An Augmented Lagrangian Method for Solving a New Variational Model based on Gradients Similarity Measures and High Order Regularization for Multimodality Registration

Anis Theljani and Ke Chen

Abstract. In this work we propose a variational model for multi-modal image registration. It minimizes a new functional based on using reformulated normalized gradients of the images as the fidelity term and higher-order derivatives as the regularizer. We first present a theoretical analysis of the proposed model. Then, to solve the model numerically, we use an augmented Lagrangian method (ALM) to reformulate it to a few more amenable subproblems (each giving rise to an Euler-Lagrange equation that is discretized by finite difference methods) and solve iteratively the main linear systems by the fast Fourier transform; a multilevel technique is employed to speed up the initialisation and avoid likely local minima of the underlying functional. Finally we show the convergence of the ALM solver and give numerical results of the new approach. Comparisons with some existing methods are presented to illustrate its effectiveness and advantages.

Key words. Variational model; Optimization; Multi-modality images; Similarity measures; Mapping; High order regularisation; Inverse Problem; Augmented Lagrangian; Multilevel.

AMS subject classifications.

1. Introduction. Image registration consists in 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 of astronomy, optics, biology, chemistry, medical imaging and remote sensing and particularly in medical imaging. For an overview of image registration methodology and approaches, we refer to [11, 23, 24, 31]. Here, we focus on deformable image registration for multi-modality images using variational approaches which belong to the class of the widely used methods ([2, 5, 7, 14, 21, 22, 36]) and aim to find a better gradients-based model than the standard gradient models.

It is informative to illustrate the notation of the image registration modelling by considering a pair of mono-modal images: Given a fixed image (also called reference) and a moving image (also called template), which are represented by scalar functions $T, R : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$, find a reasonable geometric transformation $\varphi(u)(x) = x + u(x), \ u : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that:

$$T[\varphi(u)] = T(x + u(x)) = R.$$  

This is an equation of the unknown displacement field $u$, which is supposed to be sought in a properly chosen functional space. The reconstruction model (1.1) is an ill-posed inverse problem and thus regularisation techniques are needed to overcome ill-posedness. Generally, the regularisation technique turns an ill-posed problem such as model (1.1) into a well-posed one which minimizes an energy compromised of a regularisation term (mostly a semi-norm of
a functional space that is fixed a priori) and a data fidelity term. In summary, the desired
displacement \( u \), in some appropriate space \( \mathcal{H} \), is a minimizer of the following joint energy
functional:

\[
\min_{u \in \mathcal{H}} \{ J(u) = S(u) + \frac{\lambda}{2} D(T(u), R) \}.
\]

This model may be used for registering both mono-modality and multi-modality images.

Here in (1.2), the first term \( S(u) \) is a regularisation term which controls the smoothness of
\( u \) and reflects our expectations by penalising unlikely transformations. Many works tackled the
question of how to choose the best regularisation term that gives the more possible plausible
transformation. Various regularizers have been proposed, such as first-order derivatives based
on total variation [4, 16], diffusion [9] and elastic regularizer registration models and higher-
order derivatives-based on linear curvature [10], mean curvature [6] and Gaussian curvature
[17] models; we can refer also to [5, 22, 39, 40, 41].

The second term \( D(T(u), R) \) is a fidelity measure, which quantifies distance or similarity
of the transformed template image \( T(u) \) and the reference \( R \), whereas \( \lambda \) is a positive weight
which controls the trade-off between them. In the case of mono-modal images, the fixed and
the moving images have similar features and same intensity ranges. Thus, either the \( L^1 \)–
distance (Sum of Absolute Differences) \( D = \| T - R \|_1 \) or the well-known choice \( L^2 \)–
distance (Sum of Squared Differences) between \( R \) and \( T(u) \) i.e. \( D = \| T - R \|_2^2 = \int_\Omega (T(u) - R)^2 dx \) may
be used as a similarity measure. Clearly such a measure only makes sense for mono-modal
images.

