Journal of Mathematical Research with Applications
Nov., 2019, Vol. 39, No. 6, pp. 633–656
DOI:10.3770/j.issn:2095-2651.2019.06.010
Http://jmre.dlut.edu.cn

On a New Diffeomorphic Multi-Modality Image Registration Model and Its Convergent Gauss-Newton Solver

Daoping ZHANG, Anis THELJANI, Ke CHEN*
EPSRC Liverpool Centre for Mathematics in Healthcare, Centre for Mathematical Imaging Techniques and Department of Mathematical Sciences, The University of Liverpool, Peach Street, Liverpool L69 7ZL, United Kingdom

Dedicated to the Memory of Professor L. C. HSU on the Occasion of His 100th Birthday

Abstract In this work, we propose a new variational model for multi-modal image registration and present an efficient numerical implementation. The model minimizes a new functional based on using reformulated normalized gradients of the images as the fidelity term and higher-order derivatives as the regularizer. A key feature of the model is its ability of guaranteeing a diffeomorphic transformation which is achieved by a control term motivated by the quasi-conformal map and Beltrami coefficient. The existence of the solution of this model is established. To solve the model numerically, we design a Gauss-Newton method to solve the resulting discrete optimization problem and prove its convergence; a multilevel technique is employed to speed up the initialization and avoid likely local minima of the underlying functional. Finally, numerical experiments demonstrate that this new model can deliver good performances for multi-modal image registration and simultaneously generate an accurate diffeomorphic transformation.

Keywords Multi-modal image registration; variational model; diffeomorphic transformation

MR(2010) Subject Classification 65K10; 68U10

1. Introduction

Image registration can be used to either find differences between a pair of similar images for aiding artificial intelligence or fuse complementary information from two related images to enhance imaging capability which is otherwise not possible with a single modality. It consists of finding a reasonable spatial geometric transformation between given two images of the same object taken at different times or acquired using different devices. It is a challenging task required in diverse fields such as astronomy, optics, biology, chemistry, life sciences, remote sensing and particularly in medical imaging. For an overview of image registration, methodologies and approaches, especially for registering images acquired by the same modality (e.g., CT-CT), we refer to [1–5]. For a more recent survey one can refer to [6]. Here our emphasis is on registering two images from different modalities (e.g., CT-MRI or digital-infrared).

Received August 1, 2019; Accepted October 28, 2019
* Corresponding author, Web: http://www.liv.ac.uk/∼cmchenke
E-mail addresses: dpzhang@math.cuhk.edu.hk (Daoping ZHANG); theljani@liverpool.ac.uk (Anis THELJANI); k.chen@liverpool.ac.uk (Ke CHEN)

The image registration problem can be described as follows: given a fixed image R (the reference) and a moving image T (the template), both represented by scalar function mappings over $\Omega \subset \mathbb{R}^d \longrightarrow \mathbb{R}$, find a suitable geometric transformation $\boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), ..., y_d(\boldsymbol{x})) : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that the deformed template $T(\boldsymbol{y})$ is similar to the reference R. In reality, according to specific applications, image registration can be classified into two categories: mono-modal registration and multi-modal registration. In this work, we focus on 2D deformable image registration models ($d = 2$) for multi-modal images using a variational approach.

In order to measure the difference between the deformed template $T(\boldsymbol{y})$ and the reference R, a fidelity term $\mathcal{D}(T(\boldsymbol{y}), R)$ should be defined. Minimizing this fidelity term alone is an ill-posed inverse problem and thus regularization techniques are needed to overcome this ill-posedness [7–14]. Generally speaking, a regularization technique turns an ill-posed problem into a well-posed optimization model:

$$\min_{\boldsymbol{u}\in\mathcal{H}} \mathcal{J}(\boldsymbol{u}) = \mathcal{S}(\boldsymbol{u}) + \lambda\mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R), \tag{1.1}$$

where the displacement $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{x}$ is a minimizer of this joint energy functional and λ is a positive parameter.

Here the first term, the regularizer, $\mathcal{S}(\boldsymbol{u})$ controls the smoothness of $\boldsymbol{u}$ and reflects our expectations in penalizing unlikely transformations. Various regularizers have been proposed, such as first-order derivatives-based total variation [15, 16], diffusion [17] and elastic [18] regularizers, and higher-order derivatives-based linear curvature [19], mean curvature [20] and Gaussian curvature [21] regularizers; refer also to [8, 13, 22–24].

The second term, the fidelity measure, $\mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R)$ quantifies the distance or similarity between the transformed template image $T(\boldsymbol{x}+\boldsymbol{u})$ and the reference R. For mono-modal registration, a widely-used data fidelity term is the sum of squared differences, defined by $\mathcal{D} = \|T(\boldsymbol{x}+\boldsymbol{u}) - R\|_2^2 \equiv \mathrm{SSD}(T(\boldsymbol{x}+\boldsymbol{u}), R)$ to measure the difference between the reference image R and the deformed template image $T(\boldsymbol{x}+\boldsymbol{u})$. However for multi-modality registration, the choice of $\mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R)$ is more challenging since the images are captured from different machineries and SSD no longer makes sense. Hence, how to design the right (or rather better) similarity measures that can identify the difference (in features, colors, gradients, illumination etc) between images from different modalities is a main issue. Various measures have been proposed and tested in the literature. Designing a measure which is based on the geometric information such as the gradients of the images is a good choice. See for instance, the normalized gradient field (**NGF**) [25–27], edges sketching registration [28], normalized gradient fitting [25, 29] and mutual information (**MI**) [30–32]. Recently [33] proposed a cross-correlation similarity measure based on reproducing kernel Hilbert spaces and found advantages over **MI**.

However, most of the models of type (1.1) are insufficient in ensuring that $\boldsymbol{y}(\boldsymbol{x})$ is a diffeomorphic map, as they only control the smoothness of transformation (see remarks and illustrations in [34, 35]). They do not exclude non-natural transformations such as folding or tearing between the two images, unless λ is small (small λ implying a poor registration fidelity error). Over the last decade, more and more researchers have focused on diffeomorphic image registration to

ensure that folding measured by the local invertibility quantity $\det(\nabla \boldsymbol{y})$ is reduced or avoided where $\det(\nabla \boldsymbol{y})$ is the Jacobian determinant of $\boldsymbol{y}$. Under desired assumptions, obtaining a one-to-one mapping is a natural choice [7,10,36,37]. There are several studies on model (1.1) in dealing with diffeomorphic maps for the mono-modal registration, but fewer theoretical or experimental studies for the multi-modal registration.

In this work, we aim to design a diffeomorphic registration model for multi-modal images. The formulation consists of minimizing a new functional based on reformulated normalized gradients of the images as the fidelity term [29], higher-order derivatives and a new Beltrami coefficient based term [38, 39]. An effective, iterative scheme is also presented and numerical results show that the new registration model has a good performance. An alternative approach to imposing a constraint is the so-called inverse consistent mapping in which both T to R and R to T are registered in the same model; see [34, 35, 40] and many references therein.

The rest of this work is organized as follows. In Section 2, we propose our new multi-modal registration model. In Section 3, we illustrate the solution algorithm. Numerical results are shown in Section 4, and finally a conclusion is summarized in Section 5.

2. The proposed variational model

We aim to design a variational diffeomorphic model to cope with a multi-modality image registration. Since (1.1) usually cannot guarantee a diffeomorphic registration, one direct method is to add a constraint and obtain the following constrained optimization model:

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathcal{H}} \mathcal{J}(\boldsymbol{u}) = \mathcal{S}(\boldsymbol{u}) + \lambda \mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R) \\ &\text{s.t. } \det(\nabla \boldsymbol{y}) > 0. \end{aligned} \tag{2.1}$$

Here the above constraint is highly nonlinear and is non-trivial to implement using any commonly used framework of optimization solvers. Below we shall discuss first how to deal with this constraint and then the choice of $\mathcal{S}, \mathcal{D}$ terms.

Our idea is reformulation of this constraint and an effective way is to control the Beltrami coefficient [39] instead of controlling the Jacobian determinant of the transformation directly. The Beltrami coefficient measurement is defined by

$$|\mu(\boldsymbol{y})|^2 = \frac{\|\nabla \boldsymbol{y}\|_F^2 - 2\det(\nabla \boldsymbol{y})}{\|\nabla \boldsymbol{y}\|_F^2 + 2\det(\nabla \boldsymbol{y})}. \tag{2.2}$$

From this equation, we can see the equivalence: $\det(\nabla \boldsymbol{y}) > 0 \iff |\mu(\boldsymbol{y})|^2 < 1$. The key point here is that the former constraint cannot be easily built into an unconstrained optimization but the latter can be easily built [39]. In this work we consider the following equivalent formulation to (2.1):

$$\begin{aligned} &\min_{\boldsymbol{u}\in\mathcal{H}} \mathcal{J}(\boldsymbol{u}) = \mathcal{S}(\boldsymbol{u}) + \lambda \mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R) \\ &\text{s.t. } |\mu(\boldsymbol{y})|^2 < 1. \end{aligned} \tag{2.3}$$

In order to solve (2.3), as remarked, we solve the following unconstrained optimization prob-

lem (note the notation change from (1.1) with λ for $\mathcal{D}$ replaced by β_j for $\mathcal{S}$)

$$\min_{\boldsymbol{u}\in\mathcal{H}} \mathcal{J}(\boldsymbol{u}) = \mathcal{S}(\boldsymbol{u}) + \mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R) + \gamma\mathcal{C}(\boldsymbol{u}) \tag{2.4}$$

where $\phi(t) = t$ is a choice (but we shall recommend another choice) in the third (control) term:

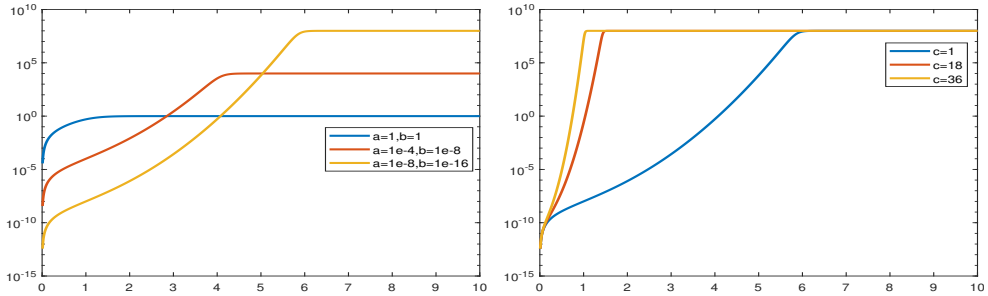
$$\mathcal{C}(\boldsymbol{u}) = \int_\Omega \phi(|\mu(\boldsymbol{u})|^2)\mathrm{d}\boldsymbol{x}, \quad |\mu(\boldsymbol{u})|^2 = \frac{(\partial_{x_1}u_1 - \partial_{x_2}u_2)^2 + (\partial_{x_2}u_1 + \partial_{x_1}u_2)^2}{(\partial_{x_1}u_1 + \partial_{x_2}u_2 + 2)^2 + (\partial_{x_2}u_1 - \partial_{x_1}u_2)^2}. \tag{2.5}$$

Here, we expect that ϕ possesses the following properties:

- $\phi(0) = 0$;
- ϕ has a big jump at 1;
- ϕ is twice continuously differentiable;
- ϕ is monotonically increasing in $[0, 1)$;
- ϕ has an upper bound.

Then we can see that ϕ promotes $|\mu(\boldsymbol{u})| < 1$ and the upper bound can help to build the existence of the solution of (2.4). For our previous work [39], ϕ is chosen as an unbounded function, so we cannot build the existence of the solution.

