of Christensen and his colleagues [7–9] contains two diffeomorphisms f and g:

$$\begin{aligned}\mathcal{J}((I_1, I_2), f, g) &= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega + \int_{\Omega_2} (I_1 \circ g - I_2)^2 \, \omega \\ &\quad + \lambda \left\{ \int_{\Omega_1} (g \circ f - id)^2 \omega + \int_{\Omega_2} (f \circ g - id)^2 \omega \right\}, \end{aligned} \tag{26}$$

where $\lambda > 0$ is a constant and id is the identity function. The terms multiplied by λ are penalty terms, and as $\lambda \to \infty$ the minimizing functions f and g become inverses of each other. For a finite value of λ, the minimizing functions f and g need not be inverses of each other and hence this objective function is only approximately symmetric.

There is a simple relation between the objective function of (26) and the $\mathcal{GL}_2$ objective function of (24). To see this, we rewrite the $\mathcal{GL}_2$ objective function as

$$\begin{aligned}&\mathcal{J}((I_1, I_2), f) \\ &= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, (\det J_f + 1)\omega \\ &= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega + \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \det J_f \omega.\end{aligned}$$

Changing the domain of the second integral on the right hand side of the above equation from Ω_1 to Ω_2 by using $f^{-1} : \Omega_2 \to \Omega_1$ gives

$$\begin{aligned}&\mathcal{J}((I_1, I_2), f) \\ &= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega + \int_{\Omega_2} (I_1 \circ f^{-1} - I_2 \circ f \circ f^{-1})^2 \\ &\quad \times \det J_f \circ f^{-1} \, \det J_{f^{-1}} \omega \\ &= \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \, \omega + \int_{\Omega_2} (I_1 \circ f^{-1} - I_2)^2 \, \omega. \end{aligned} \tag{27}$$

This is identical to the objective function of (26) when f and g are inverses of each other, showing that the $\mathcal{GL}_2$ objective function is the "exactly symmetric" form of the objective function of (26).

7.1.2 Composition Methods

Composition methods [11, 14–16, 19] have objective functions which do not directly involve the forward or the backward map. These methods require a third domain in addition to Ω_1 and Ω_2. Let this domain be Ω, and let $\phi_1 : \Omega \to \Omega_1$ and $\phi_2 : \Omega \to \Omega_2$ be two diffeomorphisms in Diff^+. Ignoring regularization, the objective function of this method is

$$\mathcal{J}((I_1, I_2), \phi_1, \phi_2) = \int_{\Omega} (I_1 \circ \phi_1 - I_2 \circ \phi_2)^2 \omega. \tag{28}$$

The objective function is minimized with respect to ϕ_1, ϕ_2 and the warp and its inverse are taken to be $f = \phi_2 \circ \phi_1^{-1}$ and $f^{-1} = \phi_1 \circ \phi_2^{-1}$. The complication in analyzing this objective function is seen by changing the domain of the integral in (28) to Ω_1 using $\phi_1^{-1} : \Omega_1 \to \Omega$:

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), \phi_1, \phi_2) \\
&\quad = \int_{\Omega} (I_1 \circ \phi_1 - I_2 \circ \phi_2)^2 \omega \\
&\quad = \int_{\Omega_1} (I_1 - I_2 \circ \phi_2 \circ \phi_1^{-1})^2 \det J_{\phi_1^{-1}} \omega \\
&\quad = \int_{\Omega_1} (I_1 - I_2 \circ f)^2 \det J_{\phi_1^{-1}} \omega. \qquad (29)
\end{aligned}$$

Note that for any $g \in \mathrm{Diff}^+$, the functions $\phi_1 \circ g$ and $\phi_2 \circ g$ give the same warp as ϕ_1, ϕ_2 (since $\phi_2 \circ g \circ (\phi_1 \circ g)^{-1} = f$). However (29) shows that $\mathcal{J}((I_1, I_2), \phi_1 \circ g, \phi_2 \circ g)$ may not be necessarily identical to $\mathcal{J}((I_1, I_2), \phi_1, \phi_2)$. Thus the composition objective function can change value even when the warping function stays fixed. What this means geometrically is not clear. This comment should not be taken as a criticism of these methods. It is only meant to point out that these methods may not have a natural geometric interpretation.

It is easy to modify the objective function of (28) so that it has a natural geometric interpretation. The modification is to multiply the integrand with a $(\det J_{\phi_1} + \det J_{\phi_2})$ term.

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), \phi_1, \phi_2) \\
&\quad = \int_{\Omega} (I_1 \circ \phi_1 - I_2 \circ \phi_2)^2 (\det J_{\phi_1} + \det J_{\phi_2}) \omega. \qquad (30)
\end{aligned}$$

As before, changing the domain of the integral from Ω to Ω_1 using ϕ_1^{-1} gives

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), \phi_1, \phi_2) \\
&\quad = \int_{\Omega} (I_1 \circ \phi_1 - I_2 \circ \phi_2)^2 (\det J_{\phi_1} + \det J_{\phi_2}) \omega \\
&\quad = \int_{\Omega_1} (I_1 - I_2 \circ \phi_2 \circ \phi_1^{-1})^2 (\det J_{\phi_1} + \det J_{\phi_2}) \circ \phi_1^{-1} \\
&\qquad \times \det J_{\phi_1^{-1}} \omega \\
&\quad = \int_{\Omega_1} (I_1 - I_2 \circ f)^2 (1 + \det J_{\phi_2} \circ \phi_1^{-1} \det J_{\phi_1^{-1}}) \omega \\
&\quad = \int_{\Omega_1} (I_1 - I_2 \circ f)^2 (1 + \det J_f) \omega,
\end{aligned}$$

since $\det J_{\phi_2} \circ \phi_1^{-1} \det J_{\phi_1^{-1}} = \det J_f$, because $\phi_2 \circ \phi_1^{-1} = f$. This is, of course, identical to the $\mathcal{GL}_2$ objective function that we have been discussing so far.

8 The IT Objective Functions

Having symmetrized $\mathcal{L}_2$, we turn our attention to IT objective functions. Many different versions of IT objective functions exist in practice. To treat them uniformly we use the model of Fig. 3. Central to the model is a joint intensity histogram which is calculated from I_1, I_2 and f. This is shown in the left half of Fig. 3. The calculated joint histogram belongs to the space $\mathcal{H}$ defined below. A functional $\Phi : \mathcal{H} \to \mathbb{R}$ maps the space of joint histograms onto the real line, and the value of Φ for the histogram calculated from I_1, I_2 and f is taken as the value of the objective function. Below, in Sect. 8.2, we show that most of the commonly used IT objective functions conform to this model. Different objective functions correspond to different choices of Φ.

IT objective functions that conform to this model are in general asymmetric and, as we show below, the asymmetry comes from the calculation of the joint histogram rather than Φ. Just as for the $\mathcal{L}_2$ objective function, it turns out that the asymmetry in the joint histogram is due to the use of the standard volume form. Replacing the standard volume form with the volume form of Theorem 2 symmetrizes the joint histogram calculation and hence symmetrizes the IT objective functions.

So far we have not imposed any restrictions on the images I_1 and I_2 other than differentiability. The co-domain—which we also call the *range space*—of such images is, of course, $\mathbb{R}$. In practice, digital images have a range space which is a bounded interval and defining histograms for a bounded range space is simple. To exploit this simplicity, we will assume below that the range space of all images is $[0, L)$ for some finite fixed L.

8.1 The Model for IT Objective Functions

The model of Fig. 3 is defined mathematically as follows.

Definition For $N > 0$, let $\mathcal{H}$ be the set of $N \times N$ matrices with non-negative entries that sum to one.

Thus, $H \in \mathcal{H}$ means $H_{ij} \geq 0$ and $\sum_{ij} H_{ij} = 1$. Also, $H \in \mathcal{H}$ implies that $H^T \in \mathcal{H}$.

Definition An *IT functional* is any functional $\Phi : \mathcal{H} \to \mathbb{R}$. The IT functional is *positive semi-definite* if $\Phi(H) \geq 0$ for any $H \in \mathcal{H}$. The IT functional is *symmetric* if $\Phi(H) = \Phi(H^T)$ for any $H \in \mathcal{H}$.

For the rest of this paper, we restrict attention to IT functionals that are positive semi-definite and symmetric. Therefore we will drop these adjectives, except in definitions and statements of theorems.

Next, we define the process of calculating the joint histogram from I_1, I_2 and f.

Definition The *joint intensity space* C is $[0, L) \times [0, L)$ where $[0, L)$ is the range space of images.

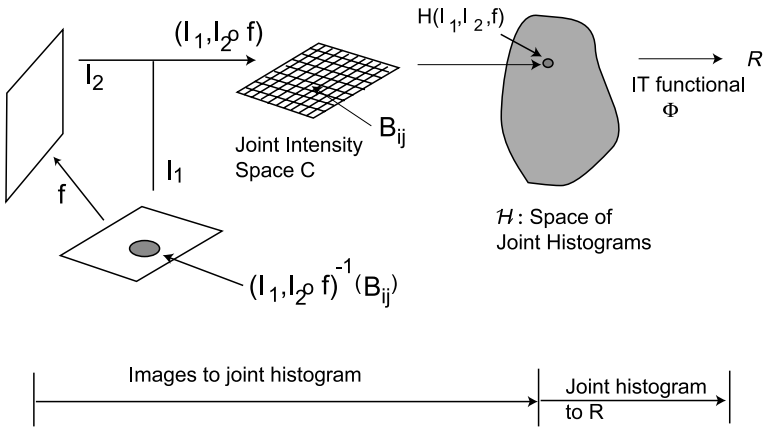


Fig. 3 The model for IT objective functions

If $(u, v) \in C$, then $(v, u) \in C$, i.e. the joint intensity space is symmetric.

Definition A collection of sets $\{B_{ij},\ i, j = 1, \ldots, N\}$ is a *binning* of C if

1. $\{B_{ij},\ i, j = 1, \ldots, N\}$ partitions C, and
2. Every B_{ij} is $\mathbb{R}^2$-measurable with a measure strictly greater than zero.

The sets B_{ij} are referred to as *bins*. Further, the binning is *symmetric* if $(u, v) \in B_{ij}$ implies $(v, u) \in B_{ji}$.

Henceforth, we always assume $\{B_{ij}\}$ to be a symmetric binning of C. We do not assume that the bins are square or have equal measure.