For a pair of multi-modal images \( T, R \) (generated from independent imaging techniques),
unfortunately, one cannot minimize \( \| T - R \| \) since values of \( T, R \) are not directly comparable.
That is, only the patterns of \( T, R \) bear some resemblance to each other, not their values (so
called intensity values). Therefore, intensities of the same object in different images are not
similar which makes the problem much harder than the mono-modality case. Hence many
good models as from [23] for mono-modal images and also the elegant mathematical approach
of optimal transport [8] cannot be used. For multi-modal images, various similarity measures
have been used and include Mutual Information [20, 26, 33] and Normalised Gradient Field
[15, 18, 30]. Recently [3] proposed a cross-correlation similarity measure based on reproducing
kernel Hilbert spaces and found advantages over Mutual Information. Below, we briefly review
these two commonly used measures: mutual information and normalized gradient fields.

**Mutual Information (MI).** It takes its origin from the theory of information and was firstly
proposed in [33]. Several variants of MI approach were proposed in recent years (see [20,
26]), showcasing its great capability as well as limitations. The basic idea behind MI is the
comparison of the histograms of the two images instead of comparing their intensities. The
Mutual information between the two images if given by the following quantity:

\[
D_{MI}(T(u), R) = -\int_{\mathbb{R}^2} p_{T,R}(t,r) \log \frac{p_{T,R}(t,r)}{p_T(t)p_R(r)} \, dt\,dr,
\]

where \( p_T, p_R \) are probability distributions of the gray values in \( R \) and \( T \), whereas \( p_{T,R} \) is the
joint probability of the gray values which can be derived from the joint histogram. As the
MI measure involves histograms, its inherent disadvantages are how to choose the size of bins
and how to remedy the lack of spatial relationships to avoid mis-registrations. In addition, the measure also fails when features with different intensities in the first image have similar intensities in the second one [19], which is the case in perfusion imaging.

**Normalised Gradient Field (NGF).** The basic idea of the Normalised Gradient Field (NGF) [15, 18, 30] is the use of a derived information from the image intensity, i.e., the gradient. Similarity measures depending in the gradients or geometry of the images, which naturally encode information about the shape, can be better. The key idea behind the NGF measure is to align the gradients \( \nabla T(u) \) and \( \nabla R \) by minimizing the cosines distance between them. More precisely, on each point \( x \in \Omega \), try to find a displacement \( u(x) \) such that \( \cos \Theta = 1 \) where \( \Theta \) is the angle between \( \nabla T(x + u(x)) \) and \( \nabla R(x) \). Therefore, the NGF consists in minimization of the following energy:

\[
D_{NGF}(T(u), R) = \int_{\Omega} (1 - (\cos \Theta)^2) \, dx = \int_{\Omega} (1 - (\nabla_n T(u) \cdot \nabla_n R)^2) \, dx, \tag{1.4}
\]

where \( \nabla_n T(u) = \nabla T(u)/\|\nabla T(u)\| \) and \( \nabla_n R = \nabla R/\|\nabla R\| \) are normalised unit vectors. As the NGF uses the product scalar between the two vectors \( \nabla_n R \) and \( \nabla_n T(u) \), it will not work well when the gradients are null or very weak. In other words, suppose that in a large region of the image \( T \), we have \( \nabla T \perp \nabla R \) and then \( 1 - (\nabla_n T \cdot \nabla_n R)^2 \approx 1 \), which means that solving the optimization problem (1.2) is equivalent to only smoothing the deformation \( u \) in this region whereas the similarity measure does not play a role in the energy, which is not reliable. As an example, we consider the images in the Fig 1(a-b) where \( \nabla_n T \cdot \nabla_n R = 0 \) a.e in \( \Omega \) due to one of \( \nabla_n T, \nabla_n R \) being zero, we see that if we use the NGF in (1.2), there is no change in the template image because of the reason mentioned before, so \( T(u) \), obtained using the NGF as measure, shown in Fig 1(c) is not correct. If we use the ratio \#N of the number of pixels where \( \nabla_n T \cdot \nabla_n R \neq 0 \) over the total number of pixels, we have observed the current NGF would not give a good registration result if \( \#N \leq 25\% \). In this work, believing in the elegance of geometric fitting, we aim to improve the above NGF for these cases. we are primarily motivated to explore the potential of normalised gradients beyond its standard form. Our question is whether or not a better normalised gradients-based model than the well-known form [15, 18, 30] is possible.