It is easy to see that $\phi_1(v) = \frac{v^2}{(1-v)^2+\epsilon}$ for a small ϵ, such as $\epsilon = 10^{-8}$, satisfies the requirements and has an upper bound $\frac{1}{\epsilon}$. In addition, we also consider another function $\phi_2(v) = \frac{ax^2}{bx^2+\mathrm{e}^{1-cx^2}}$, where a and b control the jump, c adjusts the position of the jump and the upper bound is $\frac{a}{b}$. Figure 1 shows the function $\phi_2(v)$ from varying a, b and c (with $c = 1$ on the left plot, and $a = 10^{-8}, b = 10^{-16}$ on the right fixed). Numerical tests in Section 4 will show that these two choices of ϕ lead to very similar performances and other choices of ϕ remain to be explored.



(a) ϕ_2 with the same $c = 1$ (b) ϕ_2 with the same $a = 10^{-8}, b = 10^{-16}$

Figure 1 The function ϕ_2: the left plot fixes c and the right plot fixes a, b

Data fitting. We consider a similarity measure $\mathcal{D}$ based on the gradient information [29]. This measure is motivated by the standard NGF [25, 41] and it primarily explores the potential of normalized gradients beyond its standard form. We shall combine the idea of normalized gradients fitting combined with a measure based on the triangular similarity inequality. More precisely, we consider the following fitting term

$$\mathcal{D}(T(\boldsymbol{x}+\boldsymbol{u}), R) = \mathcal{D}^{GF}(\boldsymbol{u}) + \alpha D^{TM}(\boldsymbol{u}) \tag{2.6}$$

where GF stands for 'gradient field difference' and TM for 'Triangular Measure' with

$$D^{GF}(\boldsymbol{u}) = \int_\Omega |\nabla_n T(\boldsymbol{x}+\boldsymbol{u}) - \nabla_n R|^2 \mathrm{d}\boldsymbol{x},$$
$$D^{TM}(\boldsymbol{u}) = \int_\Omega (|\nabla T(\boldsymbol{x}+\boldsymbol{u})| + |\nabla R| - |\nabla T(\boldsymbol{x}+\boldsymbol{u}) + \nabla R|)^2 \mathrm{d}\boldsymbol{x}.$$

Here the new fitting term, (2.6) of GF and TM, improves, in a major way, on the standard NGF [25, 41] which is defined by

$$D^{NGF}(\boldsymbol{u}) = \int_\Omega (1 - \nabla_n T(\boldsymbol{x}+\boldsymbol{u}) \cdot \nabla_n R)^2 \mathrm{d}\boldsymbol{x}.$$

Regularization. A regularizer controls the smoothness. Our primary choice for smoothness control is the diffusion regularizer [17] which uses first-order derivatives and promotes smoothness. As affine linear transformations are not included in the kernel of the H^1-regularizer, we desire a regularizer which can penalize such transformations. As such, we add the regularizer based on second-order derivatives (Laplacians) to the model which removes the need of any pre-registration step by an affine transformation. The second-order derivatives lead to smooth transformations [24]. Our recommended regularizer is given by

$$S(\boldsymbol{u}) = \frac{\beta_1}{2} S_1(\boldsymbol{u}) + \frac{\beta_2}{2} S_2(\boldsymbol{u}). \tag{2.7}$$

Mathematical analysis of model (2.4). Most registration models are non-convex with respect to $\boldsymbol{u}$ and consequently, if solutions exist, there are local minimizers or solutions are generally not unique. Below we prove the existence of a solution for the problem (2.4). Before stating the main result, we first consider the concept of the Carathéodory function.

Definition 2.1 *Let $\Omega \subset \mathbb{R}^2$ be an open set and let $f : \Omega \times \mathbb{R}^n \times \mathbb{R}^{2\times n} \times \mathbb{R}^{2\times 2\times n} \to [0, +\infty)$. Then f is a Carathéodory function if:*

(1) $f(\boldsymbol{x}, \cdot, \cdot, \cdot)$ is continuous for almost every $\boldsymbol{x} \in \Omega$.

(2) $f(\boldsymbol{x}, \boldsymbol{u}, \psi, \Theta)$ is measurable in $\boldsymbol{x}$ for every $(\boldsymbol{u}, \psi, \Theta) \in \mathbb{R}^n \times \mathbb{R}^{2\times n} \times \mathbb{R}^{2\times 2\times n}$.

We will use some theory about integrals of higher-order. It also sets up assumptions with which our optimization problem (2.4) admits a minimizer.

Lemma 2.2 ([42]) *Let $\Omega \subset \mathbb{R}^2$ be an open set and $f : \Omega \times \mathbb{R}^n \times \mathbb{R}^{2\times n} \times \mathbb{R}^{2\times 2\times n} \to [0, +\infty)$ satisfies the following assumptions:*

(i) f is a Carathéodory function.

(ii) $f(\boldsymbol{x}, \boldsymbol{u}, \psi, \Theta)$ is quasi-convex with respect to Θ.

(iii) $0 \le f(\boldsymbol{x}, \boldsymbol{u}, \psi, \Theta) \le a(\boldsymbol{x}) + C(|\boldsymbol{u}|^2 + |\psi|^2 + |\Theta|^2)$ where $a(\boldsymbol{x}) \in L^1(\Omega)$, $C > 0$.

*Then $\mathcal{J}(\boldsymbol{u})$ is weak lower semi-continuous (denoted by **wlsc** in $W^{2,2}(\Omega)$.*

To analyze the proposed model (2.4), we first consider the solution space where $\mathcal{H} = \{\boldsymbol{u} \in W^{2,2}(\Omega); |\int_\Omega \boldsymbol{u}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}| \le |\Omega|(M + \mathrm{diam}(\Omega))\}$. Then, it is convenient to rewrite the energy $\mathcal{J}(\cdot)$ by merging all terms under one integral in the following form:

$$\mathcal{J}(\boldsymbol{u}) = \int_\Omega f(\boldsymbol{x}, \boldsymbol{u}, \nabla \boldsymbol{u}, \nabla^2 \boldsymbol{u})\,\mathrm{d}\boldsymbol{x}, \tag{2.8}$$

where

$$f(\boldsymbol{x},\boldsymbol{u},\psi,\Theta) = \frac{\beta_1}{2}|\psi|^2 + \frac{\beta_2}{2}|\Theta|^2 + |\nabla_n T(\boldsymbol{u}) - \nabla_n R|^2 + \alpha(|\nabla T(\boldsymbol{u})| + |\nabla R| - |\nabla T(\boldsymbol{u}) + \nabla R|)^2 + \gamma\phi(|\mu(\boldsymbol{u})|^2).$$

To apply Lemma 2.2, we assume that the magnitude of $|\nabla R|$ and $|\nabla T|$ are bounded by a constant $c > 0$. Then, we have the following result:

Lemma 2.3 *The energy functional $\mathcal{J}(\cdot)$ is **wlsc** in $\mathcal{H}$.*

Proof We verify that the functional $f(\cdot)$ fulfils the assumptions in Lemma 2.2:

i) Since the gradients of the fixed and the moving images ∇R and ∇T are assumed to be continuous, and $\boldsymbol{u} \in \mathcal{H}$, i.e., continuous, we have $\nabla T(\boldsymbol{u})$ is continuous. Thus, the function $f(\cdot)$ is a Carathéodory function.

ii) It is easy to check that $f(\boldsymbol{x},\boldsymbol{u},\psi,\Theta)$ is convex with respect to Θ, clearly implying that it is quasi-convex.

iii) For condition (iii), we have $|\nabla_n T(\mathbf{u})| \leq 1$ and $|\nabla_n R| \leq 1$, which means that:

$$|\nabla_n T(\boldsymbol{u}) - \nabla_n R|^2 \leq (|\nabla_n T(\boldsymbol{u})| + |\nabla_n R|)^2 \leq 4. \tag{2.9}$$

Moreover, since $|\nabla R|$ and $|\nabla T|$ are bounded by a constant $c > 0$ and $\boldsymbol{u}$ is continuous, then $\nabla T(\boldsymbol{u})$ is also bounded and we have

$$\alpha(|\nabla T(\boldsymbol{u})| + |\nabla R| - |\nabla T(\boldsymbol{u}) + \nabla R|)^2 \leq \alpha(|\nabla T(\boldsymbol{u})| + |\nabla R|)^2 \leq 4\alpha c^2. \tag{2.10}$$

Moreover, from the definition of the function $\phi(\cdot)$, we clearly have

$$\phi(|\mu(\boldsymbol{u})|^2) \leq \frac{1}{\epsilon}. \tag{2.11}$$

Therefore, using inequalities (2.9)–(2.11), we have

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{u},\psi,\Theta) =& \frac{\beta_1}{2}|\psi|^2 + \frac{\beta_2}{2}|\Theta|^2 + |\nabla_n T(\boldsymbol{u}) - \nabla_n R|^2 + \\ & \alpha(|\nabla T(\boldsymbol{u})| + |\nabla R| - |\nabla T(\boldsymbol{u}) + \nabla R|)^2 + \gamma\phi(|\mu(\boldsymbol{u})|^2) \\ \leq& \frac{\beta_1}{2}|\psi|^2 + \frac{\beta_2}{2}|\Theta|^2 + 4 + 4\alpha c^2 + \frac{\gamma}{\epsilon}. \end{aligned}$$

Then, the function $f(\cdot)$ fulfils the condition (iii) of Lemma 2.2 with $a(\boldsymbol{x}) \equiv 4 + 4\alpha c^2 + \frac{\gamma}{\epsilon}$ which implies that the energy $\mathcal{J}(\cdot)$, is ***wlsc*** in $\mathcal{H}$. □

We are now ready to prove the existence of a solution for the minimization model (2.4). Based on Lemmas 2.2 and 2.3, we have the following result

Theorem 2.4 *The minimization problem (2.4) admits at least one solution in the space $\mathcal{H}$.*

Proof Using the positivity of $\mathcal{C}(\boldsymbol{u})$, $\mathcal{D}^{GF}(T(\boldsymbol{u}),R)$ and $\mathcal{D}^{TM}(T(\boldsymbol{u}),R)$ and the generalized Poincaré inequality [43] for $\mathcal{S}$, there exist two constants $C, K \in \mathbb{R}$ such that

$$\mathcal{J}(\boldsymbol{u}) \geq C\|\boldsymbol{u}\|_{\mathcal{H}}^2 + K. \tag{2.12}$$

Now, we consider a minimizing sequence $(\boldsymbol{u}_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ of $\mathcal{J}(\cdot)$, i.e.,

$$\mathcal{J}(\boldsymbol{u}_n) \underset{n\to\infty}{\longrightarrow} \inf_{\boldsymbol{u}\in\mathcal{H}} \mathcal{J}(\boldsymbol{u}).$$

The inequality (2.12) guarantees that the sequence $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ is uniformly bounded in $\mathcal{H}$. Thus, there exists a subsequence, still denoted $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$, such that $\boldsymbol{u}_n \underset{n\to\infty}{\rightharpoonup} \boldsymbol{u}$ weakly in $\mathcal{H}$. Using the weak lower semi-continuity of $\mathcal{J}(\cdot)$, we obtain that the limit $\boldsymbol{u}$ is a minimizer of $\mathcal{J}(\cdot)$. □

3. The solution algorithm

In this section, we shall develop a numerical algorithm to solve model (2.4). Here, we choose first-discretize-then-optimize method, namely directly discretize the variational model to get a discrete optimization problem and then use optimization methods to solve this resulting optimization problem.