Define $\mu_{\Omega_1}(A) = \int_A \omega / \int_{\Omega_1} \omega$ for any measurable set $A \subseteq \Omega_1$. Note that μ_{Ω_1} is a probability measure. Using the function $(I_1, I_2 \circ f) : \Omega_1 \to C$ define an $N \times N$ matrix $H(I_1, I_2 f)$ by

$$H(I_1, I_2, f)_{ij} = \mu_{\Omega_1}((I_1, I_2 \circ f)^{-1}(B_{ij})). \tag{31}$$

Definition Given images I_1, I_2 and a diffeomorphism $f : \Omega_1 \to \Omega_2$, the *registration joint histogram*, or simply *the joint histogram*, is the $N \times N$ matrix $H(I_1, I_2, f)$ defined above.

Note the emphasis in notation for the joint histogram. An arbitrary element of $\mathcal{H}$ is still referred to as H. However, when the element comes from non-rigid registration, it is referred to as $H(I_1, I_2, f)$.

We now have all of the ingredients for defining an IT objective function

Definition An *IT objective function* $\mathcal{J}$ is any function of the form

$$\mathcal{J}((I_1, I_2), f) = \Phi(H(I_1, I_2, f)), \tag{32}$$

where Φ is a positive semi-definite IT functional and $H(I_1, I_2, f)$ is the joint histogram.

8.2 Commonly Used IT Objective Functions

Two classes of IT objective functions are commonly used and we now show that they conform to the model defined above. Both classes of objective functions use mutual information [26, 28] and normalized mutual information [27] as IT functionals, and we define these first.

Two *marginal vectors* of size $N \times 1$ and $1 \times N$ are associated to $H \in \mathcal{H}$. These vectors, denoted H^1 and H^2, are defined by $H^1_i = \sum_j H_{ij}$ and $H^2_j = \sum_i H_{ij}$. The Shannon entropies of H, H^1 and H^2 are defined by

$$\begin{aligned} S(H) &= -\sum_{ij} H_{ij} \log H_{ij}, \\ S(H^1) &= -\sum_{i} H^1_i \log H^1_i, \\ S(H^2) &= -\sum_{j} H^2_j \log H^2_j. \end{aligned} \tag{33}$$

The mutual information (MI) and the normalized mutual information (NMI) of H are

$$\mathrm{MI}(H) = -S(H) + S(H^1) + S(H^2), \tag{34}$$

$$\mathrm{NMI}(H) = \frac{S(H^1) + S(H^2)}{S(H)}. \tag{35}$$

Transposing H does not change $S(H)$, but exchanges the values of $S(H^1)$ and $S(H^2)$. Because $S(H^1)$ and $S(H^1)$ only appear as a sum in MI and NMI, transposing H leaves MI and NMI unaltered, i.e. $\mathrm{MI}(H) = \mathrm{MI}(H^T)$, and $\mathrm{NMI}(H) = \mathrm{NMI}(H^T)$.

The first class of IT objective functions we consider are unsmoothed histogram-based objective functions.

8.2.1 Unsmoothed Histogram-Based Objective Functions

Unsmoothed histogram-based objective functions use equal-sized square bins

$$B_{ij} = [(i-1)L/N, iL/N) \times [(j-1)L/N, jL/N), \quad i, j = 1, \ldots, N. \tag{36}$$

This binning is symmetric, and the bins have a measure (area) strictly greater than zero. Therefore, they conform to the definition used in our model.

The IT functional in these methods is either the negative MI or the negative NMI (These algorithms usually maximize MI or NMI. We take the negative of these values since we are minimizing the objective function.) The negative mutual information functional is

$$\Phi_{\mathrm{MI}}(H) = -\mathrm{MI}(H), \tag{37}$$

and, the negative normalized mutual information functional is

$$\Phi_{\mathrm{NMI}}(H) = -\mathrm{NMI}(H). \tag{38}$$

The functionals Φ_{MI} and Φ_{NMI} can be made positive semi-definite by adding constants and by adding a small positive constant to the denominator in (35). Further, since $\mathrm{MI}(H) = \mathrm{MI}(H^T)$ and $\mathrm{NMI}(H) = \mathrm{NMI}(H^T)$, the functionals Φ_{MI} and Φ_{NMI} are symmetric (adding a positive constant to the functional or to the denominator in (35) does not change this). Hence the above functionals conform to the model of Sect. 8.1.

The objective functions $\mathcal{J}_{\mathrm{MI}}$ and $\mathcal{J}_{\mathrm{NMI}}$ of these methods are given by

$$\mathcal{J}_{\mathrm{MI}}((I_1, I_2), f) = \Phi_{\mathrm{MI}}(H(I_1, I_2, f)) \quad \text{and} \tag{39}$$

$$\mathcal{J}_{\mathrm{NMI}}((I_1, I_2), f) = \Phi_{\mathrm{NMI}}(H(I_1, I_2, f)). \tag{40}$$

These too conform to the model of Sect. 8.1.

Variants of the above objective functions using other entropies have also been proposed [30] and it is straightforward to check that, with minor modifications, they also conform to the model of Sect. 8.1.

The second class of IT objective functions we consider are Parzen kernel-based objective functions.

8.2.2 Parzen Kernel-Based Objective Functions

The objective functions in the second class we consider follow a slightly different philosophy. They are not histogram-based, but instead calculate an estimate of joint and marginal probability densities (continuous functions defined on a non-discretized domain, but about which we have only discrete data in practice) using a convolution function known as a *Parzen kernel* [26, 28, 29]. The joint and marginal densities are used to calculate differential entropies from which mutual and normalized mutual information are calculated and used as the objective function.

These objective functions are difficult to analyze exactly, but it is possible to closely approximate them with the model proposed in Sect. 8.1. The approximation starts from the observation that in practice, most images are digitized so that their gray levels take only integer values (typically $0, \ldots, 255$). For such images there are only finitely many joint intensity values (typically 256^2). Suppose we choose bins in C of the type given in (36) such that each bin contains exactly one of the finitely many joint intensity values. Let the center of the ij^{th} bin be $(a_{ij}, b_{ij}) \in C$. Then the Parzen kernel-based mutual and normalized information IT functional can be approximated as the following functionals of the joint histogram:

$$\Phi_{\mathrm{pMI}}(H) = -\mathrm{MI}(H * K), \quad \text{and,} \tag{41}$$

$$\Phi_{\mathrm{pNMI}}(H) = -\mathrm{NMI}(H * K), \tag{42}$$

where $H * K$ is the smoothed joint histogram

$$(H * K)_{ij} = \sum_{mn} K((a_{ij}, b_{ij}), (a_{mn}, b_{mn}))H_{mn} \tag{43}$$

and $K : C \times C \to \mathbb{R}$ is the Parzen kernel.

Two properties of the Parzen kernels that are used in practice are relevant to our analysis—(1) K is non-negative and (2) K is symmetric, i.e.

$$K((u, v), (r, s)) \geq 0, \quad \text{for all } (u, v) \in C,\ (s, r) \in C,$$

$$K((u, v), (r, s)) = K((v, u), (s, r)), \quad \text{for all } (u, v) \in C,\ (s, r) \in C.$$

Using these properties and arguing as in the above section, it is simple to show that IT functionals in (41) and (42) are symmetric and can be trivially altered to be positive semi-definite. The objective functions $\mathcal{J}_{\mathrm{pMI}}$ and $\mathcal{J}_{\mathrm{pNMI}}$ of these methods are given by

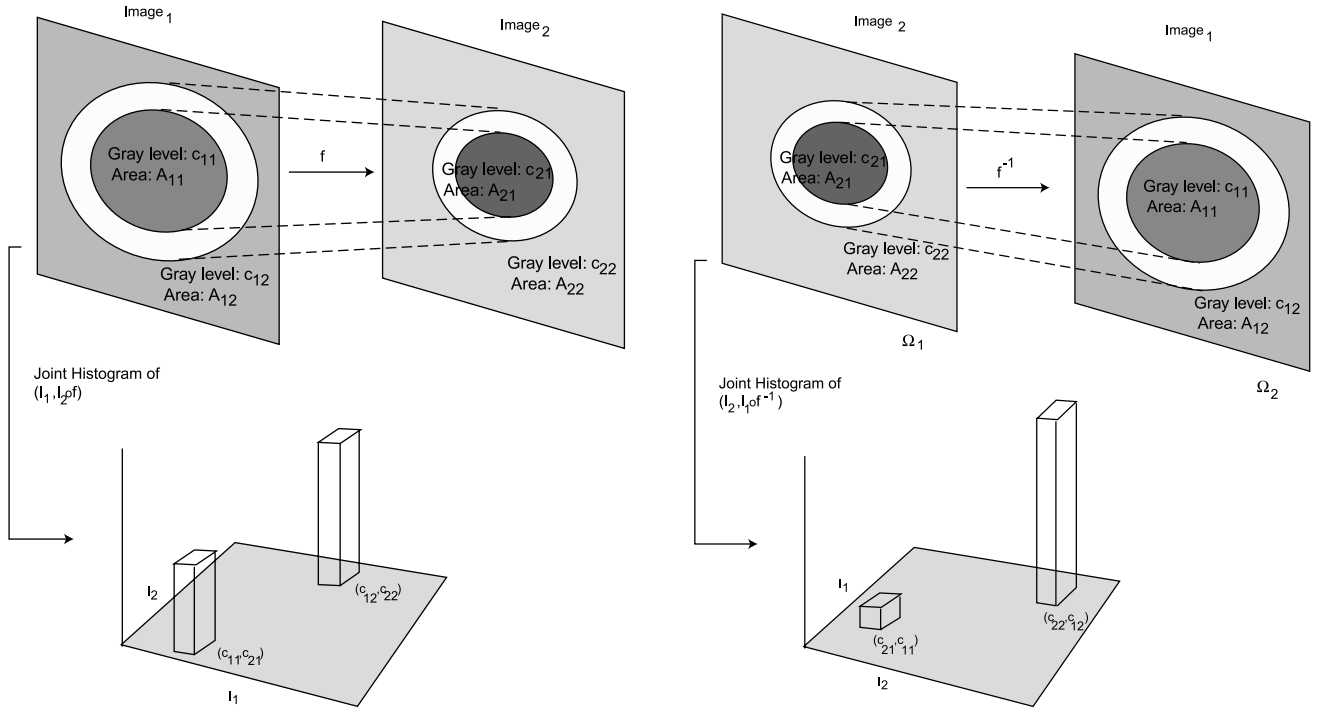
$$\mathcal{J}_{\mathrm{pMI}}((I_1, I_2), f) = \Phi_{\mathrm{pMI}}(H(I_1, I_2, f)) \quad \text{and} \tag{44}$$

$$\mathcal{J}_{\mathrm{pNMI}}((I_1, I_2), f) = \Phi_{\mathrm{pNMI}}(H(I_1, I_2, f)). \tag{45}$$

They conform to the model of Sect. 8.1.