The outline of the paper is as follows. In Section 2, we propose our variational model which minimizes an energy with new similarity measures and we prove by variational techniques the existence of a minimizer. Section 3 is dedicated to the numerical solution of the proposed model by an augmented Lagrangian approach and analysis of convergence. Finally, Section 4 concerns the implementation and the presentation of several numerical examples to test the efficiency and robustness of the proposed approach.

### 2. The new multi-modality model.
Since our formulation consists of two building blocks: a similarity measure \( D \) and a regularization term \( S \), we now discuss our choice of regularizers and the distance measure. Because our emphasis is on the latter, almost all regularizers suitable for variational registration models of mono-modal images may also be used.

**Choice of Similarity Measure.** To motivate our proposed measure \( D \), consider the NGF example in Fig 1. For this specific example, note that where \( \nabla_n T \cdot \nabla_n R = 0 \) we have
Figure 1. Example of Reference and Template images where $\nabla_n T \cdot \nabla_n R = 0$ (or one of $\nabla_n T$, $\nabla_n R$ is zero) a.e in $\Omega$.

$\|\nabla_n T - \nabla_n R\| \neq 0$. This suggests a revised NGF model

$$\min_u \{ S(u) + \frac{\lambda}{2} D^{GF}(T(u), R) \},$$

with the new measure $D^{GF}$ replacing the standard NGF measure $D^{NGF}$:

$$D^{GF}(T(u), R) = \int_{\Omega} GF(T(u), R) dx,$$

where

$$GF(T(u), R) = |\nabla_n T - \nabla_n R|^2.$$

As expected, such a model can solve the example from Fig 1(a-b) with acceptable registration result shown in Fig 1(d). This suggests that a better choice of normalised gradients as similarity measure is possible for multi-modal registration scenario. Moreover, to enhance this idea, we use Fig. 2 to show that alignment of two vectors $X = \nabla T, Y = \nabla R$ from a large discrepancy on the left to the small discrepancy on the right amounts to minimization of the distance $|X| + |Y| - |X + Y|$ (which is similar to minimizing $\cos \theta(X, Y)$ as in $D^{NGF}$). Below we shall combine the ideas of minimizing both $|X - Y|$ and $|X| + |Y| - |X + Y|$.

Figure 2. Three examples of the triangle inequality for triangles with sides $X$, $Y$ and $Z$. The left example shows a case where $|Z|$ is much less than the sum $|X| + |Y|$ of the other two sides, and the right example shows a case where $|Z|$ is only slightly less than $|X| + |Y|$.
Choice of a Regularizer. As mentioned, there is a large class of possible regularizers that we could choose from. Here we choose a robust regularizer that allows large and smooth deformation, comprised of both first order and second order derivatives for the deformation field.

Based on the new measure, we propose to register the two functions \( R, T \) from different image modalities by solving the following minimization problem:

\[
\begin{align*}
\min_{u \in \mathcal{W}} \{ J_1(u) &= S(u) + \frac{\lambda}{2} D^{GF}(T(u), R) + \frac{\lambda}{2} D^{TM}(T(u), R), \\
\text{w.r.t } \mathcal{C}(u) &= \det (I + \nabla u) > 0,
\end{align*}
\]

(2.1)

where \( \mathcal{W} = W^{1,2}_0(\Omega) \cap W^{2,2}(\Omega) \), \( \mathcal{C}(u) = (1 + \frac{\partial u_1}{\partial x})(1 + \frac{\partial u_2}{\partial y}) - \frac{\partial u_1 \partial u_2}{\partial x} \partial y \) and

\[
S(u) = \frac{\alpha}{2} \int_\Omega |\nabla u|^2 dx + \frac{\alpha_1}{2} \int_\Omega |\nabla^2 u|^2 dx,
\]