3.1. Discretization

In the implementation, we set $\Omega = [0,1]^2$, employ the nodal grid and define a spatial partition $\Omega_h^n = \{\boldsymbol{x}^{i,j} \in \Omega | \boldsymbol{x}^{i,j} = (x_1^i, x_2^j) = (ih, jh), 0 \le i \le n, 0 \le j \le n\}$, where $h = \frac{1}{n}$ and the discrete domain consists of n^2 cells of size $h \times h$. We discretize the displacement field $\boldsymbol{u}$ on the nodal grid, namely $\boldsymbol{u}^{i,j} = (u_1^{i,j}, u_2^{i,j}) = (u_1(x_1^i, x_2^j), u_2(x_1^i, x_2^j))$. By lexicographical ordering, we reshape four matrices to two long vectors of dimension $\mathbb{R}^{2(n+1)^2\times 1}$

$$\begin{aligned} X &= (x_1^0, ..., x_1^n, ..., x_1^0, ..., x_1^n, x_2^0, ..., x_2^0, ..., x_2^n, ..., x_2^n)^T, \\ U &= (u_1^{0,0}, ..., u_1^{n,0}, ..., u_1^{0,n}, ..., u_1^{n,n}, u_2^{0,0}, ..., u_2^{n,0}, ..., u_2^{0,n}, ..., u_2^{n,n})^T. \end{aligned}$$

Discretization of fitting term. Firstly, set $\vec{R} = \vec{R}(PX) \in \mathbb{R}^{n^2\times 1}$ as the discretized reference image and $\vec{T}(PX + PU) \in \mathbb{R}^{n^2\times 1}$ as the discretized deformed template image, where $P \in \mathbb{R}^{2n^2\times 2(n+1)^2}$ is an average matrix from the nodal grid to the cell-centered grid [6]. In order to discretize ∇T and ∇R, we introduce two discrete operators: $D_1 = I_n \otimes \partial_h^1$ and $D_2 = \partial_h^1 \otimes I_n$, where

$$\partial_h^1 = \frac{1}{2h}\begin{bmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & ... & ... & ... & \\ & & -1 & 0 & 1 \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

Hence, the discretized ∇T and ∇R are $[D_1\vec{T}, D_2\vec{T}]$ and $[D_1\vec{R}, D_2\vec{R}]$ respectively. Set $\mathrm{LT} = (\sum_{i=1}^2 D_i\vec{T} \odot D_i\vec{T} + \epsilon)^{\cdot 1/2}$, $\mathrm{LR} = (\sum_{i=1}^2 D_i\vec{R} \odot D_i\vec{R} + \epsilon)^{\cdot 1/2}$ and $\mathrm{LTR} = (\sum_{i=1}^2 D_i(\vec{T}+\vec{R}) \odot D_i(\vec{T}+\vec{R}) + \epsilon)^{\cdot 1/2}$, where $\odot$ indicates the component-wise product and $(\cdot)^{\cdot 1/2}$ indicates the component-wise square root. Then for $\mathcal{D}^{GF}(\boldsymbol{u})$ and $\mathcal{D}^{TM}(\boldsymbol{u})$, we have the following discretizations:

$$\mathcal{D}^{GF}(\boldsymbol{u}) \approx h^2 p_1^T p_1, \quad \mathcal{D}^{TSM}(\boldsymbol{u}) \approx h^2 p_2^T p_2, \tag{3.1}$$

where $p_1 = (D_1\vec{T}./\mathrm{LT} - D_1\vec{R}./\mathrm{LR}; D_2\vec{T}./\mathrm{LT} - D_2\vec{R}./\mathrm{LR}; D_3\vec{T}./\mathrm{LT} - D_3\vec{R}./\mathrm{LR})$, $p_2 = \mathrm{LT} + \mathrm{LR} - \mathrm{LTR}$ and ./ indicates the component-wise division.

Discretization of regularization term. By using the forward difference, central difference, mid-point rule and Dirichlet boundary conditions, for the first-order and second-order regularization terms, we have the following approximations:

$$\mathcal{S}_1(\boldsymbol{u}) \approx h^2 \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=1}^{2} \big(\frac{u_l^{i+1,j}-u_l^{i,j}}{h}\big)^2 + \big(\frac{u_l^{i,j+1}-u_l^{i,j}}{h}\big)^2 \tag{3.2}$$

and

$$\begin{aligned}\mathcal{S}_2(\boldsymbol{u}) \approx & h^2 \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\sum_{l=1}^{2} \big(\frac{u_l^{i+1,j}-2u_l^{i,j}+u_l^{i-1,j}}{h^2}\big)^2 + \big(\frac{u_l^{i,j+1}-2u_l^{i,j}+u_l^{i,j-1}}{h^2}\big)^2 + \\ & h^2 \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\sum_{l=1}^{2} \big(\frac{u_l^{i,j}-u_l^{i+1,j}-u_l^{i,j+1}+u_l^{i+1,j+1}}{h^2}\big)^2. \end{aligned} \tag{3.3}$$

By introducing the discrete differential operator and using the Kronecker product, we can build two matrices A_1 and A_2 and have the following compact form:

$$\mathcal{S}_1(\boldsymbol{u}) \approx h^2 U^T A_1 U, \quad \mathcal{S}_2(\boldsymbol{u}) \approx h^2 U^T A_2 U. \tag{3.4}$$

Discretization of control term. Note that $\phi(|\mu(\boldsymbol{u})|^2)$ involves only first order derivatives and all $\boldsymbol{u}^{i,j}$ are available at vertex pixels. Then we shall divide each cell (Figure 2) into 2 triangles. In each triangle, we construct two linear interpolation functions to approximate the u_1 and u_2. Consequently, all partial derivatives are locally constants or $\phi(|\mu(\boldsymbol{u})|^2)$ is constant in each triangle.

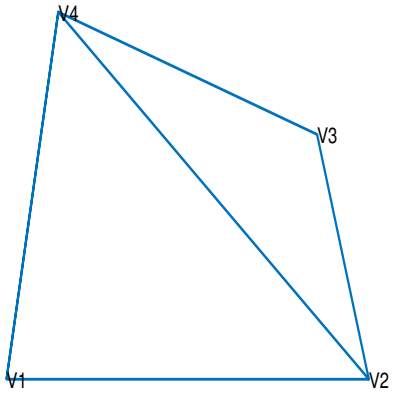


Figure 2 Partition of a cell. $\triangle V_1V_2V_4$ is $\Omega_{i,j,k}$

Set $\mathbf{L}^{i,j,k}(\boldsymbol{x}) = (L_1^{i,j,k}(\boldsymbol{x}), L_2^{i,j,k}(\boldsymbol{x})) = (a_1^{i,j,k}x_1 + a_2^{i,j,k}x_2 + a_3^{i,j,k}, a_4^{i,j,k}x_1 + a_5^{i,j,k}x_2 + a_6^{i,j,k})$, which is the linear interpolation for $\boldsymbol{u}$ in the $\Omega_{i,j,k}$. Note that $\partial_{x_1}L_1^{i,j,k} = a_1^{i,j,k}, \partial_{x_2}L_1^{i,j,k} = a_2^{i,j,k}, \partial_{x_1}L_2^{i,j,k} = a_4^{i,j,k}$ and $\partial_{x_2}L_2^{i,j,k} = a_5^{i,j,k}$. Then according to the partition in Figure 2, we have

$$\mathcal{C}(\boldsymbol{u}) = \int_\Omega \phi(|\mu(\boldsymbol{u})|^2)\mathrm{d}\boldsymbol{x} \approx \frac{h^2}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{2}\phi\big(\frac{(a_1^{i,j,k}-a_5^{i,j,k})^2+(a_2^{i,j,k}+a_4^{i,j,k})^2}{(a_1^{i,j,k}+a_5^{i,j,k}+2)^2+(a_2^{i,j,k}-a_4^{i,j,k})^2}\big). \tag{3.5}$$

To simplify (3.5), define 3 vectors $\vec{\boldsymbol{r}}(U), \vec{\boldsymbol{r}}^1(U), \vec{\boldsymbol{r}}^2(U) \in \mathbb{R}^{2n^2}$ by $\vec{\boldsymbol{r}}(U)_\ell = \vec{\boldsymbol{r}}^1(U)_\ell \vec{\boldsymbol{r}}^2(U)_\ell$,

$$\begin{aligned}\vec{\boldsymbol{r}}^1(U)_\ell &= (a_1^{i,j,k} - a_5^{i,j,k})^2 + (a_2^{i,j,k} + a_4^{i,j,k})^2,\\ \vec{\boldsymbol{r}}^2(U)_\ell &= 1/[(a_1^{i,j,k} + a_5^{i,j,k} + 2)^2 + (a_2^{i,j,k} - a_4^{i,j,k})^2]\end{aligned}$$

where $\ell = (k-1)n^2 + (j-1)n + i \ \in [1, 4n^2]$.

Hence, (3.5) becomes

$$\mathcal{C}(\boldsymbol{u}) \approx \frac{h^2}{2}\boldsymbol{\phi}(\vec{\boldsymbol{r}}(U))e^T \tag{3.6}$$

where $\boldsymbol{\phi}(\vec{\boldsymbol{r}}(U)) = (\phi(\vec{\boldsymbol{r}}(U)_1), ..., \phi(\vec{\boldsymbol{r}}(U)_{2n^2}))$ denotes the pixel-wise discretization of u_1, u_2 at all cell centers, and $e = (1, ..., 1) \in \mathbb{R}^{2n^2 \times 1}$.

Finally, combining the above three parts (3.1), (3.4) and (3.6), we get the discretization formulation for model (2.4):

$$\min_U J(U) := h^2 p_1^T p_1 + \alpha h^2 p_2^T p_2 + \frac{\beta_1 h^2}{2} U^T A_1 U + \frac{\beta_2 h^2}{2} U^T A_2 U + \frac{\gamma h^2}{2}\boldsymbol{\phi}(\vec{\boldsymbol{r}}(U))e^T. \tag{3.7}$$

3.2. Optimization method for the discretized problem (3.7)

In the numerical implementation, we choose a line search method to solve the resulting unconstrained optimization problem (3.7). Here, the basic iterative scheme based on the search direction δU^k step k is

$$U^{k+1} = U^k + \theta \delta U^k, \tag{3.8}$$

where θ is the step length. The key point here is to obtain a descent search direction. A commonly used approach to generate δU^k is to apply a standard Newton method

$$H^e \delta U^k = -d_J$$

where H^e and d_J are respectively the exact Hessian and the gradient of (3.7). Unfortunately the above equation does not generate a descent search direction since H^e is indefinite. Below we shall discuss, in a Gauss-Newton framework, how to construct a positive, definite and approximate Hessian H so that a descent direction is generated. Consequently a converging method is obtained.

Gradient and approximated Hessian of (3.7). Firstly, we consider computing the gradient and approximated Hessian of the discretized fitting term $h^2 p_1^T p_1 + \alpha h^2 p_2^T p_2$. Its gradient and approximated Hessian are respectively:

$$\begin{cases} d_1 = 2h^2 P^T(\mathrm{d}p_1^T p_1 + \alpha \mathrm{d}p_2^T p_2),\\ \hat{H}_1 = h^2 P^T(\mathrm{d}p_1^T \mathrm{d}p_1 + \alpha \mathrm{d}p_2^T \mathrm{d}p_2)P, \end{cases} \tag{3.9}$$

where $\tilde{\tilde{p}}_1 = [\Lambda D_1 - \mathrm{sdiag}(D_1\vec{T}./t)\Gamma; \Lambda D_2 - \mathrm{sdiag}(D_2\vec{T}./t)\Gamma]$, $\tilde{\tilde{p}}_2 = \sum_{i=1}^2 \mathrm{sdiag}(D_i\vec{T}./\mathrm{LT} - D_i(\vec{T} + \vec{R})./\mathrm{LTR})D_i$, $\Lambda = \mathrm{sdiag}(1./\mathrm{LT})$, $t = \mathrm{LT}^{.3}$, $\Gamma = \sum_{i=1}^2 \mathrm{sdiag}(D_i\vec{T})D_i$ and $\mathrm{sdiag}(v)$ is a diagonal with v on its main diagonal. Here the special notation './' and '$(\cdot)^{.3}$' refer to pointwise division and power function respectively. Quantities $\mathrm{LT}, \mathrm{LTR}, D_1, D_2$ are as defined in the previous subsection.