8.3 Asymmetry in IT Objective Functions

All objective functions conforming to the model of Sect. 8.1 are asymmetric. As a simple example of this, consider the MI objective function of (39) for the two images shown in Fig. 4a with two concentric circles in each image. The regions inside the inner circle and outside the outer circle have areas (as calculated by the standard volume form)



(a) Original (b) Swapped

Fig. 4 Asymmetry in IT objective functions

$A_{11}, A_{12}, A_{21}, A_{22}$ as shown. The gray levels inside the inner circle and outside the outer circle are constant and take values $c_{11}, c_{12}, c_{21}, c_{22}$ as shown. In the zone between the two circles, the gray levels make a smooth transition. Figure 4a also shows a diffeomorphism $f : \Omega_1 \to \Omega_2$ which takes the circles in Ω_1 to the circles in Ω_2. Assume that the histogram bins in C are small enough that all distinct joint combinations of the intensities $\{c_{11}, c_{12}\} \times \{c_{21}, c_{22}\}$ fall into distinct bins. Then the joint histogram of $I_1, I_2 \circ f$ has two peaks, one at the location (c_{11}, c_{21}) and another at the location (c_{12}, c_{22}). We are ignoring the contribution to the histogram from the transition region because the area of the transition region can be made arbitrarily small. Swapping the two images and using f^{-1} swaps the circles and gives Fig. 4b.

The probability measures of the areas inside the inner circle and outside the outer circle are relevant to our discussion. They are

$$\begin{aligned}
p_{11} &= \mu_{\Omega_1}\ (\text{inside region of inner circle}) = A_{11}/\|\Omega_1\|,\\
p_{12} &= \mu_{\Omega_1}\ (\text{outside region of outer circle}) = A_{12}/\|\Omega_1\|\\
&= 1 - p_{11},\\
p_{21} &= \mu_{\Omega_2}\ (\text{inside region of inner circle}) = A_{21}/\|\Omega_2\|,\\
p_{22} &= \mu_{\Omega_2}\ (\text{outside region of outer circle}) = A_{22}/\|\Omega_2\|\\
&= 1 - p_{21},
\end{aligned}$$

where $\|\Omega_1\| = \|\Omega_2\|$ are the areas of Ω_1 and Ω_2 calculated according to the standard volume form.

Evaluating the objective function gives

$$\begin{aligned}
\mathcal{J}_{\mathrm{MI}}((I_1, I_2), f) &= \{p_{11}\log p_{11} + p_{12}\log p_{12}\}\\
&\quad - \{p_{11}\log p_{11} + p_{12}\log p_{12}\}\\
&\quad - \{p_{11}\log p_{11} + p_{12}\log p_{12}\}\\
&= -(p_{11}\log p_{11} + p_{12}\log p_{12})\\
&= -(p_{11}\log p_{11} + (1-p_{11})\log(1-p_{11}))
\end{aligned}$$

After swapping the images, the objective function with f^{-1} evaluates to

$$\begin{aligned}
&\mathcal{J}_{\mathrm{MI}}((I_2, I_1), f^{-1})\\
&\quad = -(p_{21}\log p_{21} + (1-p_{21})\log(1-p_{21})).
\end{aligned}$$

Thus, unless coincidentally p_{11} happens to be equal to p_{21} or $1 - p_{21}$, the value of $\mathcal{J}_{\mathrm{MI}}((I_1, I_2), f)$ will not be equal to the value of $\mathcal{J}_{\mathrm{MI}}((I_2, I_1), f^{-1})$. This shows that the objective function is asymmetric.

Speaking more generally, the problem is that joint histogram in Fig. 4a has a bin count p_{11} in the bin at (c_{11}, c_{21}) and a bin count of p_{12} in the bin at (c_{12}, c_{22}) whereas the joint histogram in Fig. 4b has a bin count p_{21} in the bin at (c_{21}, c_{11}) and a bin count of p_{22} in the bin at (c_{22}, c_{12}).

Since the bin counts depend on the standard area of the circles, they can be made arbitrarily different by simply choosing circles with different areas. This shows that the ultimate cause of asymmetry is the standard volume form.

The above argument can be made to work (i.e we can always find a pair of images which demonstrates asymmetry) for the model of Sect. 8.1 as long as the IT functional is not a constant functional.

9 Graph-Based IT Objective Functions

The IT objective functions can be symmetrized by making the histogram calculation graph-based using the volume form of (25).

Let $\mathcal{G}_f$ be the graph of the diffeomorphism $f : \Omega_1 \to \Omega_2$. For any measurable $A \subseteq \mathcal{G}_f$, let

$$\tilde{\mu}_{\mathcal{G}_f}(A) = \int_A \alpha \Big/ \int_{\mathcal{G}_f} \alpha$$

be the of probability measure of A, where α is the volume form of (25). The object $\tilde{\mu}_{\mathcal{G}_f}$ is a probability measure on $\mathcal{G}_f$. It is easy to show that $\int_{\mathcal{G}_f} \alpha$ evaluates to $\int_{\mathcal{G}_f} \alpha = \int_{\Omega_1} \omega + \int_{\Omega_2} \omega = 2\ \mathrm{vol}(\Omega_1)$ independent of f, where $\mathrm{vol}(\Omega_1)$ is the Euclidean volume of Ω_1. Thus,

$$\tilde{\mu}_{\mathcal{G}_f}(A) = \int_A \alpha \Big/ \int_{\mathcal{G}_f} \alpha = \frac{\int_A \alpha}{2\mathrm{vol}(\Omega_1)}.$$

Below, we use this formula in calculations.

The function $(I_1 \circ p_1, I_2 \circ p_2)$ maps $\mathcal{G}_f$ to the joint intensity space C which is tessellated by bins B_{ij}, $i, j = 1, \ldots, N$, motivating the following definition:

Definition For any images I_1, I_2 and diffeomorphism $f : \Omega_1 \to \Omega_2$, the *graph-based joint histogram* is the $N \times N$ matrix $H_g(I_1, I_2, f)$, where $H_g(I_1, I_2, f)_{ij}$ defined by

$$H_g(I_1, I_2, f)_{ij} = \tilde{\mu}_{\mathcal{G}_f}((I_1 \circ p_1, I_2 \circ p_2)^{-1}(B_{ij})),$$
$$\text{for } i, j = 1, \ldots, N,$$

is the fractional area of the set of points of $\mathcal{G}_f$ that are mapped by $(I_1 \circ p_1, I_2 \circ p_2)$ into the bin B_{ij}.

It is easy to check that all elements of any graph-based joint histogram are non-negative and sum to one. Hence graph-based joint histograms are elements of $\mathcal{H}$, and any IT functional can be used with them.

Definition A *graph-based IT* (GIT) *objective function* $\mathcal{J}$ is any function of the form

$$\mathcal{J}((I_1, I_2), f) = \Phi(H_g(I_1, I_2, f)), \tag{46}$$

where Φ is a positive semi-definite and symmetric IT functional.

We now proceed to investigate properties of GIT objective functions.

9.1 Positive Semi-Definiteness

A modification is required in the definition of positive-definiteness. For $\mathcal{L}_2$ and $\mathcal{GL}_2$, if the objective function equals zero, then I_1 and $I_2 \circ f$ are equal almost everywhere. Because the GIT objective functions use joint histograms, this property cannot be guaranteed anymore. Instead we require that the GIT objective function be positive semi-definite.

Definition An objective function $\mathcal{J}$ is positive semi-definite if $\mathcal{J}((I_1, I_2), f) \geq 0$ for all I_1, I_2, f.

Proposition 5 *Every GIT objective function is positive semi-definite.*

Proof This follows from the positive semi-definiteness of Φ (in (46)). □

All other definitions are identical to the previous ones and will not repeated.

The next two subsections will show that the GIT objective function is symmetric and invariant under SDiff^+. The proofs of these properties require us to consider the volume form α on the graphs of two different diffeomorphisms. Since the volume form depends on the function we need notation to distinguish between the two volume forms. We will explicitly denote the volume form α on the graph of a diffeomorphism f by $\alpha^{(f)}$. This will distinguish it, for example, from the volume form on the graph of a diffeomorphism g, since the latter volume form will be denoted $\alpha^{(g)}$.

9.2 Symmetry

Proposition 6 *Every GIT objective function is symmetric.*

Proof We will compare $\mathcal{J}((I_1, I_2), f)$ and $\mathcal{J}((I_2, I_1), f^{-1})$ with reference to Fig. 5. The bottom part of the figure shows the calculation of $\mathcal{J}((I_1, I_2), f)$ and the top part shows the calculation of $\mathcal{J}((I_2, I_1), f^{-1})$. Because the graphs $\mathcal{G}_f$ and $\mathcal{G}_{f^{-1}}$ are not identical, their projection functions to Ω_1 and Ω_2 are different and are shown as p_{11}, p_{12} for $\mathcal{G}_f$ and p_{21}, p_{22} for $\mathcal{G}_{f^{-1}}$. The joint intensity spaces for the top and bottom part of the figure are both C, and the two joint intensity spaces are tesselated by the symmetric binning $\{B_{ij}\}$.