(2.2)

\[
D^{GF}(T(u), R) = \int_\Omega GF(T(u), R) dx,
\]

(2.3)

\[
D^{TM}(T, R) = \int_\Omega TM(T(u), R) dx.
\]

(2.4)

where \( TM(T(u), R) = (|\nabla T(u)| + |\nabla R| - |\nabla T(u) + \nabla R|)^2 \). Here in the term \( D^{GF} \), we must use the normalized gradients rather than the usual gradients because the difference in the magnitude of gradients of \( R \) and \( T(u) \) is large in multi-modality images. Moreover, we can easily prove that minimizing the length of \( TM(T(u), R) = |\nabla T(u)| + |\nabla R| - |\nabla T(u) + \nabla R| \) is equivalent to minimize the angle \( \theta \) between the vectors \( \nabla T(u) \) and \( \nabla R \), which leads to the alignment of the edges of \( R \) and \( T(u) \); note that an alternative to minimizing the above \( TM \) is to minimize \( TM_n(T(u), R) = |\nabla_n T(u)| + |\nabla_n R| - |\nabla_n T(u) + \nabla_n R| \) based on normalized gradients. However, this will lead to a more difficult problem to solve numerically due to higher non-linearity. Our primary choice for regularization is the diffusion model [9] 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 transformation. As such, we add the regularizer based on second-order derivatives (LLT) to the model which allows to remove the need of any preregistration step of affine transformation. The second-order derivatives allows also getting smooth transformations [41]. The constraint \( \mathcal{C}(u) > 0 \) on the determinant in the minimization problem (2.1) guarantees that the resulting deformation field \( \varphi = x + u \) suffers no mesh folding and thus is physically plausible; see also [12, 13, 28]. Different alternatives were proposed to ensure invertibility by adding another regularisation term depending on the determinant of the transformation to the registration objective function; see [2].

Mathematical analysis of the proposed model. Most registration models are non-convex with respect to \( u \) and consequently, if solutions exist, there are local minimizers or solutions are generally not unique. Below we prove the existence of a minimizer for problem (2.1). Before stating the main result, we first consider the concept of Carathéodory functions.
Definition 2.1. Let $\Omega \subset \mathbb{R}^d$ be an open set and let $f : \Omega \times \mathbb{R}^n \times \mathbb{R}^{d \times n} \times \mathbb{R}^{d \times d \times n} \to [0, +\infty)$. Then $f$ is a Carathéodory function if:

1. $f(x, \cdot, \cdot, \cdot)$ is continuous for almost every $x \in \Omega$.
2. $f(x, u, \psi, \Theta)$ is measurable in $x$ for every $(u, \psi, \Theta) \in \mathbb{R}^n \times \mathbb{R}^{d \times n} \times \mathbb{R}^{d \times d \times n}$.

We will use some theory about integrals of higher-order. It also sets up assumptions with which our optimisation problem (2.1) admits a minimiser.

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

1. $f$ is a Carathéodory function.
2. $f(x, u, \psi, \Theta)$ is quasi-convex with respect to $\Theta$.
3. $0 \leq f(x, u, \psi, \Theta) \leq a(x) + C(|\psi|^p + |\Theta|^p)$ where $a(x) \in L^1(\Omega)$, $C > 0$.

Then $J(u)$ is weak lower semi-continuous (denoted by $\text{wlsc}$) in $W$.

To analyse the proposed model (2.1), it is convenient to rewrite the energy $J(\cdot)$ by merging all terms under one integral in the following form:

$$J(u) = \int_{\Omega} f(x, u, \nabla u, \nabla^2 u) \, dx,$$

where $f(x, u, \psi, \Theta) = \frac{\alpha}{2} |\psi|^2 + \frac{\alpha_1}{2} |\Theta|^2 + \frac{\lambda}{2} |\nabla_n T(u) - \nabla_n R|^2$

$$+ \frac{\lambda}{2} (|\nabla T(u)| + |\nabla R| - |\nabla T(u) + \nabla R|)^2.$$

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

Lemma 2.3. The energy functional $J(\cdot)$ is coercive and $\text{wlsc}$ in $W$.