Remark 3.1 Evaluating the deformed template image $\vec{T}$ must involve interpolation because $\tilde{\mathbf{U}}$ are not in general pixel points. Here in our implementation, we choose B-splines for the interpolation.

For the discretized diffusion regularizer $\frac{\beta_1 h^2}{2}U^T A_1 U + \frac{\beta_2 h^2}{2}U^T A_2 U$, its gradient and Hessian are respectively

$$\begin{cases} d_2 = h^2(\beta_1 A_1 + \beta_2 A_2)U, \\ H_2 = h^2(\beta_1 A_1 + \beta_2 A_2). \end{cases} \tag{3.10}$$

Finally, for the discretized Beltrami term $\frac{\beta h^2}{2}\boldsymbol{\phi}(\vec{\boldsymbol{r}}(U))e^T$, the gradient and approximated Hessian are as follows:

$$\begin{cases} d_3 = \frac{\beta h^2}{2}\mathrm{d}\vec{\boldsymbol{r}}^T \mathrm{d}\boldsymbol{\phi}(\vec{\boldsymbol{r}}), \\ \hat{H}_3 = \frac{\beta h^2}{2}\mathrm{d}\vec{\boldsymbol{r}}^T \mathrm{d}^2\boldsymbol{\phi}(\vec{\boldsymbol{r}})\mathrm{d}\vec{\boldsymbol{r}}. \end{cases} \tag{3.11}$$

where $\mathrm{d}\boldsymbol{\phi}(\vec{\boldsymbol{r}}) = (\phi'(\vec{\boldsymbol{r}}_1), ..., \phi'(\vec{\boldsymbol{r}}_{4n^2}))^T$ is the vector of derivatives of $\boldsymbol{\phi}$ at all cell centers,

$$\begin{cases} \mathrm{d}\vec{\boldsymbol{r}} = \mathrm{diag}(\vec{\boldsymbol{r}}^1)\mathrm{d}\vec{\boldsymbol{r}}^2 + \mathrm{diag}(\vec{\boldsymbol{r}}^2)\mathrm{d}\vec{\boldsymbol{r}}^1, \\ \mathrm{d}\vec{\boldsymbol{r}}^1 = 2\mathrm{diag}(A_{31}U)A_{31} + 2\mathrm{diag}(A_{32}U)A_{32}, \\ \mathrm{d}\vec{\boldsymbol{r}}^2 = -\mathrm{diag}(\vec{\boldsymbol{r}}^2 \odot \vec{\boldsymbol{r}}^2)[2\mathrm{diag}(A_{33}U + 2)A_{33} + 2\mathrm{diag}(A_{34}U)A_{34}], \end{cases} \tag{3.12}$$

$\odot$ denotes a Hadamard product, $\tilde{\vec{\boldsymbol{r}}}, \tilde{\vec{\boldsymbol{r}}}^1, \tilde{\vec{\boldsymbol{r}}}^2$ are the Jacobian of $\vec{\boldsymbol{r}}, \vec{\boldsymbol{r}}^1, \vec{\boldsymbol{r}}^2$ with respect to U respectively, $[\tilde{\boldsymbol{\phi}}(\vec{\boldsymbol{r}})]_\ell$ is the ℓth component of $\tilde{\boldsymbol{\phi}}(\vec{\boldsymbol{r}})$ and $\tilde{\tilde{\boldsymbol{\phi}}}(\vec{\boldsymbol{r}})$ is the Hessian of $\boldsymbol{\phi}$ with respect to $\vec{\boldsymbol{r}}$, which is a diagonal matrix whose ith diagonal element is $\phi''(\vec{\boldsymbol{r}}_i),\ 1 \le i \le 2n^2$. More details about $\vec{\boldsymbol{r}}^1$, $\vec{\boldsymbol{r}}^2$, A_{31}, A_{32}, A_{33} and A_{34} are shown in Appendix A.

Therefore, combining the above results for 3 terms, we can obtain the gradient

$$d_J = d_1 + d_2 + d_3 \tag{3.13}$$

and the approximated Hessian of (3.7):

$$H = \hat{H}_1 + H_2 + \hat{H}_3. \tag{3.14}$$

Search direction. With the above approximated Hessian (3.14), in each outer (nonlinear) iteration, we solve the Gauss-Newton system

$$H\delta U = -d_J \tag{3.15}$$

to obtain the search direction δU for (3.7). Since we impose the Dirichlet boundary conditions, H is symmetric and positive definite. In our implementation, we choose MINRES as the numerical solver [44, 45].

Step length. Here, we choose a popular inexact line search condition, Armijo condition, which determines a step length θ that satisfies the following sufficient decrease condition:

$$\mathcal{J}(U + \theta\delta U) < \mathcal{J}(U) + \theta\eta {d_{\mathcal{J}}}^T \delta U. \tag{3.16}$$

Here, we set $\eta = 10^{-4}$ and use the backtracking approach to find a suitable θ. In addition, we need to check that $\vec{\boldsymbol{r}}(U)$ is smaller than 1 which is the norm of the discretized Beltrami coefficient. For more details, please refer to [46–48].

Stopping criteria. In the implementation, we choose the stopping criteria used in [3]:

(1.a) $\|J(U^{k+1}) - J(U^k)\| \leq \tau_J(1 + \|J(U^0)\|)$,
(1.b) $\|U^{k+1} - U^k\| \leq \tau_W(1 + \|X + U^0\|)$,
(1.c) $\|d_J\| \leq \tau_G(1 + \|J(U^0)\|)$,
(2) $\|d_J\| \leq \text{eps}$,
(3) $k \geq \text{MaxIter}$.

Here, eps is the machine precision and MaxIter is the maximal number of outer iterations. We set $\tau_J = 10^{-3}$, $\tau_W = 10^{-2}$ and $\tau_G = 10^{-2}$. If any one of (1) (2) and (3) is satisfied, the iterations are terminated. Hence, a Gauss-Newton numerical scheme with Armijo line search can be developed and summarized in Algorithm 1. The global convergence of Algorithm 1 is discussed in Appendix B.

Algorithm 1 Gauss-Newton scheme by using Armijo line search for Image Registration: $U \leftarrow \text{GNAIR}(\alpha, \beta_1, \beta_2, \gamma, U^0, T, R)$

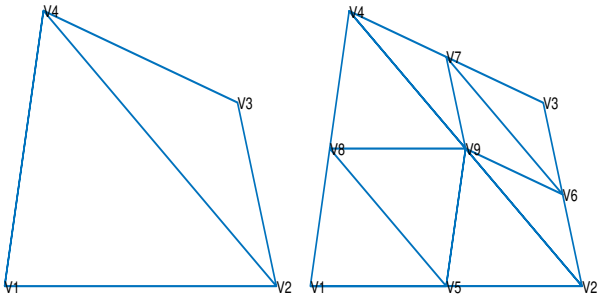
Step 1: Set $k = 0$ at the solution point $U^k = U^0$.
Step 2: For (3.7), compute the energy functional $J(U^k)$, its gradient d_J^k and the approximated Hessian H^k by (3.14).
while "none of the listed 3 stopping criteria are satisfied" **do**
  Solve the Gauss-Newton equation: $H^k \delta U^k = -d_J^k$;
  Use Line Search to find step length θ;
  $U^{k+1} = U^k + \theta \delta U^k$;
  $k = k + 1$;
  Compute $J(U^k)$, d_J^k and H^k;
end while

Multi-Level strategy. A multi-level strategy is a standard technique in image registration. In the multi-level strategy, we firstly coarsen the template T and the reference R by L levels. Then we can obtain U_1 by solving our model (2.4) on the coarsest level. In order to give a good initial guess for the finer level, we adopt an interpolation operator on U_1 to obtain U_2^0 as the initial guess for the next level. We repeat this process and can get the final registration on the finest level. The most important advantage of the multi-level strategy is that it can save computation time because of less variables on the coarser level than on the fine level. In addition, it can help to avoid to trap into a local minimum.

Here, we use finite element interpolation (Figure 3) to obtain U_2^0 rather than bilinear interpolation since it can keep consistent between the coarse level and fine level.

4. Numerical results

In this section, we will show some numerical results to illustrate the performances of our proposed model (2.4). We set the zero vector as the initial guess U^0. In order to measure the performance of our model (2.4), we introduce three terms to measure the deformed template



(a) Coarse Level (b) Fine Level

Figure 3 Interpolation from coarse level to fine level

with the reference:

$$\text{GFer} = \frac{D^{GF}(\boldsymbol{u})}{D^{GF}(\boldsymbol{u}^0)}, \tag{4.1}$$

$$\text{NGFer} = \frac{D^{NGF}(\boldsymbol{u})}{D^{NGF}(\boldsymbol{u}^0)}, \tag{4.2}$$

$$\text{MIer} = -D^{MI}(\boldsymbol{u}). \tag{4.3}$$

The good result means that it can lead to small GFer, small NGFer and large MIer.

All of the tests are coded by Matlab 2018a and performed on a PC with 3.20 GHz Intel(R) Core(TM) i5-6500 microprocessor and with installed memory (RAM) of 8 GB.

4.1. Example 1 – Registration of a pair of MR T1 and T2 images

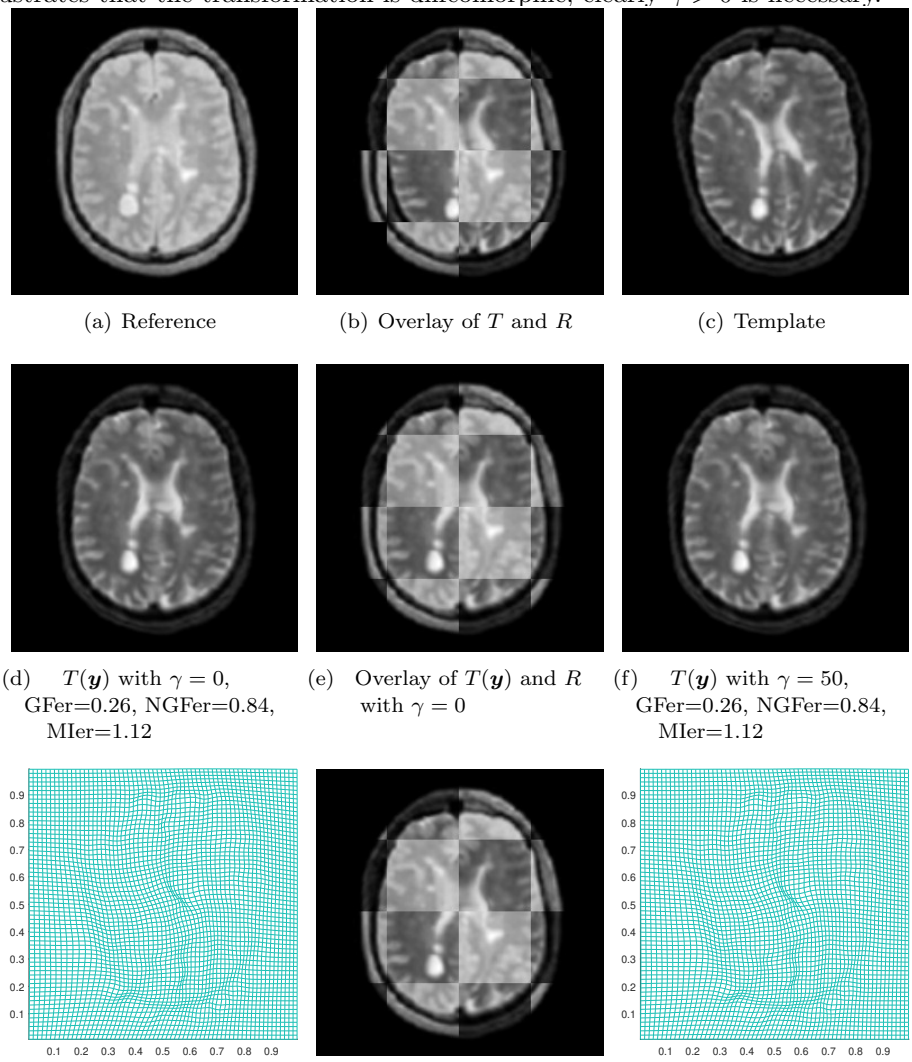
In this example, we consider a pair of MRI images (T1 and T2) in Figure 4 (a, b) whose resolution is 256×256. We choose a six-step multilevel strategy. In order to choose the parameter easily, we fix α and set $\alpha = 1$.