Consider the bin B_{ij} in the joint intensity space on the bottom part of Fig. 5 and the bin B_{ji} in the joint intensity

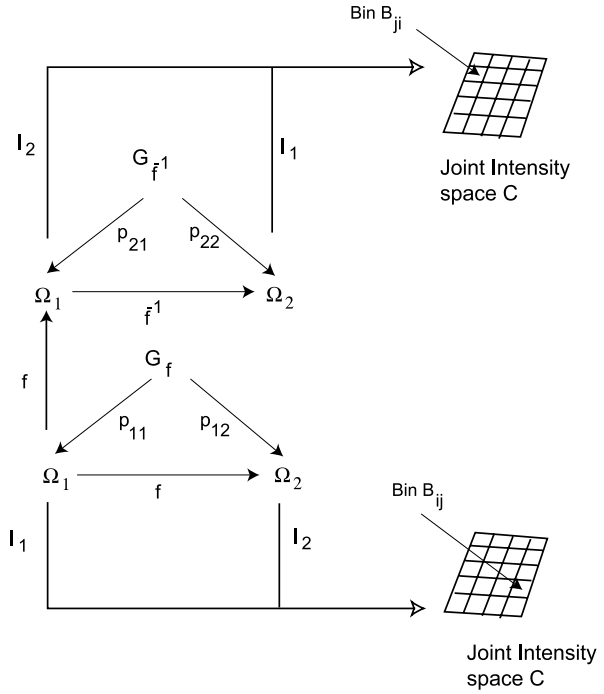


Fig. 5 Analysis of symmetry for GIT

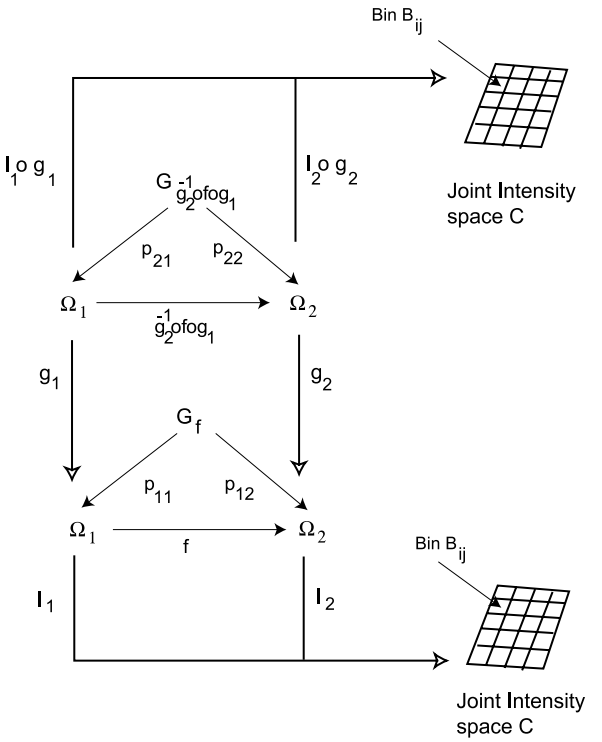


Fig. 6 Analysis of invariance under SDiff^+ for GIT

space on the top part of Fig. 5. We first want to establish that

$$\int_{(I_2\circ p_{21}, I_1\circ p_{22})^{-1}(B_{ji})} \alpha^{(f^{-1})} = \int_{(I_2\circ p_{11}, I_1\circ p_{12})^{-1}(B_{ij})} \alpha^{(f)}, \tag{47}$$

where the first integral is on $\mathcal{G}_{f^{-1}}$ and the second integral is on $\mathcal{G}_f$.

Starting with the first integral in (47), we express it as an integral on the top copy of Ω_1 and then pull it back on the bottom copy of Ω_1 via f.

$$\int_{(I_2\circ p_{21}, I_1\circ p_{22})^{-1}(B_{ji})} \alpha^{(f^{-1})}$$

$$= \int_{O_2} (1 + \det J_{f^{-1}})\, \omega,$$

where $O_2 = \{u \in \Omega_1 \mid (I_2(u), I_1 \circ f^{-1}(u)) \in B_{ji}\}$

$$= \int_{\tilde{O}_2} ((1 + \det J_{f^{-1}}) \circ f) \det J_f\, \omega,$$

where $\tilde{O}_2 = f^{-1}(O_2)$

$$= \{u \in \Omega_1 \mid (I_2 \circ f(u), I_1(u)) \in B_{ji}\}$$

$$= \int_{\tilde{O}_2} (\det J_f + 1)\, \omega, \tag{48}$$

since $(\det J_{f^{-1}} \circ f) \det J_f = 1$.

The second integral in (47) expressed as an integral on the bottom copy of Ω_1 gives

$$\int_{(I_2\circ p_{11}, I_1\circ p_{12})^{-1}(B_{ij})} \alpha^{(f)} = \int_{\tilde{O}_1} (1 + \det J_f)\omega, \tag{49}$$

where $\tilde{O}_1 = \{u \in \Omega_1 \mid (I_1(u), I_2 \circ f(u)) \in B_{ij}\}$. However, because the binning is symmetric, $(I_2 \circ f(u), I_1(u)) \in B_{ji}$ if and only if $(I_1(u), I_2 \circ f(u)) \in B_{ij}$, and we have $\tilde{O}_1 = \tilde{O}_2$. This and (48) and (49) establish the equality in (47).

Dividing both sides of (47) by $2\mathrm{vol}(\Omega_1)$ gives $H_g(I_2, I_1, f^{-1})_{ji} = H_g(I_1, I_2, f)_{ij}$, which shows that the joint histogram $H_g(I_2, I_1, f^{-1})$ is the transpose of the joint histogram $H_g(I_1, I_2, f)$. Then, the symmetry of the IT functional gives symmetry of the GIT objective function. This completes the proof. □

9.3 Invariance under SDiff^+

Proposition 7 *Every GIT objective function is invariant under* SDiff^+.

Proof The proof is similar in spirit to the proof of symmetry.

We compare $\mathcal{J}((I_1, I_2), f)$ and $\mathcal{J}((I_1 \circ g_1, I_2 \circ g_2), g_2^{-1} \circ f \circ g^{-1})$ with the help of Fig. 6. The bottom part of the figure shows the diffeomorphism $f : \Omega_1 \to \Omega_2$, the graph $\mathcal{G}_f$, and the map to the joint intensity space. The top part shows the diffeomorphism $g_2^{-1} \circ f \circ g_1 : \Omega_1 \to \Omega_2$, the graph

$\mathcal{G}_{g_2^{-1}\circ f\circ g_1}$, and the map to the joint intensity space. Because the two graphs are not identical, their projection functions to Ω_1 and Ω_2 are different and are shown as p_{11}, p_{12} for $\mathcal{G}_f$ and p_{21}, p_{22} for $\mathcal{G}_{g_2^{-1}\circ f\circ g_1}$. The joint intensity spaces for the top and bottom part of the figure are both C.

We first show that

$$\int_{(I_1\circ p_{11}, I_2\circ p_{12})^{-1}(B_{ij})} \alpha^{(f)} = \int_{(I_1\circ g_1\circ p_{21}, I_2\circ g_2\circ p_{22})^{-1}(B_{ij})} \alpha^{(g_2^{-1}\circ f\circ g_1)} \tag{50}$$

where the first integral is on $\mathcal{G}_f$ and the second on $\mathcal{G}_{g_2^{-1}\circ f\circ g_1}$. In contrast to the proof of symmetry, the same bin appears on both sides of the equation.

Starting with the integral on the left hand of (50) and evaluating it on the bottom copy of Ω_1 gives

$$\int_{(I_1\circ p_{11}, I_2\circ p_{12})^{-1}(B_{ij})} \alpha^{(f)} = \int_{\tilde{O}_1} (1+\det J_f)\omega, \tag{51}$$

where $\tilde{O}_1 = \{u \in \Omega_1 \mid (I_1(u), I_2\circ f(u)) \in B_{ij}\}$.

Evaluating the integral on the right hand side of (50) on the top copy of Ω_1 gives

$$\int_{(I_1\circ g_1\circ p_{21}, I_2\circ g_2\circ p_{22})^{-1}(B_{ij})} \alpha^{(g_2^{-1}\circ f\circ g_1)} = \int_{O_2} (1+\det J_{g_2^{-1}\circ f\circ g_1})\omega, \tag{52}$$

where $O_2 = \{u \in \Omega_1 \mid (I_1\circ g_1(u), I_2\circ g_2\circ g_2^{-1}\circ f\circ g_1(u)) \in B_{ij}\}$. The Jacobian of a composition of functions is the product of the Jacobians of the functions evaluated at appropriate points. Also, $g_1, g_2 \in \text{SDiff}^+$ and $\det J_{g_2^{-1}}$, $\det J_{g_1}$ and $\det J_{g_1^{-1}} = 1$ are identically 1. Thus the integral in (52) reduces to $\int_{O_2}(1+(\det J_f)\circ g_1)\omega$. Pulling this integral back to the bottom copy of Ω_1 by g_1^{-1} we find that it is equal to

$$\int_{\tilde{O}_2} (1+\det J_f)\omega, \tag{53}$$

where $\tilde{O}_2 = \{u \in \Omega_1 \mid g_1^{-1}(u) \in O_2\} = \{u \in \Omega_1 \mid (I_1(u), I_2\circ f(u)) \in B_{ij}\}$. Hence $\tilde{O}_1 = \tilde{O}_2$ and the equalities in (51) and (53) give (50).

Dividing (50) on both sides by $2\text{vol}(\Omega_1)$, we get $H_g(I_1, I_2, f)_{ij} = H_g(I_1\circ g_1, I_2\circ g_2, g_2^{-1}\circ f\circ g_1^{-1})_{ij}$, showing that the joint histograms $H_g(I_1, I_2, f)$ and $H_g(I_1\circ g_1, I_2\circ g_2, g_2^{-1}\circ f\circ g_1^{-1})$ are equal. By the definition of the GIT, this in turn shows that $\mathcal{J}((I_1, I_2), f) = \mathcal{J}((I_1\circ g_1, I_2\circ g_2), g_2^{-1}\circ f\circ g^{-1})$ completing the proof. □

9.4 Lack of Bias

Proposition 8 *Every GIT objective function is unbiased.*

Proof Suppose I_1, I_2 are constant images. Then for any $f \in \text{Diff}^+$, only a single bin (which contains the constant gray level values of the two images) has a non-zero histogram value. Thus $J((I_1, I_2), f) = \Phi(H_g(I_1, I_2, f))$, which is a constant independent of f, showing that the GIT is unbiased. □

Putting the results of Propositions 5, 6, 7 and 8 together, we have

Theorem 3 *Every GIT objective function is positive semi-definite, symmetric, invariant under SDiff$^+$, and unbiased.*

10 Regularization

As mentioned in Sect. 2.3, it is common to add a regularization term to the objective function to bias the warp towards desirable solutions. With the regularization term, the net objective function becomes

$$\mathcal{J}_{net}((I_1, I_2), f) = \mathcal{J}((I_1, I_2), f) + \lambda\mathcal{J}_{reg}(f),$$

where $\mathcal{J}$ is an objective function of the type we discussed in the paper so far, $\mathcal{J}_{reg}$ is the regularization term, and $\lambda > 0$ is the *regularization constant*.

Typically, regularization biases the registration towards smoother warps, where *smoothness* is defined using the square of the $\mathcal{L}_2$ norm of the first or higher-order derivatives of the warp. A commonly used regularization term, for example, is:

$$\mathcal{J}_{reg}(f) = \int_{\Omega_1} \sum_k (D^k f)^2 \omega, \tag{54}$$

where $k = (k_1, \ldots, k_n)$ is a multi-index with $k_i \geq 2$ for all i, and $D^k = \frac{\partial^n}{\partial x_{k_1}\ldots\partial x_{k_n}}$ is a partial derivative operator.