First, we consider the model without Beltrami control term, namely $\gamma = 0$. For the parameters of regularizers, we set $\beta_1 = 100$ and $\beta_2 = 70$. Its resulting deformed template and transformation are shown in Figure 4 (d, g). From Figure 4 (f), we can find that the deformed template is visually good and this choice gives GFer = 0.2622, NGFer = 0.8433 and MIer = 1.1149. In addition, the resulting transformation is smooth and has no folding (since the minimum of the Jacobian determinant of the transformation is positive).

Now, we investigate the sensitivity of β_1 and β_2 without Beltrami control term i.e., $\gamma = 0$. From Figure 5 (a-c), we can find that when we fix α and change β_1 and β_2, GFer, NGFer and MIer are robust. However, according to Figure 5 (d), although GFer, NGFer and MIer are robust, the minima of the Jacobian determinants of the transformations are not consistent and can be negative, which means that the resulting transformation has folding. This phenomenon is clear because in the model, there is no restriction about how to control the transformation.

Second, in order to overcome this drawback when $\gamma = 0$, we keep β_1, β_2 unchanged and choose a suitable $\gamma > 0$. Here, we set $\gamma = 50$ and choose $\phi = \phi_1$. Figure 4 (f, i) shows the corresponding deformed template and transformation. We see that the deformed template (f) is similar visually to the previous one (d) and the measurements are also similar (GFer=0.2592, NGFer=0.8420

and MIer=1.1154). The minimum of the Jacobian determinant of the transformation is positive, which illustrates that the transformation is diffeomorphic; clearly $\gamma > 0$ is necessary.



(a) Reference (b) Overlay of T and R (c) Template

(d) $T(\boldsymbol{y})$ with $\gamma = 0$, GFer=0.26, NGFer=0.84, MIer=1.12 (e) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 0$ (f) $T(\boldsymbol{y})$ with $\gamma = 50$, GFer=0.26, NGFer=0.84, MIer=1.12

(g) Map $\boldsymbol{y}$ with $\gamma = 0$, $\min \det(\nabla \boldsymbol{y}) = 0.14$ (h) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 50$ (i) Map $\boldsymbol{y}$ with $\gamma = 50$, $\min \det(\nabla \boldsymbol{y}) = 0.12$

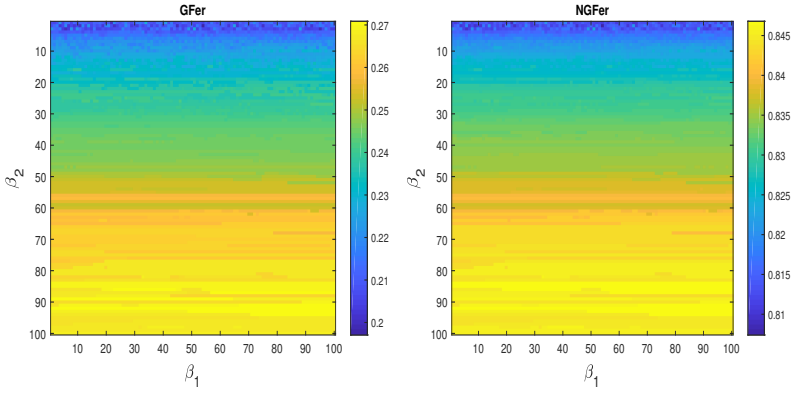
Figure 4 Example 1: a pair of MRI images (T1 and T2). By using the control term, the resulting transformation is diffeomorphic and the deformed template is also visually satisfactory

We also investigate the sensitivity of the parameters β_1 and β_2 under this fixed γ. Figure 6 clearly demonstrates that no matter how we change β_1 and β_2, the measurements of the results are robust. In addition, since the minimum of the Jacobian determinant of the transformations is always positive, our proposed model is not sensitive with respect to the β_1 and β_2 under a suitable γ.

For the computing time, the model with $\gamma = 0$ needs 14.4 seconds and the model with $\gamma = 50$ needs 31.7 seconds. Although introducing the Beltrami control term can increase the running

time, the extra computation cost is worthy the effort since it can guarantee that the resulting transformation is regular and has no folding. We also investigate the convergence of the algorithm for our model. Here, we force the relative norm of the gradient of the approximated solution to reach 10^{-3} although it only runs several iterations by using the practical stopping criteria. Here, according to Figure 7, we can find that the algorithm for our model is convergent.

Hence, this example illustrates that our new control term can effectively control the transformation and lead to an accurate registration. Meanwhile, the new control term can make this model more robust.



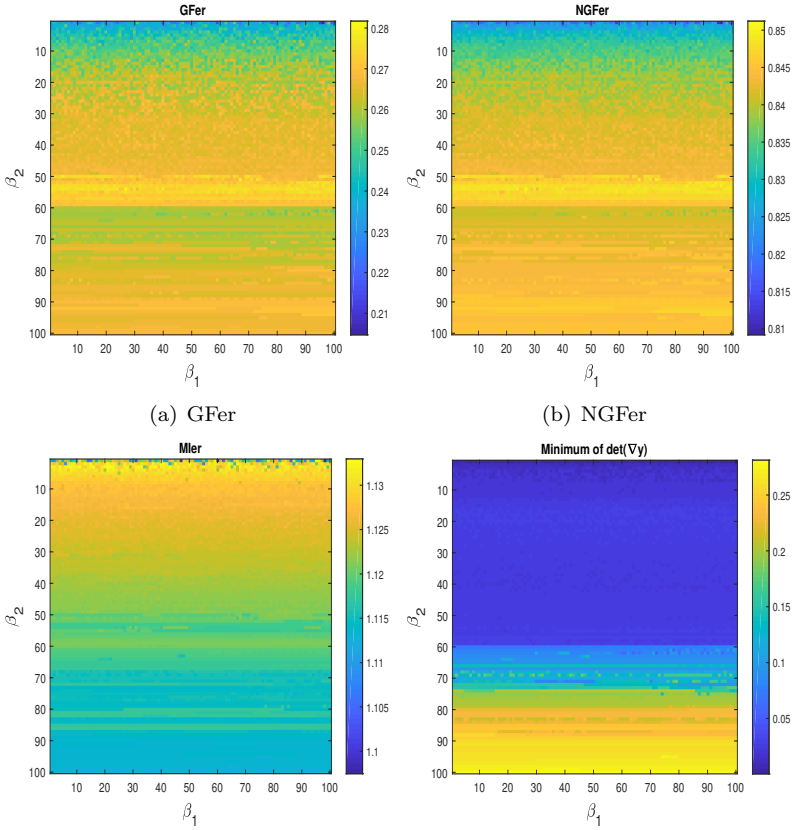
(a) GFer (b) NGFer

(c) MIer (d) Minimum of det $\nabla \boldsymbol{y}$

Figure 5 Example 1: sensitivity of β_1 and β_2 without Beltrami control term. Obviously, this model is sensitive with respect to β_1 and β_2 since the minimum of the Jacobian determinant of the transformation can be negative with some specific β_1 and β_2

4.2. Example 2 – Registration of a second pair of MRI images

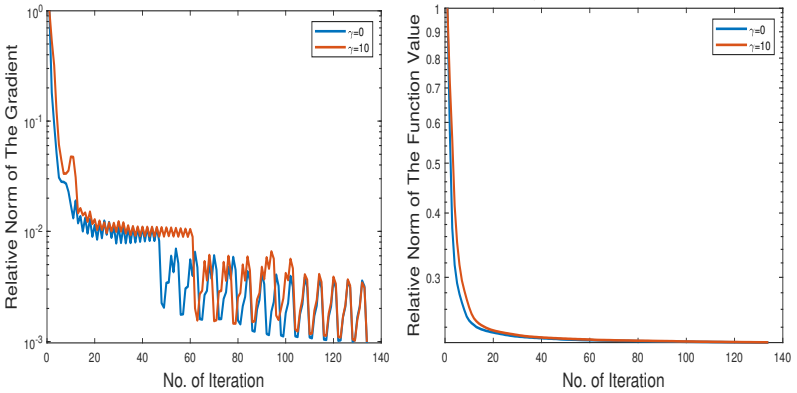
Here, we consider another pair of MRI images (T1 and T2) in Figure 8 (a,b) for verifying our observation for Example 1. The resolution is 256×256 and we also choose a six-step multilevel



(a) GFer (b) NGFer

(c) MIer (d) Minimum of det $\nabla \boldsymbol{y}$

Figure 6 Example 1: sensitivity of β_1 and β_2 with Beltrami control term. Clearly, this model is not sensitive and robust with respect to β_1 and β_2 under a suitable γ



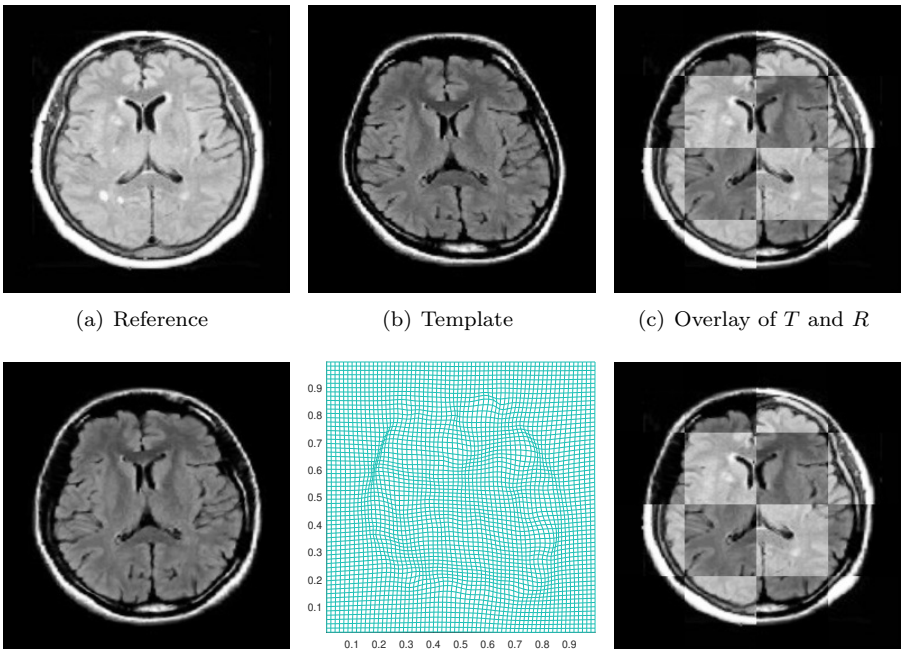
(a) Relative norm of the gradient (b) Relative norm of the function value

Figure 7 Example 1: Relative norm of the gradient and relative norm of the function value with or without γ. Here, we can notice that our algorithm is convergent

strategy. Again, in order to reduce the complexity of choosing parameters, we fix $\alpha = 10^{-5}$ in this example.

Firstly, we consider the model with $\gamma = 0$ i.e., test the model made up of NGF and the same first- and second-order regularizer. From Figure 8 (d-f), although the deformed template is visually satisfactory, we can find that the resulting transformation has folding (since the minimum of the Jacobian determinant is negative). Figure 9 (a) shows the details about the mesh folding. As a comparison to later, we set $\beta_1 = 50, \beta_2 = 0.01$ and $\gamma = 0$. Here, we test three pairs of (β_1, β_2) and the corresponding results are shown in Figure 10. We can find that for our proposed model with $\gamma = 0$, it is very hard to choose the suitable parameters to get a good registration, namely, simultaneously a diffeomorphic transformation and a visually pleasing deformed template.

Secondly, in order to overcome this difficulty by non-zero γ, we keep β_1, β_2 unchanged and choose γ as 0.1, 1 and 10 separately. Here, we test both ϕ_1 and ϕ_2. Figures 11 and 12 show that they can all generate visually satisfactory deformed template and diffeomorphic transformations. In addition, according to Tables 1 and 2, we can see that the measurements obtained by these choices are very similar, which again demonstrates that the new model is robust due to the use of the Beltrami control term.



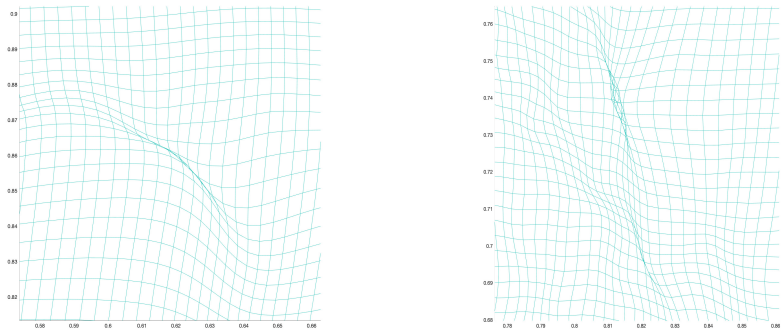
(a) Reference (b) Template (c) Overlay of T and R

(d) $T(\boldsymbol{y})$: $\gamma = 0$, GFer = 0.73 NGFer=0.96, MIer=0.78 (e) $\boldsymbol{y}$ with $\gamma = 0$, $\min\det(\nabla \boldsymbol{y})$=**−0.1114** (f) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 0$

Figure 8 Example 2: a pair of MRI images (T1 and T2). Without using the control term, the resulting transformation is **not** diffeomorphic although $T(\boldsymbol{y})$ is visually satisfactory

5. Conclusion

Image registration is an increasingly important and often challenging image processing task



(a) Folding mesh by new model without control term (b) Folding mesh by NGF

Figure 9 Example 2: zoom in the transformations obtained by the new model without control term (left) and NGF (right) respectively. Clearly, the transformation has folding

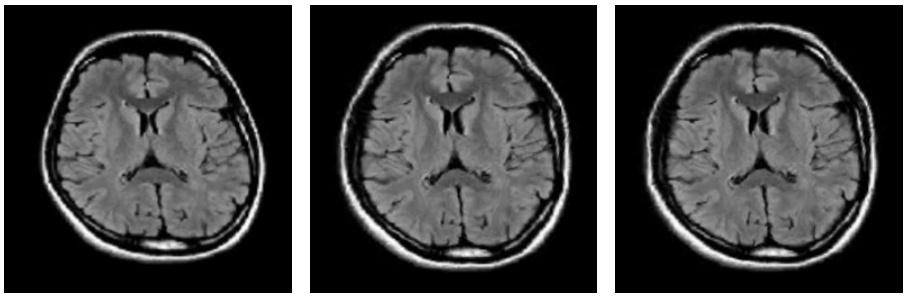
with a broad range of applications such as in astronomy, optics, biology, chemistry and medical imaging. In this paper to improve the multi-modality registration model based on the normalized gradients of the images, we propose a new gradients-based variational model using a regularization term which combines first- and second-order derivatives of the displacement. More importantly, in order to control the transformation, we add a novel control term motivated by Beltrami coefficient, which can lead to a diffeomorphic transformation. By employing first-discretize-then-optimize method, we design an effective solver to solve our proposed model numerically. Experimental tests confirm that our proposed model performs well in multi-modality images registration. In addition, with the help of the Beltrami control term, the proposed model is more robust with respect to the parameters. Future work will consider generalizations to 3 dimensions and registration of images that do not have dominant gradients. In addition, since the Beltrami coefficient is defined in the complex space which can be considered as a 2 dimensional space, how to extend the definition of Beltrami coefficient to 3 dimensional space is also one of our future directions.

γ	time(s)	GFer	NGFer	MIer	$\min\det(\nabla \boldsymbol{y})$	$\max\det(\nabla \boldsymbol{y})$
0	65.6	0.7341	0.9564	0.7835	−0.1114	1.8740
0.1	147.0	0.7352	0.9561	0.7823	0.1188	2.2557
1	331.1	0.7327	0.9553	0.7833	0.1494	2.2922
10	135.3	0.7349	0.9561	0.7824	0.2339	2.2788

Table 1 Example 2: Sensitivity test on γ when $\alpha = 10^{-5}, \beta_1 = 50$ and $\beta_2 = 0.01$ for ϕ_1

γ	time(s)	GFer	NGFer	MIer	$\min\det(\nabla \boldsymbol{y})$	$\max\det(\nabla \boldsymbol{y})$
0	65.6	0.7341	0.9564	0.7835	−0.1114	1.8740
0.1	168.7	0.7356	0.9558	0.7832	0.1127	2.2322
1	162.8	0.7364	0.9562	0.7828	0.1136	2.2232
10	202.9	0.7355	0.9558	0.7833	0.1188	2.2240

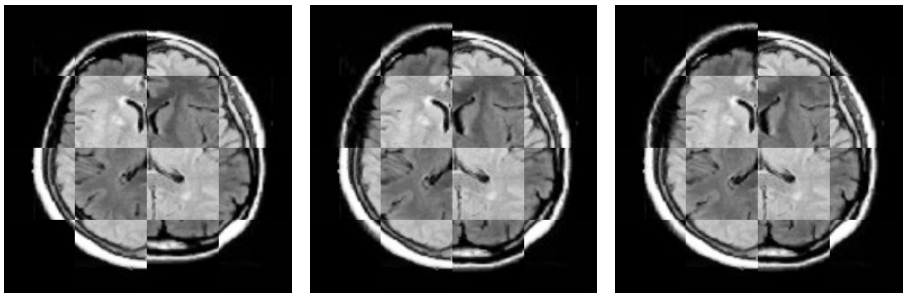
Table 2 Example 2: Sensitivity test on γ when $\alpha = 10^{-5}, \beta_1 = 50$ and $\beta_2 = 0.01$ for ϕ_2



(a) $T(\boldsymbol{y})$: $(\beta_1, \beta_2) = (50, 0.01)$ GFer=1.02, NGFer=0.997, MIer = 0.55

(b) $T(\boldsymbol{y})$: $(\beta_1, \beta_2) = (1, 10^{-4})$ GFer = 1.01, NGFer = 0.92, MIer = 0.597

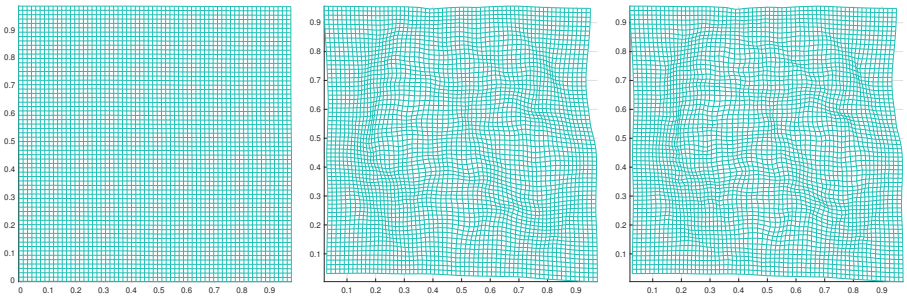
(c) $T(\boldsymbol{y})$: $(\beta_1, \beta_2) = (1, 10^{-5})$ GFer = 1.003, NGFer = 0.897, MIer = 0.599



(d) Overlay of $T(\boldsymbol{y})$ and R with $(\beta_1, \beta_2) = (50, 0.01)$

(e) Overlay of $T(\boldsymbol{y})$ and R with $(\beta_1, \beta_2) = (1, 10^{-4})$

(f) Overlay of $T(\boldsymbol{y})$ and R with $(\beta_1, \beta_2) = (1, 10^{-5})$



(g) $\boldsymbol{y}$: $(\beta_1, \beta_2) = (50, 0.01)$, $\min\det(\nabla \boldsymbol{y}) = 0.9752$

(h) $\boldsymbol{y}$: $(\beta_1, \beta_2) = (1, 10^{-4})$, $\min\det(\nabla \boldsymbol{y}) = 0.1944$

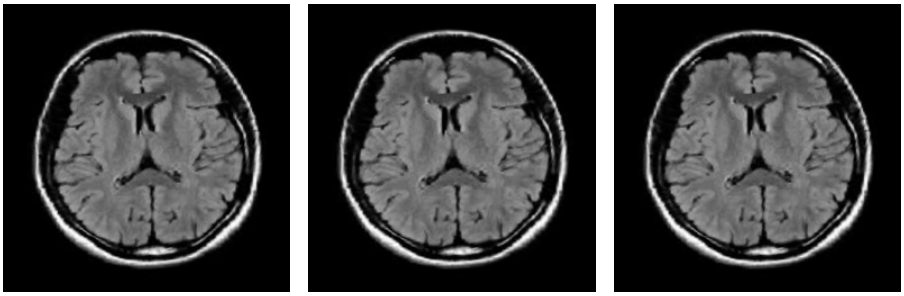
(i) $\boldsymbol{y}$: $(\beta_1, \beta_2) = (1, 10^{-5})$, $\min\det(\nabla \boldsymbol{y})$=**–0.264**

Figure 10 Example 2: a pair of MRI images (T1 and T2). The deformed templates and transformations are obtained by the model made up of NGF and the first- and second-order regularizers. Here, we do not impose a control term (so the third column gives the wrong results)

Appendix A. Computation of the vector $\vec{\boldsymbol{r}}(U)$. First of all, denote the 3 vertices of this triangle by $V_1 = \boldsymbol{x}^{1,1}$, $V_2 = \boldsymbol{x}^{2,1}$ and $V_4 = \boldsymbol{x}^{1,2}$ in Figure 2. Set $\mathbf{L}(V_1) = (u_1^{1,1}, u_2^{1,1})$, $\mathbf{L}(V_2) = (u_1^{2,1}, u_2^{2,1})$ and $\mathbf{L}(V_4) = (u_1^{1,2}, u_2^{1,2})$ at the vertex pixels. Here the linear approximations are $\mathbf{L}(x_1, x_2) = (a_1x_1 + a_2x_2 + a_3, a_4x_1 + a_5x_2 + a_6)$.