In a Bayesian setting, any regularization term is interpreted as the log of the prior likelihood of the warp. We adopt this interpretation and ask what properties we may require of the term without excessively limiting the choice of priors for a user. In particular, we ask whether the regularization term should be required to satisfy the four properties we required of the image dependent term. Consider each property in turn:

1. **Positive-definiteness:** A regularization term which is positive-definite would be zero only when the warp is the identity function. Such a term would bias the registration towards the zero function—a highly undesirable effect. Thus it is appropriate to relax this property and ask that the regularization term only be positive semi-definite so that it can be zero on a richer set of warps.

2. **Symmetry:** Symmetry of the regularization term should be required since we want the combined objective function to be symmetric.
3. **Invariance under SDiff$^+$:** Invariance under SDiff$^+$ can limit the choice of a prior significantly. Let $f : \Omega_1 \to \Omega_2$ be a warp and let $g : \Omega_2 \to \Omega_2$ be a volume preserving distortion of the image on Ω_2. Invariance under SDiff$^+$ would require that the value of the regularization term for $f \circ g$ be the same as the value for f. But it is possible to find $g \in \text{SDiff}^+$ such that $f \circ g$ is less smooth than f (smoothness measured by a functional such as in (54)), and it is no longer clear that invariance under SDiff$^+$ should be a required property.
4. **Lack of bias:** The whole point of the regularization term is to introduce a bias towards smoother warps. The resulting net objective function will vary with f even for constant images I_1, I_2. Hence we should not ask that the regularized objective function be unbiased in the sense defined in Sect. 3.

Therefore we only require that the regularization term be positive semi-definite and symmetric. We present some useful positive semi-definite and symmetric regularization terms below. This is not an exhaustive list, but it does contain terms that are either commonly used or are symmetrized versions of terms that are commonly used:

1. Regularization terms which are symmetric versions of the objective function of (54):

$$\mathcal{J}_{reg}(f) = \int_{\Omega_1} \sum_k (D^k f)^2 \omega + \int_{\Omega_2} \sum_k (D^k f^{-1})^2 \omega.$$

   It is easy to check that this objective function is positive semi-definite and symmetric. One drawback of this term is that it requires the evaluation of the inverse of f.
2. A regularization term which does not require evaluation of the inverse function is proposed in [31]:

$$\mathcal{J}_{reg}(f) = \int_{\Omega_1} (\det J_f - 1) \log(\det J_f)\, \omega.$$

   It is easy to check that this objective function is positive semi-definite and symmetric. Positive semi-definiteness follows from the property that both $x - 1$ and $\log x$ have the same sign (positive if $x > 1$ and negative if $x < 1$).
3. Often the warps f are restricted to be B-splines or piecewise-linear functions that use control points or node points. It is possible to regularize such warps by just using functions of the control/node points. One example is discussed in Sect. 11.3 in detail.

11 Numerical Symmetry

11.1 Discretization

In practice, the objective function $\mathcal{J}_{net}((I_1, I_2), f)$ has to be discretized for numerical calculation. In this section, we show how a commonly used discretization scheme can introduce asymmetry and how this can be fixed.

11.2 $\mathcal{GL}_2$ Objective Functions

A commonly used discretization procedure can be applied to the $\mathcal{GL}_2$ objective function $\mathcal{J}((I_1, I_2), f) = \int_{\omega_1} (I_1 - I_2 \circ f)^2 (1 + \det J_f)\omega$ in two steps:

1. Discretize f: Express f in terms of a finite basis as $f = \sum_i a_i \phi_i$, where ϕ_i are the basis functions to be defined later, and a_i are the coefficients.
2. Discretize the integral: Replace the integral with a finite sum taken on a uniform grid in Ω_1. We will call this grid the integration grid.

The discretized objective function is

$$\mathcal{J}\left((I_1, I_2), \sum_i a_i \phi_i\right) = \sum_k \left\{ I_1(x_k) - \left(I_2 \circ \sum_i a_i \phi_i\right)(x_k) \right\}^2 \times (1 + \det J_{\sum a_i \phi_i}(x_k))\Delta,$$

where, x_k are the uniform integration grid points in Ω_1 and Δ is the volume of a single grid element.

This objective function is asymmetric, and there are two causes of asymmetry:

1. The inverse of $f = \sum_i a_i \phi_i$ may not lie in the span of $\{\phi_i\}$, so that minimizing $\mathcal{J}((I_1, I_2), \sum_i a_i \phi_i)$ with respect to a_i may not produce the inverse of the function minimizing $\mathcal{J}((I_2, I_1), \sum_i a_i \phi_i)$ with respect to a_i.
2. During discretized integration, the image I_1 is sampled on the uniform integration grid but I_2 is not. On the other hand, while calculating $\mathcal{J}((I_2, I_1), \sum_i a_i \phi_i)$, the image I_2 is sampled on a uniform integration grid (since I_2 is now defined on Ω_1), but I_1 is not.

To eliminate the above asymmetries we adopt the following strategies:

1. We approximate f by a piecewise linear map defined on a grid as shown in Fig. 7. The grid is a triangular grid formed by taking a square grid and connecting the same diagonal in every square. In each triangle, the function f is approximated by a linear function that maps the triangle in Ω_1 to the corresponding triangle in Ω_2 (vertices go to vertices). The function f is changed by

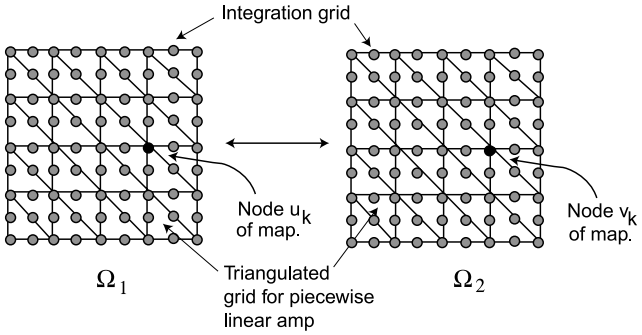


Fig. 7 Piecewise linear map

moving the nodes of the grid in Ω_1 as well as Ω_2. Notationally, if $u_1, \ldots, u_n$ are nodes in Ω_1 and $v_1, \ldots, v_n$ are nodes in Ω_2, then $f_{u_1,\ldots,u_n,v_1,\ldots,v_n}$ is a piecewise linear diffeomorphism form Ω_1 to Ω_2. The inverse of f is the function obtained by swapping the nodes of the grid in Ω_1 with the nodes of the grid in Ω_2. That is $f^{-1}_{u_1,\ldots,u_n,v_1,\ldots,v_n} = f_{v_1,\ldots,v_n,u_1,\ldots,u_n}$.

2. To remove asymmetry due to the integration grid, we impose two integration grids, one on Ω_1 and the other on Ω_2 (Fig. 7). Suppose x_k are node points of the grid in Ω_1 and y_k are corresponding node points of the grid in Ω_2, then we take

$$\begin{aligned}
&\mathcal{J}((I_1, I_2), f_{u_1,\ldots,u_n,v_1,\ldots,v_n}) \\
&\quad = 1/2\Bigg[\sum_k \{I_1(x_k) - I_2 \circ f(x_k)\}^2 (1 + \det J_f(x_k)) \\
&\qquad + \sum_k \left\{I_1 \circ f^{-1}(y_k) - I_2(y_k)\right\}^2 \\
&\qquad \times (1 + \det J_f(f^{-1}(y_k)))\Bigg]\Delta.
\end{aligned}$$

That is to say that we sample the integrand first on the uniform grid in Ω_1, then on the uniform grid in Ω_2 and take their mean. This discretization is symmetric (as can be easily verified).

11.3 Regularization

We use a simple regularizer which is the energy in the 2nd finite difference of the node points. For 2-D problems, is given by

$$\begin{aligned}
&\mathcal{J}_{reg}(f_{u_1,\ldots,u_n,v_1,\ldots,v_n}) \\
&\quad = \sum_k \|\acute{u}_k - 2u_k + \grave{u}_k\|^2 + \|\hat{u}_k - 2u_k + \check{u}_k\|^2 \\
&\qquad + \|\acute{v}_k - 2v_k + \grave{v}_k\|^2 + \|\hat{v}_k - 2v_k + \check{v}_k\|^2,
\end{aligned}$$

where, $\|\cdots\|^2$ is the square of the Euclidean norm, the superscript symbols $\hat{}$, $\check{}$, $\acute{}$, $\grave{}$ refer to the node above, below, to the left of, and to the right of the node under the sign, and the sum is over all non-boundary nodes of the grid. The 3-D version is similar.

Thus, the actual objective function being minimized is

$$\begin{aligned}
&\mathcal{J}_{net}((I_1, I_2), f_{u_1,\ldots,u_n,v_1,\ldots,v_n}) \\
&\quad = \mathcal{J}((I_1, I_2), f_{u_1,\ldots,u_n,v_1,\ldots,v_n}) + \lambda \mathcal{J}_{reg}(f_{u_1,\ldots,u_n,v_1,\ldots,v_n}),
\end{aligned}$$

where λ is a very small positive regularization constant. The minimization is with respect to the nodes $u_1, \ldots, u_n, v_1, \ldots, v_n$.

11.4 GIT Objective Functions

The GIT objective function is discretized in the same way. The function f is approximated as a piecewise linear map from the triangular grid in Ω_1 to the triangular grid in Ω_2. Two histograms are created—first by sampling Ω_1 on a uniform grid, and then sampling Ω_2 on a uniform grid. During the sampling, the joint intensity $(I_1(x_k), I_2 \circ f(x_k))$ (or $(I_1 \circ f^{-1}(x_k), I_2(x_k))$ for the Ω_2 grid) indicates the bin to which a histogram value of $(1 + \det J_f(x_k))\Delta$ is added. After both histograms are formed, their average is taken as a symmetric histogram and the objective function calculated from it.

Internal energy is added as above and the minimum is sought with respect to $u_1, \ldots, u_n, v_1, \ldots, v_n$.

11.5 Optimization Strategy

We use a simple paired-node gradient descent strategy where the objective function $J_{net}((I_1, I_2), f_{u_1,\ldots,u_n,v_1,\ldots,v_n})$ is minimized iteratively. In each iteration the function is minimized with respect to one node pair u_k, v_k at a time, for $k = 1, \ldots, n$. The minimization is carried out via constrained gradient descent where the constraint maintains a positive determinant. For the triangular mesh this constraint does not allow the triangles to flip orientation. In practice we find that the constraint is almost always satisfied, and thus the numerical procedure is equivalent to gradient descent on node pairs.

To speed up the performance, we use a multi-resolution strategy. The grid defining the piecewise linear f is initially chosen to be coarse. After the objective function is minimized with the coarse grid, the grid is made finer and another iteration of the minimization carried. The process stops at a pre-determined level of fineness. The multi-resolution grid sizes used in all experiments are 4×4, 5×5, 7×7, 10×10, 14×14.