After substituting V_1, V_2 and V_5 into $\mathbf{L}$, we get

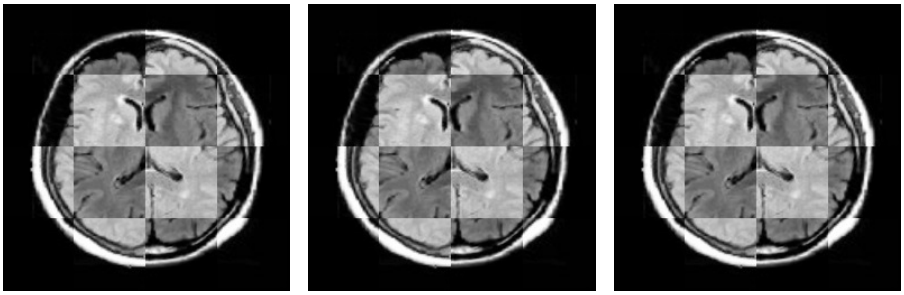
$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \frac{1}{\det} \begin{pmatrix} x_2^1 - x_2^2 & -x_2^1 + x_2^2 \\ -x_1^2 + x_1^1 & x_1^1 - x_1^1 \end{pmatrix} \begin{pmatrix} u_1^{1,1} - u_1^{1,2} \\ u_1^{2,1} - u_1^{1,2} \end{pmatrix}, \tag{5.1}$$



(a) $T(\boldsymbol{y})$ with $\gamma = 0.1$, GFer=0.74, NGFer = 0.96, MIer = 0.7823

(b) $T(\boldsymbol{y})$ with $\gamma = 1$, GFer=0.73, NGFer = 0.96, MIer = 0.7833

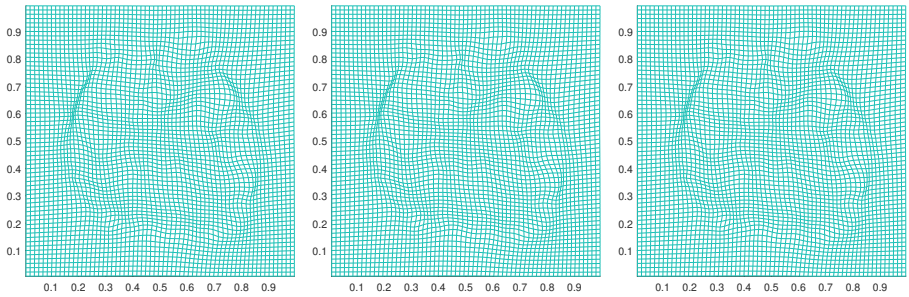
(c) $T(\boldsymbol{y})$ with $\gamma = 10$, GFer=0.73, NGFer = 0.96, MIer = 0.7824



(d) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 0.1$

(e) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 1$

(f) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 10$



(g) $\boldsymbol{y}$ with $\gamma = 0.1$, $\min\det(\nabla\boldsymbol{y}) = 0.1188$

(h) $\boldsymbol{y}$ with $\gamma = 1$, $\min\det(\nabla\boldsymbol{y}) = 0.1494$

(i) $\boldsymbol{y}$ with $\gamma = 10$, $\min\det(\nabla\boldsymbol{y}) = 0.2339$

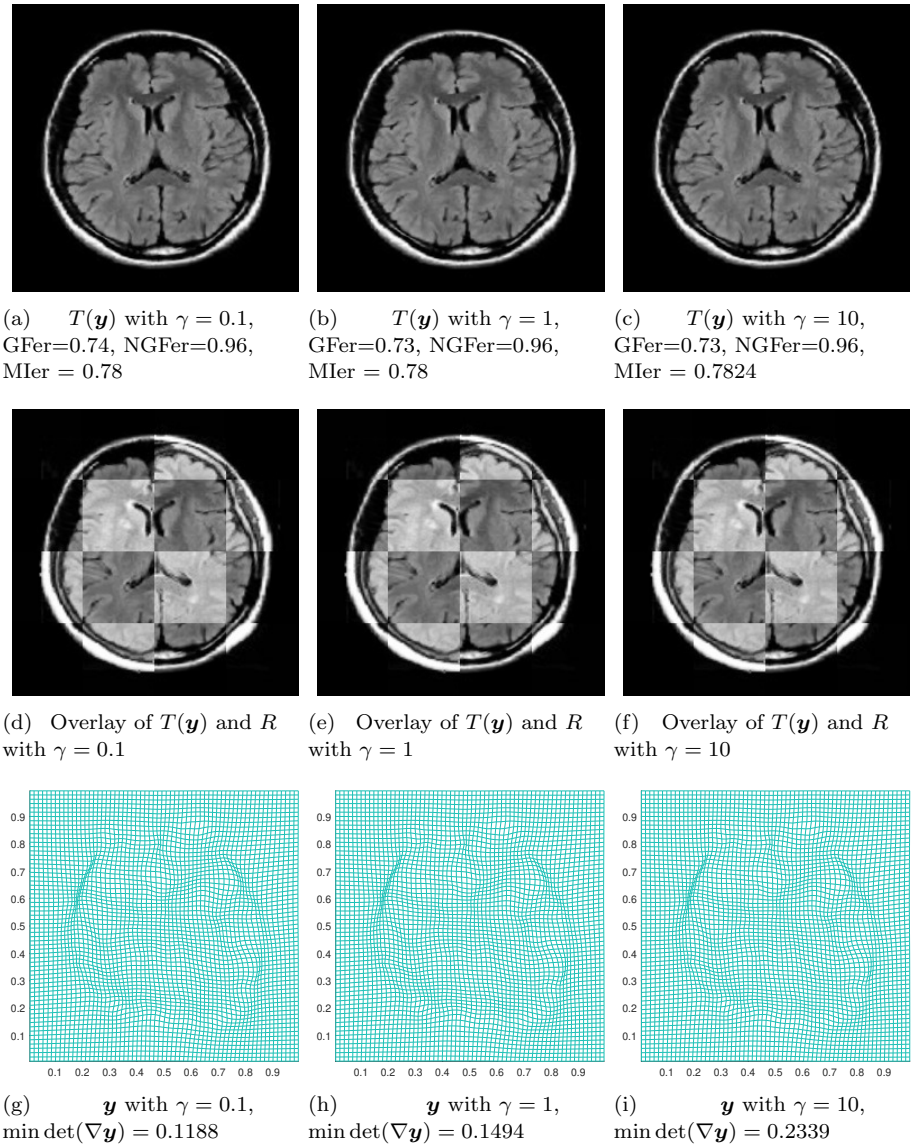
Figure 11 Example 2: a pair of MRI images (T1 and T2). By using the control term with ϕ_1, the resulting transformation is diffeomorphic and the deformed template is also visually satisfactory

$$\begin{pmatrix} a_4 \\ a_5 \end{pmatrix} = \frac{1}{\det} \begin{pmatrix} x_2^1 - x_2^2 & -x_2^1 + x_2^2 \\ -x_1^2 + x_1^1 & x_1^1 - x_1^1 \end{pmatrix} \begin{pmatrix} u_2^{1,1} - u_2^{1,2} \\ u_2^{2,1} - u_2^{1,2} \end{pmatrix}, \tag{5.2}$$

where $\det = \begin{vmatrix} x_1^1 - x_1^1 & x_2^1 - x_2^2 \\ x_1^2 - x_1^1 & x_2^1 - x_2^2 \end{vmatrix}$.

According to (5.1) and (5.2), we can formulate two matrices $D1 \in \mathbb{R}^{2n^2 \times (n+1)^2}$ and $D2 \in \mathbb{R}^{2n^2 \times (n+1)^2}$ such that

$$A_{31} = [D1, -D2],\ A_{32} = [D2, D1],\ A_{33} = [D1, D2],\ A_{34} = [D2, -D1].$$



(a) $T(\boldsymbol{y})$ with $\gamma = 0.1$, GFer=0.74, NGFer=0.96, MIer = 0.78

(b) $T(\boldsymbol{y})$ with $\gamma = 1$, GFer=0.73, NGFer=0.96, MIer = 0.78

(c) $T(\boldsymbol{y})$ with $\gamma = 10$, GFer=0.73, NGFer=0.96, MIer = 0.7824

(d) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 0.1$

(e) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 1$

(f) Overlay of $T(\boldsymbol{y})$ and R with $\gamma = 10$

(g) $\boldsymbol{y}$ with $\gamma = 0.1$, $\min\det(\nabla \boldsymbol{y}) = 0.1188$

(h) $\boldsymbol{y}$ with $\gamma = 1$, $\min\det(\nabla \boldsymbol{y}) = 0.1494$

(i) $\boldsymbol{y}$ with $\gamma = 10$, $\min\det(\nabla \boldsymbol{y}) = 0.2339$

Figure 12 Example 2: a pair of MRI images (T1 and T2). By using the control term with ϕ_2, the resulting transformation is diffeomorphic and the deformed template is also visually satisfied

Then using the Hadamard product $\odot$, we get a compact form for

$$\begin{cases} \vec{\boldsymbol{r}}^1(U) = A_{31}U \odot A_{31}U + A_{32}U \odot A_{32}U, \\ \vec{\boldsymbol{r}}^2(U) = 1./((A_{33}U + 2) \odot (A_{33}U + 2) + A_{34}U \odot A_{34}U), \\ \vec{\boldsymbol{r}}(U) = \vec{\boldsymbol{r}}^1 \odot \vec{\boldsymbol{r}}^2 \quad \in \mathbb{R}^{2n^2 \times 1}. \end{cases} \tag{5.3}$$

Appendix B. The Global Convergence of Algorithm 1. In order to discuss the global convergence result of Algorithm 1 for the discretized optimization problem (3.7), we first review two lemmas.

Lemma 5.1 ([46]) *For the unconstrained optimization problem*

$$\min_U J(U)$$

let an iterative sequence be defined by $U^{k+1} = U^k + \theta\delta U^k$, *where* $\delta U^k = -(H^k)^{-1}d_J(U^k)$ *and* θ *is obtained by Armijo condition. Assume that the following three conditions are met:*

(i) d_J *is Lipschitz continuous;*

(ii) the matrices H^k *are SPD;*

(iii) there exist constants $\bar{\kappa}$ *and* M *such that the condition number* $\kappa(H^k) \leq \bar{\kappa}$ *and the norm* $||H^k|| \leq M$ *for all* i.

Then either $J(U^k)$ *is unbounded from below or*

$$\lim_{i\to\infty} d_J(U^k) = 0 \tag{5.4}$$

and hence any limit point of the sequence of iterates is a stationary point.

Lemma 5.2 *Consider the matrix* $H = H_1 + H_2 + H_3$ *comprised of* 3 *submatrices* $\{H_j\}$. *If* H_1 *and* H_2 *are symmetric positive semi-definite and* H_3 *is SPD, then* H *is SPD with* $\lambda_{H_3} \leq \lambda_H$, *where* λ_{H_3} *and* λ_H *are the minimum eigenvalues of* H_3 *and* H *separately.*

Proof According to Rayleigh quotient, we can find a vector v such that

$$\lambda_H = \frac{v^T H v}{v^T v}. \tag{5.5}$$

Then we have

$$\lambda_{H_3} \leq \frac{v^T H_1 v}{v^T v} + \frac{v^T H_2 v}{v^T v} + \frac{v^T H_3 v}{v^T v} = \frac{v^T H v}{v^T v} = \lambda_H, \tag{5.6}$$

which completes the proof. □

In addition, define an important set $\mathcal{X} := \{U|\vec{\boldsymbol{r}}(U)_\ell \leq 1-\epsilon, 1 \leq \ell \leq 2n^2\}$ for small ϵ. So $U \in \mathcal{X}$ means that the transformation is diffeomorphic. Under the suitable γ, we assume that each U^k generated by Algorithm 1 is in the $\mathcal{X}$.

Theorem 5.3 *Assume that* T *and* R *are twice continuously differentiable. For (3.7), by using Algorithm 1, we obtain*

$$\lim_{i\to\infty} d_J(U^k) = 0 \tag{5.7}$$

and hence any limit point of the sequence of iterates produced by Algorithm 1 is a stationary point.

Proof It suffices to show that Algorithm 1 satisfies the requirements of Lemma 5.1. Recall $\vec{\boldsymbol{r}}(U)$ and we can see that it is continuous. Here, we use the Dirichlet boundary conditions and then $\|U\|$ is bounded. Hence, $\vec{\boldsymbol{r}}(U)$ is a continuous mapping from a compact set to $\mathbb{R}^{2n^2\times 1}$ and $\vec{\boldsymbol{r}}(U)$ is proper. So for some small $\epsilon > 0$, $\mathcal{X}$ is compact.

Firstly, we show that in $\mathcal{X}$, d_J of (3.7) is Lipschitz continuous. The term $\boldsymbol{\phi}(\vec{\boldsymbol{r}}(U))e^T$ in the (3.7) is twice continuously differentiable with respect to $U \in \mathcal{X}$. In addition, T and R are twice continuously differentiable. So (3.7) is twice continuously differentiable with respect to $U \in \mathcal{X}$ and d_J is Lipschitz continuous.

Secondly, we show that in $\mathcal{X}$, $H^k = \hat{H}_1^k + H_2^k + \hat{H}_3^k$ is SPD. By the construction of $\hat{H}_1^k$ and $\hat{H}_3^k$, they are symmetric positive semi-definite. H_2^k is symmetric positive definite under the Dirichlet boundary conditions. Consequently, according to Lemma 5.2, H^k is SPD.

Thirdly, we show that both $\kappa(H^k)$ and $\|H^k\|$ are bounded. We notice that in each iteration, H_2^k is constant and we can set $\|H_2^k\| = M_2$. For $\hat{H}_1^k$, we get its upper bound M_1 because T is twice continuously differentiable and $\mathcal{X}$ is compact. ϕ is also twice continuously differentiable with respect to $U \in \mathcal{X}$, then we have $\|\hat{H}_3^k\| \leq M_3$. Hence, we have

$$\|H^k\| \leq \|\hat{H}_1^k\| + \|H_2^k\| + \|\hat{H}_3^k\| \leq M_1 + M_2 + M_3. \tag{5.8}$$

So set $M = M_1 + M_2 + M_3$ and $\|H^k\| \leq M$. Set σ as the minimum eigenvalue of H_2^k. According to Lemma 5.2, the smallest eigenvalue $\lambda_{\min}$ of H^k should be larger than σ. The largest eigenvalue $\lambda_{\max}$ of H^k should be smaller than M due to $\lambda_{\max} \leq \|H^k\|$. So the conditional number of H^k is smaller than $\frac{M}{\sigma}$.

Finally, we can find that (3.7) has lower bound 0. Hence, by applying Lemma 5.1, we complete the proof. □

References

[1] B. FISCHER, J. MODERSITZKI. *Ill-posed medicine: an introduction to image registration*. Inverse Problems, 2008, **24**(3): 034008.

[2] F. GIGENGACK, L. RUTHOTTO, M. BURGER. *Motion correction in dual gated cardiac pet using mass-preserving image registration*. IEEE Transactions on Medical Imaging, 2012, **31**: 698–712.

[3] J. MODERSITZKI. *FAIR: Flexible Algorithms for Image Registration*. SIAM, Philadelphia, USA, 2009.

[4] F. P. OLIVEIRA, J. M. R. TAVARES. *Medical image registration: a review*. Computer Methods in Biomechanics and Biomedical Engineering, 2014, **17**: 73–93.

[5] A. SOTIRAS, C. DAVATZIKOS, N. PARAGIOS. *Deformable medical image registration: A survey*. IEEE Transactions on Medical Imaging, 2013, **32**: 1153–1190.

[6] Ke CHEN, L. M. LUI, J. MODERSITZKI. *Image and Surface Registration*. vol. 20 of Elsevier Handbook of Numerical Analysis (2019), ch. 15, pp. 579–611.

[7] M. BURGER, J. MODERSITZKI, L. RUTHOTTO. *A hyperelastic regularization energy for image registration*. SIAM J. Sci. Comput., 2013, **35**(1): B132–B148.

[8] N. CHUMCHOB, Ke CHEN. *Improved variational image registration model and a fast algorithm for its numerical approximation*. Numer. Methods Partial Differential Equations, 2012, **28**(6): 1966–1995.

[9] M. DROSKE, W. RING. *A Mumford-Shah level-set approach for geometric image registration*. SIAM J. Appl. Math., 2006, **66**(6): 2127–2148.

[10] M. DROSKE, M. RUMPF. *A variational approach to nonrigid morphological image registration*. SIAM J. Appl. Math., 2004, **64**(2): 668–687.

[11] S. HENN. *A multigrid method for a fourth-order diffusion equation with application to image processing*. SIAM J. Sci. Comput., 2005, **27**(3): 831–849.

[12] A. MANG, G. BIROS. *An inexact Newton-Krylov algorithm for constrained diffeomorphic image registration*. SIAM J. Imaging Sci., 2015, **8**(2): 1030–1069.

[13] A. MANG, G. BIROS. *Constrained H^1-regularization schemes for diffeomorphic image registration*. SIAM Journal on Imaging Sciences, 2016, **9**: 1154–1194.

[14] Chen XING, Peihua QIU. *Intensity-based image registration by nonparametric local smoothing*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, **33**: 2081–2092.

[15] N. CHUMCHOB. *Vectorial total variation-based regularization for variational image registration*. IEEE Trans. Image Process, 2013, **22**(11): 4551–4559.

[16] Wenrui HU, Yuan XIE, Lin LI, Wensheng ZHANG. *A total variation based nonrigid image registration by combining parametric and non-parametric transformation models*. Neurocomputing, 2014, **144**: 222–237.

[17] B. FISCHER, J. MODERSITZKI. *Fast Diffusion Registration*. Contemp. Math., 313, Amer. Math. Soc., Providence, USA, 2002.

[18] I. YANOVSKY, C. LE GUYADER, A. LEOW, et al. *Unbiased volumetric registration via nonlinear elastic regularization*. in Mathematical Foundations of Computational Anatomy (online), Web https://hal.archives-ouvertes.fr/MFCA08, 2008.

[19] B. FISCHER, J. MODERSITZKI. *Curvature based image registration*. J. Math. Imaging Vision, 2003, **18**(1): 81–85.

[20] N. CHUMCHOB, Ke CHEN, C. BRITO-LOEZA. *A fourth-order variational image registration model and its fast multigrid algorithm*. Multiscale Model. Simul., 2011, **9**(1): 89–128.

[21] M. IBRAHIM, Ke CHEN, C. BRITO-LOEZA. *A novel variational model for image registration using Gaussian curvature*. Geom. Imaging Comput., 2014, **1**(4): 417–446.

[22] Jianping ZHANG, Ke CHEN. *Variational image registration by a total fractional-order variation model*. J. Comput. Phys., 2015, **293**: 442–461.

[23] Jin ZHANG, Ke CHEN, Bo YU. *An improved discontinuity-preserving image registration model and its fast algorithm*. Appl. Math. Model., 2016, **40**(23-24): 10740–10759.

[24] Jin ZHANG, Ke CHEN, Bo YU. *A novel high-order functional based image registration model with inequality constraint*. Comput. Math. Appl., 2016, **72**(12): 2887–2899.

[25] E. HODNELAND, A. LUNDERVOLD, J. RORVIK, et al. *Normalized gradient*

elds for nonlinear motion correction of DCE-MRI time series. Computerized Medical Imaging and Graphics, 2014, **38**: 202–210.

[26] L. KÖNIG, J. RÜHAAK. *A fast and accurate parallel algorithm for non-linear image registration using normalized gradient fields*. in 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE, 2014, 580–583.

[27] J. RÜHAAK, L. KÖNIG, M. HALLMANN, et al. *A fully parallel algorithm for multimodal image registration using normalized gradient fields*. in 2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI), IEEE, 2013, 572–575.

[28] A. ANGELOV, M. WAGNER. *Multimodal image registration by elastic matching of edge sketches via optimal control*. J. Ind. Manag. Optim., 2014, **10**(2): 567–590.

[29] A. THELJANI, Ke CHEN. *An augmented Lagrangian method for solving a new variational model based on gradients similarity measures and high order regularization for multimodality registration*. Inverse Probl. Imaging, 2019, **13**(2): 309–335.

[30] F. MAES, A. COLLIGNON, D. VANDERMEULEN. *Multimodality image registration by maximization of mutual information*. IEEE Transactions on Medical Imaging, 1997, **16**: 187–198.

[31] J. P. PLUIM, J. A. MAINTZ, M. A. VIERGEVER. *Mutual-information-based registration of medical images: a survey*. IEEE Transactions on Medical Imaging, 2003, **22**: 986–1004.

[32] P. VIOLA, W. M. WELLS III. *Alignment by maximization of mutual information*. International Journal of Computer Vision, 1997, **24**(2): 137–154.

[33] Yunmei CHEN, Jiangli SHI, M. RAO. *Deformable multi-modal image registration by maximizing Rényi's statistical dependence measure*. Inverse Probl. Imaging, 2015, **9**(1): 79–103.

[34] Dongyang KUANG. *On reducing negative Jacobian determinant of the deformation predicted by deep registration networks*. ArXiv 1907.00068, 2019. (https://arxiv.org/pdf/1907.00068.pdf).

[35] T. THOMPSON, Ke CHEN. *An effective diffeomorphic model and its fast multigrid algorithm for registration of lung CT images*. J. Computational Methods in Applied Mathematics, 2019, pp. (https://doi.org/10.1515/cmam-2018-0126).

[36] E. HABER, J. MODERSITZKI. *Numerical methods for volume preserving image registration*. Inverse Problems, 2004, **20**(5): 1621–1638.

[37] E. HABER, J. MODERSITZKI. *Image registration with guaranteed displacement regularity*. International Journal of Computer Vision, 2007, **71**: 361–372.

[38] Lok Ming LUI, Tsz Ching NG. *A splitting method for diffeomorphism optimization problem using Beltrami coeffcients*. Journal of Scientiec Computing, 2015, **63**: 573–611.

[39] Daoping ZHANG, Ke CHEN. *A novel diffeomorphic model for image registration and its algorithm*. Journal of Mathematical Imaging and Vision, 2018, **60**: 1261–1283.

[40] Jun ZHANG. *Inverse-consistent deep networks for unsupervised deformable image registration*. arXiv:1809.03443, 2018.

[41] J. MODERSITZKI. *Numerical Methods for Image Registration*. Oxford Science Publications. Oxford University Press, New York, 2004.

[42] X. P. ZHOU. *Weak lower semicontinuity of a functional with any order*. J. Math. Anal. Appl., 1998, **221**(1): 217–237.

[43] G. LEONI. *A First Course In Sobolev Spaces*. American Mathematical Society, Providence, USA, 2017.

[44] R. BARRETT, M. BERRY, Tony F. CHAN, J. DEMMEL , J. DONATO, J. DONGARRA , V. EIJKHOUT, R. POZO, C. ROMINE, H. VAN DER VORST. *Templates For the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, USA, 1994.

[45] C. C. PAIGE, M. A. SAUNDERS. *Solution of sparse indefinite systems of linear equations*. SIAM Journal on Numerical Analysis, 1975, **12**: 617–629.

[46] C. T. KELLEY. *Iterative Methods For Optimization*. SIAM, Philadelphia, USA, 1999.

[47] J. NOCEDAL, S. J. WRIGHT. *Numerical Optimization*. Springer, New York, 2006.

[48] Wenyu SUN, Yaxiang YUAN. *Optimization Theory and Methods*. Springer, New York, 2006.