12 Experiments

We now report a series of experiments about the performance of the numerical algorithm. This numerical algorithm

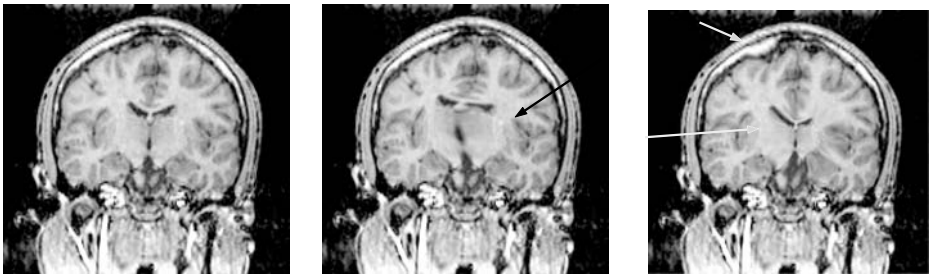


Fig. 8 Original and warped images

(a) Original

(b) Warped by first diffeo. in Table 1.

(c) Warped by second diffeo. in Table 1.

Table 1 Diffeomorphisms used in simulations

Diffeomorphism	Region of Support of non-linearity	Definition
First	$R_1 = \text{rect}((50, 68), (110, 118))$	$f(x,y) = \begin{cases} (x + 10\sin(\pi u)\cos(\pi v)\cos(\pi u), \\ y + 15\sin(\pi u)\cos(\pi v)\sin(2\pi v)) & \text{if } (x,y) \in R_1. \\ (x,y) & \text{otherwise.} \end{cases}$ where $u = (x-50)/60,\ v = (y-68)/50.$
Second	$R_1 = \text{rect}((60, 80), (120, 140))$ $R_2 = \text{rect}((20, 70), (50, 100))$	$f(x,y) = \begin{cases} (x + 15\phi((x-90)/r), \\ y + 15\phi((y-110)/r)) & \text{if } (x,y) \in R_1. \\ (x + 10\sin(\pi u)\cos(\pi v)\cos(\pi u), \\ y + 15\sin(\pi u)\cos(\pi v)\sin(2\pi v)) & \text{if } (x,y) \in R_2. \\ (x,y) & \text{otherwise.} \end{cases}$ where $r = \sqrt{(x-90)^2 + (y-110)^2}$, $\phi(u) = \min(u, 1-u)$, $u = (x-20)/30,\ v = (y-70)/30.$

and some of its results (esp. regarding the effect of symmetrization on registration accuracy) have not been reported before.

The first aim of the experiments was to test whether the new objective functions give symmetric registrations in practice. A second aim was to investigate whether there is any change in accuracy of registration when the new objective functions are used. For simplicity, all experiments were carried out on 2-D images.

A coronal slice through a 3-D brain MR image was selected as the undistorted image. The slice wask 180×180 pixels with an in-plane resolution of approximately 1 mm $\times$ 1 mm. The undistorted image was taken to be $I_2 : \Omega_2 \to \mathbb{R}$, and distorted to produce : $I_1 = I_2 \circ f$, where $f : \Omega_1 \to \Omega_2$ is the diffeomorphism. Two different diffeomorphisms were used to distort the image. Both diffeomorphisms were the identity function outside a region of support. Within the region of support (which was a rectangle R_1 for the first diffeomorphism and a union of rectangles R_1, R_2 for the second) the diffeomorphisms were non-linear. The diffeomorphisms are completely specified in Table 1. The original image and the two warped images are shown in Fig. 8. The arrows in Figs. 8b and c show the visual distortion produced by the diffeomorphisms. For future reference, we note that the maximum deviation of the first diffeomorphism from identity (i.e. $\max_{(x,y)} \|f(x,y) - (x,y)\|$) was 13.02 pixels, and of the second diffeomorphism was 7.50 pixels in R_1 and 7.51 pixels in R_2. We also measured the condition number of the Jacobian matrix of each diffeomorphism as a measure of shear. The condition number for translation, rotation and scaling is 1, so that deviation from 1 suggests the degree of shear in the non-linearity. Evaluated on a half pixel $\times$ half pixel grid, the median and maximum condition numbers for the first diffeomorphism were 1.87 and 9.05×10^3. The median and maximum condition numbers for the second diffeomorphism were 1.71 and 2.23×10^4 and 2.63 and 1.36×10^4 in regions R_1 and R_2 respectively. This shows that not only were the diffeomorphisms different from pure translation, rotation and scaling in most pixels, but also that the amount of shear varied considerably demonstrating significant non-linearity.

Table 2 Asymmetry for the first diffeomorphism measured in R_1

(a) Asymmetry in $\mathcal{L}_2$ and $\mathcal{GL}_2$

Noise σ	Quantity	$\mathcal{L}_2$		$\mathcal{GL}_2$	
		r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	$\hat{f}_2 \circ \hat{f}_1 - id$	1.0842	3.1823	1.7970×10^{-11}	1.7482×10^{-10}
	$\hat{f}_1 \circ \hat{f}_2 - id$	1.1985	3.7705	1.2071×10^{-11}	1.0550×10^{-10}
3	$\hat{f}_2 \circ \hat{f}_1 - id$	0.9645	2.7623	1.4214×10^{-13}	1.026×10^{-12}
	$\hat{f}_2 \circ \hat{f}_1 - id$	1.0665	3.2573	1.303×10^{-13}	8.7705×10^{-13}
6	$\hat{f}_2 \circ \hat{f}_1 - id$	1.2373	4.0014	8.5038×10^{-14}	5.0568×10^{-13}
	$\hat{f}_1 \circ \hat{f}_2 - id$	1.3972	4.4755	7.1633×10^{-14}	4.743×10^{-13}
9	$\hat{f}_2 \circ \hat{f}_1 - id$	1.3249	4.1363	3.4919×10^{-11}	3.7496×10^{-10}
	$\hat{f}_1 \circ \hat{f}_2 - id$	1.4465	4.3220	3.321×10^{-11}	3.338×10^{-11}

(b) Asymmetry in IT and GIT

Noise σ	Quantity	IT		GIT	
		r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	$\hat{f}_2 \circ \hat{f}_1 - id$	0.4688	1.9006	1.2906×10^{-12}	3.9278×10^{-12}
	$\hat{f}_1 \circ \hat{f}_2 - id$	0.4496	1.7978	1.3978×10^{-12}	5.0588×10^{-12}
3	$\hat{f}_2 \circ \hat{f}_1 - id$	0.4901	1.7043	1.4507×10^{-12}	5.7096×10^{-12}
	$\hat{f}_2 \circ \hat{f}_1 - id$	0.4697	1.7451	1.5207×10^{-12}	5.4017×10^{-12}
6	$\hat{f}_2 \circ \hat{f}_1 - id$	0.3968	1.3650	1.6730×10^{-12}	4.6989×10^{-12}
	$\hat{f}_1 \circ \hat{f}_2 - id$	0.4130	1.6032	1.8203×10^{-12}	5.2544×10^{-12}
9	$\hat{f}_2 \circ \hat{f}_1 - id$	0.3927	1.7353	1.8158×10^{-12}	7.4788×10^{-12}
	$\hat{f}_1 \circ \hat{f}_2 - id$	0.4220	1.7378	1.7654×10^{-12}	6.6807×10^{-12}

Zero-mean Gaussian noise was added to the image pair I_1, I_2 to produce noisy image pairs. The standard deviation of the additive noise was set to $\sigma = 0, 3, 6,$ and 9 respectively. Thus, there were 8 image pairs (2 diffeomorphisms $\times$ 4 standard deviations).

Each pair of noisy images was registered using the old non-symmetric $\mathcal{L}_2$ and IT objective functions and the new symmetric $\mathcal{GL}_2$ and GIT objective functions. This calculated diffeomorphism is denoted below by $\hat{f}_1$. The two images were swapped and registered again. This diffeomorphism is denoted $\hat{f}_2$.

12.1 Symmetry

Ideally, $\hat{f}_2 = \hat{f}_1^{-1}$, so that $\hat{f}_2 \circ \hat{f}_1 = id$, and $\hat{f}_1 \circ \hat{f}_2 = id$. If $\hat{f}_2 \circ \hat{f}_1$ and $\hat{f}_1 \circ \hat{f}_2$ deviate from the identity map, the extent of the deviation measures the asymmetry in the registration algorithm. So we examined $\hat{f}_2 \circ \hat{f}_1 - id$ and $\hat{f}_1 \circ \hat{f}_2 - id$. We quickly found that almost all of the deviation from identity occurred in the region of support of the non-linearity, and all evaluations were done in this region.

We measured the r.m.s. value of the deviation from the identity function as

$$\sqrt{\frac{1}{N}\sum_k \|(\hat{f}_2 \circ \hat{f}_1)(x_k) - x_k\|^2} \quad \text{and}$$

$$\sqrt{\frac{1}{N}\sum_k \|(\hat{f}_1 \circ \hat{f}_2)(y_k) - y_k\|^2},$$

where, $\|\ \|$ is the Euclidean norm, x_k are all pixels in the first image that belong to region R_1 and/or R_2 as defined in Table 1, and y_k are all pixels in the second image such that $f_2(y_k)$ belonged to R_1 and/or R_2, and N is the total number of pixels summed over. We also measured the maximum deviation from the identity function as:

$$\max_k \|(\hat{f}_2 \circ \hat{f}_1)(x_k) - x_k\| \quad \text{and} \quad \max_k \|(\hat{f}_1 \circ \hat{f}_2)(y_k) - y_k\|.$$

The results are shown in Tables 2 and 3 for the first and second diffeomorphism. The 'a' part of the table shows the r.m.s. and maximum deviations from identity for $\mathcal{L}_2$ and $\mathcal{GL}_2$ objective functions and the 'b' part shows the r.m.s.

Table 3 Asymmetry for the second diffeomorphism

(a) Asymmetry in $\mathcal{L}_2$ and $\mathcal{GL}_2$

Noise σ	Region	Quantity	$\mathcal{L}_2$		$\mathcal{GL}_2$	
			r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	0.8542	4.4492	3.708×10^{-13}	2.142×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.9492	3.2595	4.207×10^{-13}	2.530×10^{-12}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.6321	2.4743	1.387×10^{-13}	6.733×10^{-13}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.7027	2.6951	1.786×10^{-13}	9.505×10^{-13}
3	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	1.3887	4.0617	4.692×10^{-13}	4.500×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.5375	4.3940	5.160×10^{-13}	4.692×10^{-13}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	1.0609	4.0617	5.330×10^{-11}	4.306×10^{-10}
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.2519	4.3940	5.308×10^{-11}	4.194×10^{-10}
6	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	0.9195	4.3272	1.275×10^{-12}	1.0612×10^{-11}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.9534	2.8894	1.487×10^{-12}	1.288×10^{-11}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.6674	2.6441	1.410×10^{-12}	1.061×10^{-11}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.7414	2.8894	1.594×10^{-12}	1.288×10^{-11}
9	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	1.9854	4.0770	8.209×10^{-14}	5.736×10^{-13}
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.1553	3.5180	8.752×10^{-14}	6.751×10^{-13}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.7832	2.1569	1.155×10^{-13}	8.485×10^{-13}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.8547	2.5514	1.163×10^{-13}	8.586×10^{-13}

(b) Asymmetry in IT and GIT

Noise σ	Region	Quantity	IT		GIT	
			r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	0.3174	1.1210	1.208×10^{-12}	4.514×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.3075	1.1510	1.221×10^{-12}	5.169×10^{-12}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.2399	1.1572	1.025×10^{-12}	4.762×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.2450	1.1510	1.015×10^{-12}	4.025×10^{-12}
3	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	0.3278	1.1410	2.637×10^{-12}	1.167×10^{-11}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.3153	1.1070	2.550×10^{-12}	1.153×10^{-11}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.2438	1.2453	4.105×10^{-12}	3.110×10^{-11}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.2515	1.1357	4.057×10^{-12}	3.006×10^{-11}
6	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	0.2953	1.3482	1.595×10^{-12}	4.819×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.2832	1.0547	1.492×10^{-12}	4.389×10^{-12}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.2579	1.1152	1.781×10^{-12}	6.036×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.2571	1.0230	1.827×10^{-12}	5.798×10^{-12}
9	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	0.4220	1.5074	3.979×10^{-12}	2.505×10^{-11}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.4344	1.6539	4.251×10^{-12}	2.566×10^{-11}
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	0.3492	1.1180	1.747×10^{-12}	8.988×10^{-12}
		$\hat{f}_1 \circ \hat{f}_2 - id$	0.3634	1.3462	1.855×10^{-12}	1.001×10^{-11}

Table 4 Accuracy for the first diffeomorphism measured in R_1

(a) Accuracy of $\mathcal{L}_2$ and $\mathcal{GL}_2$

Noise σ	Quantity	$\mathcal{L}_2$		$\mathcal{GL}_2$	
		r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	$\hat{f}_1 - f$	1.4613	3.7473	1.0680	2.6929
	$\hat{f}_2 - f^{-1}$	2.0131	5.1020	0.9490	2.5975
3	$\hat{f}_1 - f$	1.7639	4.3061	1.3512	3.1772
	$\hat{f}_2 - f^{-1}$	2.2762	5.3790	1.2460	3.2098
6	$\hat{f}_1 - f$	1.8887	4.4614	1.7260	4.2699
	$\hat{f}_2 - f^{-1}$	2.8408	6.5554	1.6724	4.7475
9	$\hat{f}_1 - f$	1.6952	4.2519	1.7177	4.0022
	$\hat{f}_2 - f^{-1}$	1.6110	4.1891	2.6415	5.8287

(b) Accuracy of IT and GIT

Noise σ	Quantity	IT		GIT	
		r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	$\hat{f}_1 - f$	1.5212	6.9750	1.1954	4.9140
	$\hat{f}_2 - f^{-1}$	1.5152	7.0146	1.1280	5.3109
3	$\hat{f}_1 - f$	1.4980	6.7150	1.1931	4.9996
	$\hat{f}_2 - f^{-1}$	1.4951	6.8243	1.1112	5.1781
6	$\hat{f}_1 - f$	1.7127	7.8179	1.2710	5.1473
	$\hat{f}_2 - f^{-1}$	1.6429	7.4669	1.2134	5.8420
9	$\hat{f}_1 - f$	1.7127	7.8951	1.4516	6.3953
	$\hat{f}_2 - f^{-1}$	1.5607	7.3623	1.3772	6.3409

and maximum deviations for IT and GIT objective functions. Since there is only one region of support for the nonlinearity in the first diffeomorphism, the region of support is not shown explicitly in Table 2. There are two regions of support for the second diffeomorphism and the measured values for both are shown separately in Table 3.

Several conclusions can be drawn from these two tables:

1. The forward ($\hat{f}_2 \circ \hat{f}_1 - id$) and backward ($\hat{f}_1 \circ \hat{f}_2 - id$) asymmetry errors are of the same magnitude, suggesting that the asymmetry measures are not sensitive to the direction in which they are measured.
2. The r.m.s. and the maximum asymmetry follow similar trends, but the latter is significantly higher. We focus on the maximum asymmetry since it is the "worst case" measurement.
3. For the first diffeomorphism (Table 2), the maximum asymmetry of $\mathcal{L}_2$ is quite significant—it is approximately 4.47 pixels. This amount is comparable to 13.02 pixels which is the maximum deviation of the first diffeomorphism from the identity. The asymmetry in the IT objective function (Table 2b) is not as significant, being of the order of 2 pixels.
4. Compared to $\mathcal{L}_2$ and IT, the $\mathcal{GL}_2$ and GIT objective functions reduce the maximum asymmetry to the range 10^{-10}–10^{-14} pixels. Essentially, the asymmetry has been reduced to machine precision.
5. The same trends are apparent in Table 3. The maximum asymmetry in Table 3 for $\mathcal{L}_2$ is comparable to the maximum deviation of the diffeomorphism from identity. Recall that the maximum deviation from identity in R_1 is 7.50 pixels and in R_2 is 7.51 pixels.
6. The effectiveness of the symmetrized objective functions in eliminating asymmetry is quite clear from both tables.

12.2 Accuracy

We also measured the accuracy of registration function. Ideally $\hat{f}_1, \hat{f}_2$ should be equal to the true diffeomorphism f and its inverse f^{-1} respectively. We measured the r.m.s. and maximum values of $\|f(x_k) - \hat{f}_1(x_k)\|$ for x_k in R_1 and R_2 and the r.m.s. and maximum values of $\|f^{-1}(y_k) - \hat{f}_2(y_k)\|$ for $f^{-1}(y_k)$ in R_1 and R_2. The measured values are shown

Table 5 Accuracy for the second diffeomorphism

(a) Accuracy of $\mathcal{L}_2$ and $\mathcal{GL}_2$

Noise σ	Region	Quantity	$\mathcal{L}_2$		$\mathcal{GL}_2$	
			r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	1.3909	4.5962	1.1366	4.6232
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.2989	4.4115	0.8715	3.3903
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	1.2819	4.5164	1.0273	4.1530
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.1682	3.6940	0.9314	3.3887
3	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	2.0615	6.3711	1.2600	4.2367
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.4005	4.4261	1.0184	3.3180
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	1.8208	6.3711	1.1585	4.0111
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.3141	4.1485	1.0628	3.6667
6	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	1.8017	5.1628	1.5066	4.8696
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.6729	4.6270	1.2566	3.9152
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	1.5317	5.0180	1.2390	4.2938
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.4813	4.3843	1.1531	3.8387
9	R_1	$\hat{f}_2 \circ \hat{f}_1 - id$	1.9802	5.8152	1.6156	5.6171
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.7049	5.0381	1.3284	3.9268
	R_2	$\hat{f}_2 \circ \hat{f}_1 - id$	1.6245	5.3238	1.4930	4.7008
		$\hat{f}_1 \circ \hat{f}_2 - id$	1.6038	4.7218	1.3815	4.0747

(b) Accuracy of IT and GIT

Noise σ	Region	Quantity	IT		GIT	
			r.m.s. (pixels)	max. (pixels)	r.m.s. (pixels)	max. (pixels)
0	R_1	$\hat{f}_1 - f$	2.2502	6.0380	1.692	5.535
		$\hat{f}_2 - f^{-1}$	2.1191	5.8461	1.468	5.014
	R_2	$\hat{f}_1 - f$	1.7420	5.7271	1.555	5.464
		$\hat{f}_2 - f^{-1}$	1.5942	4.9774	1.394	4.680
3	R_1	$\hat{f}_1 - f$	2.2575	6.0380	1.785	5.552
		$\hat{f}_2 - f^{-1}$	2.1641	5.8475	1.566	4.966
	R_2	$\hat{f}_1 - f$	1.7719	5.7313	1.621	5.542
		$\hat{f}_2 - f^{-1}$	1.6500	5.0248	1.458	4.750
6	R_1	$\hat{f}_1 - f$	2.2817	5.8245	1.824	5.457
		$\hat{f}_2 - f^{-1}$	2.2199	6.0453	1.667	4.964
	R_2	$\hat{f}_1 - f$	1.7954	5.5162	1.560	5.289
		$\hat{f}_2 - f^{-1}$	1.6932	5.0805	1.419	4.525
9	R_1	$\hat{f}_1 - f$	2.2100	2.1045	1.740	1.632
		$\hat{f}_2 - f^{-1}$	2.1045	5.8932	1.632	5.235
	R_2	$\hat{f}_1 - f$	1.7323	5.4000	1.424	4.842
		$\hat{f}_2 - f^{-1}$	1.5817	4.9428	1.292	4.266

in Tables 4–5. The following observations can be made about these measurements:

1. The maximum error is significantly higher than the r.m.s. in the forward and the backward directions. As with asymmetry, we will concentrate on the maximum error.
2. The maximum error (as well as the r.m.s. error) for the symmetrized objective functions is generally lower than the maximum error for the original objective functions. The only exception is the first row of Table 5a. (however, the r.m.s. error has decreased in this row). The increase is quite small, being approximately 0.6%.

   The reason for this increase is not clear. One possibility is that the optimization algorithm is trapped in a shallow local minimum, and this local minimum disappears upon adding noise to the image (this might explain why no increase is observed for the noisy images).

Thus it appears that replacing the old objective functions with new ones does not lead to a poorer performance and might actually improve the accuracy of the registration.

13 Conclusions

In this report we analyzed the asymmetry of popular registration objective functions. We showed that the $\mathcal{L}_2$ and IT objective functions were asymmetric and that their asymmetry arose from the use of Euclidean volume forms. We also showed that there is a unique tangent-dependent volume form on the graph of the diffeomorphism that is positive-definite, symmetric, invariant under SDiff^+, and makes the objective function unbiased. This volume form symmetrizes both objective functions. We introduced the notion of numerical symmetry and showed how the objective functions can be numerically symmetrized. Experimental results show that the new objective functions reduce asymmetry to the range 10^{-14} to 10^{-10} pixels.

Appendix: Proof of Lemma 2

Lemma 2 *With notation as above*,

$$\frac{d}{dt}\int_{\Xi_2} \tilde{\zeta}\circ\left[\det J_{\Phi_t}.r_f\right]\ (\Phi_t)^*\omega\,|_{t=0} = -\int_{\Xi_2}\langle\zeta''(r_f)dr_f, X\rangle\omega. \tag{55}$$

Proof To differentiate with respect to t, first recall that the derivative of the determinant function of $n\times n$ matrices at the identity is

$$D(\det)|_I(A) = \mathrm{tr}(A), \tag{56}$$

where A, the direction of differentiation, is an arbitrary $n\times n$ matrix. Thus

$$\begin{aligned}
&\frac{d}{dt}\left(\tilde{\zeta}\circ[r_f\cdot\det(J_{\Phi_t})]\right)\Big|_{t=0}\\
&\quad=(\tilde{\zeta}'\circ r_f)r_f\ \frac{d}{dt}\det(J_{\Phi_t})\Big|_{t=0}\\
&\quad=(\tilde{\zeta}'\circ r_f)r_f\ D(\det)|_I\left(\frac{d}{dt}J_{\Phi_t}\Big|_{t=0}\right)\\
&\quad=(\tilde{\zeta}'\circ r_f)r_f\ \mathrm{tr}\left(J_{\frac{d}{dt}\Phi_t|_{t=0}}\right)\\
&\quad=(\tilde{\zeta}'\circ r_f)r_f\ \mathrm{tr}\,(J_X)\,,
\end{aligned}$$

where

$$D(\det)|_I\left(\frac{d}{dt}J_{\Phi_t}\Big|_{t=0}\right)=\mathrm{tr}\left(J_{\frac{d}{dt}\Phi_t|_{t=0}}\right),$$

because derivatives used in the Jacobian commute with derivative with respect to t.

Next, letting $\mathcal{L}_X$ denote Lie derivative by the vector field X, recall that

$$\frac{d}{dt}\Phi_t^*\omega\Big|_{t=0}=\mathcal{L}_X\omega=\mathrm{div}(X)\ \omega=\mathrm{tr}(J_X)\ \omega, \tag{57}$$

and also that

$$\mathcal{L}_X\omega=d(i_X\omega), \tag{58}$$

where $i_X\omega$ denotes contraction of the vector field X into the differential form ω.

Combining (57)–(58), we therefore have

$$\begin{aligned}
&\frac{d}{dt}\left[\left(\tilde{\zeta}\circ[r_f.\det(J_{\Phi_t})\right]\right)\Phi_t^*\omega\Big|_{t=0}\\
&\quad=\left((\tilde{\zeta}'\circ r_f)r_f+\tilde{\zeta}\circ r_f\right)\mathrm{div}(X)\omega\\
&\quad=\left((r\tilde{\zeta}(r))'|_{r=r_f}\right)d(i_X\omega)\\
&\quad=(\zeta'\circ r_f)\ d(i_X\omega).
\end{aligned}\tag{59}$$

Since $X=0$ on $\partial\Xi_2$, using Stokes' Theorem we find

$$\begin{aligned}
&\frac{d}{dt}\int_{\Xi_2}\left(\tilde{\zeta}\circ[r_f\cdot\det(J_{\Phi_t})]\right)\Phi_t^*\omega\Big|_{t=0}\\
&\quad=\int_{\Xi_2}\frac{d}{dt}\left[\left(\tilde{\zeta}\circ[r_f\cdot\det(J_{\Phi_t})]\right)\ \Phi_t^*\omega\right]_{t=0}\\
&\quad=\int_{\Xi_2}(\zeta'\circ r_f)\ d(i_X\omega)\\
&\quad=-\int_{\Xi_2}d(\zeta'\circ r_f)\wedge(i_X\omega)
\end{aligned}$$

$$= -\int_{\Xi_2} (\zeta'' \circ r_f)\, dr_f \wedge (i_X \omega)$$

$$= -\int_{\Xi_2} (\zeta'' \circ r_f) \langle dr_f, X \rangle \omega, \tag{60}$$

since $dr_f \wedge \omega = 0$ implies

$$0 = i_X(dr_f \wedge \omega) = (i_X dr_f) \wedge \omega - dr_f \wedge i_X \omega)$$
$$= \langle dr_f, X \rangle \omega - dr_f \wedge (i_X \omega).$$

Equation (22) now follows from (60). This completes the proof of the lemma. □

References

1. Hajnal, J.V., Hill, D.L.G., Hawkes, D.J.: Medical Image Registration. CRC Press, Boca Raton (2001)
2. Toga, A.: Brain Warping. Academic Press, San Diego (1999)
3. Goshtaby, A.A.: 2-D and 3-D Image Registration for Medical Remote Sensing and Industrial Applications. Wiley, Hoboken (2005)
4. Maintz, J.B.A., Viergever, M.A.: A survey of image registration. Med. Image Anal. MEDIA **2**(1), 1–36 (1998)
5. Zitová, B., Flusser, J.: Image registration methods: a survey. Image Vis. Comput. **21**(11), 977–1000 (2003)
6. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual information based registration of medical images: a survey. IEEE Trans. Med. Imag. **22**(8), 986–1004 (2003)
7. Christensen, G.E., Johnson, H.J.: Consistent image registration. IEEE Trans. Med. Imag. **20**(7), 568–582 (2001)
8. Johnson, H.J., Christensen, G.E.: Consistent landmark and intensity-based image registration. IEEE Trans. Med. Imag. **21**(5), 450–461 (2002)
9. Christensen, G.E., Johnson, H.J.: Invertibility and transitivity analysis for non-rigid image registration. J. Electron. Imag. **12**(1), 106–117 (2003)
10. Ashburner, J., Andersson, J.L.R., Friston, K.J.: High-dimensional image registration using symmetric prior. NeuroImage **9**(6), 619–628 (1999)
11. Beg, M.F., Khan, A.: Symmetric data attachment terms for large deformation image registration. IEEE Trans. Med. Imag. **26**(9), 1179–1189 (2007)
12. Cachier, P., Rey, D.: Symmetrization of the non-rigid registration problem using inversion-invariant energies: application to multiple sclerosis. In: Medical Image Computing and Computer Aided Intervention (MICCAI), Pittsburgh, PA, October 2000, pp. 472–481
13. Rogelj, P., Kovacic, S.: Symmetric image registration. Med. Image Anal. **10**(3), 484–493 (2006)
14. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23**(1), S151–S160 (2004). Supplement issue on Mathematics in Brain Imaging
15. Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., Joshi, S.: Multi-modal image set registration and atlas formation. Med. Image Anal. MEDIA **10**(3), 440–451 (2006)
16. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)
17. Tagare, H.D., O'Shea, D., Groisser, D.: Shape based non-rigid correspondence for plane curves. J. Math. Imag. Vis. **16**, 57–68 (2002)
18. Tagare, H.D.: Shape-based non-rigid correspondence with applications to heart motion analysis. IEEE Trans. Med. Imag. **18**(7), 570–579 (1999)
19. Lorenzen, P., Davis, B., Joshi, S.: Model based symmetric information theoretic large deformation multi-model image registration. In: Intl. Symp. Biomedical Imaging, pp. 720–723 (2004)
20. Lorenzen, P., Davis, B., Joshi, S.: Unbiased atlas formation via large deformations metric mapping. In: J.S. Duncan, G. Gerig (eds.) Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 411–418 (2005)
21. Magnotta, V.A., Bockholt, H.J., Johnson, H.J., Christensen, G.E., Andreasen, N.C.: Subcortical, cerebellar and MR based consistent brain image registration. NeuroImage **19**(2), 233–245 (2003)
22. Christensen, G.E., Johnson, H.J., Vannier, M.W.: Synthesizing average 3D anatomical shapes. NeuroImage **32**, 146–158 (2006)
23. Li, B., Christensen, G.E., McLennan, G., Hoffman, E.A., Reinhardt, J.M.: Establishing a normative atlas of the human lung: Inter-subject warping and registration of volumetric CT. Acad. Radiology **10**(3), 255–265 (2003)
24. Lu, W., Parikh, P., El Naqa, I., Nystrom, M., Hubenschmidt, J., Wahab, S., Mutic, S., Sing, A., Christensen, G.E., Bradley, J.D., Low, D.A.: Quantitation of the four-dimensional computed tomography process for lung cancer patients. Med. Phys. **32**(4), 890–901 (2005)
25. Christensen, G.E., He, J., Dill, J.A., Rubinstein, J.T., Vannier, M.W., Wang, G.: Automatic measurement of the labyrinth using image registration and a deformable inner ear atlas. Acad. Radiology **10**(9), 988–999 (2003)
26. Wells, W.M. III, Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Med. Image Anal. **1**(1), 35–51 (1996)
27. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. **32**(1), 71–86 (1999)
28. Viola, P., Wells, W.M. III: Alignment by maximization of mutual information. Int. J. Comput. Vis. **24**(2), 137–154 (1997)
29. D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: A viscous fluid model for multimodal non-rigid image registration using mutual information. Med. Image Anal. **7**, 565–575 (2003)
30. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: F-information measures in medical image registration. IEEE Trans. Med. Imag. **23**(12), 1508–1516 (2004)
31. Yanovsky, I., Thompson, P., Osher, S., Leow, A.: Topology preserving log-unbiased non-linear image registration: theory and implementation. In: IEEE Conf. on Computer Vision and Pattern Recognition, June 2007

Hemant D. Tagare is an Associate Professor in the Department of Diagnostic Radiology and the Department of Biomedical Engineering at Yale University. He received his Ph.D. in 1990 in Electrical Engineering from Rice Unviersity in Houston, TX. His research interest are in bio-medical image processing, especially the application of geometry and probability to image analysis problems.

David Groisser is an Associate Professor of Mathematics at the University of Florida. Dr. Groisser conducts research in pure and applied differential geometry. His current interests include the geometry of shape spaces and applications of differential geometry to image-analysis.

Oskar Skrinjar received the B.S. degree from University of Novi Sad, Novi Sad, Serbia, in 1996 and the M.S., M.Phil., and Ph.D. degrees from Yale University, New Haven, CT, in 1998, 1999, and 2002, respectively, all in electrical engineering. After receiving the Ph.D. degree on the subject of deformable models in image-guided neurosurgery, he joined the Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA where currently he is an assistant professor. His research interests span image processing, computer vision, computer graphics and mechanical modeling applied to biomedical problems. His major projects include cardiac image analysis, image-guided neurosurgery planning and navigation, and general image registration. Dr. Skrinjar serves on the Editorial Board of the Journal of Electronic Imaging as an Associate Editor for medical imaging.