Proof. The coercivity can easy obtained using the Poincaré inequality. In fact, the later guarantees that

$$||u||_W = (||\nabla u||_2^2 + ||\nabla^2 u||_2^2)^{1/2}$$

defines a norm in the space $W$. Using the positivity of $D^{GF}(T(u), R)$ and $D^{TM}(T(u), R)$, we have:

$$J(u) \geq \frac{\min(\alpha, \alpha_1)}{2} ||u||_W^2,$$

which directly gives the coercivity of $J(\cdot)$. For the weak lower semi-continuity, we now verify that the functions $f(\cdot)$ fulfills the assumptions in Lemma 2.2:

i) Since the gradient of the fixed and the moving image $\nabla R$ and $\nabla T(u)$ are assumed to be continuous, $f(\cdot)$ is Carathéodory function.

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

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

$$\frac{\lambda}{2} |\nabla_n T(u) - \nabla_n R|^2 \leq \frac{\lambda}{2} (|\nabla_n T(u)| + |\nabla_n R|)^2 \leq 2\lambda.$$
Moreover, using the fact that $|\nabla R|$ and $|\nabla T(u)|$ are bounded almost everywhere by a constant $c > 0$, we get

\begin{equation}
\frac{\lambda}{2}(|\nabla T(u)| + |\nabla R| - |\nabla T(u) + \nabla R|)^2 \leq \frac{\lambda}{2}(|\nabla T(u)| + |\nabla R|)^2 \leq 2\lambda c^2.
\end{equation}

Therefore, using inequalities (2.5) and (2.6), we have:

$$f(x, u, \psi, \Theta) = \frac{\alpha}{2} |\psi|^2 + \frac{\alpha_1}{2} |\Theta|^2 + \frac{\lambda}{2} |\nabla_T(u) - \nabla R|^2$$

$$+ \frac{\lambda}{2}(|\nabla T(u)| + |\nabla R| - |\nabla T(u) + \nabla R|)^2$$

$$\leq \frac{\alpha}{2} |\psi|^2 + \frac{\alpha_1}{2} |\Theta|^2 + 2\lambda c^2 + 2\lambda.$$

Then, the function $f(\cdot)$ fulfils the condition (iii) of Lemma 2.2 with $a(x) \equiv \lambda c^2 + 2\lambda$ which implies that the energy $J(\cdot)$, is $\text{wlsc}$ in $W$.

We are now ready to prove the existence of a solution for the minimization model (2.1). Based on Lemma 2.2 and Lemma 2.3, we have the following result:

**Proposition 2.4.** The minimization problem (2.1) admits at least one solution in the space $A = \{ u \in W; C_\epsilon(u) \geq 0 \}$ where $\epsilon > 0$ is a small parameter, $C_\epsilon(u) = C(u) - \epsilon$, and $C(\cdot)$ is given in (2.1).

**Proof.** Consider a minimizing sequence $(u_n)_n \subset A$ of $J(\cdot)$, i.e.,

$$J(u_n) \longrightarrow \inf_{u \in A} J(u).$$

The coercivity of $J(\cdot)$ guarantees that the sequence $(u_n)_n \in \mathbb{N}$ is uniformly bounded $W$. Thus, there exists a subsequence, still denoted $(u_n)_n \in \mathbb{N}$, such that $u_n \rightarrow u$ weakly in $W$. Using the weak lower semi-continuity of $J(\cdot)$, we obtain that the limit $u$ is a minimizer of $J(\cdot)$.

It remains to prove that $u$ fulfils the constraint $C(u) > 0$. Now, we show that $A$ is weakly closed subset of $W$. Let $u_k$ be a weakly convergent sequence to $u$ in $W$. From the definition of the space $W$, we have that $u_k$ is weakly convergent to $u$ in $W^{1,2}(\Omega)$ and $u_k$ is weakly convergent to $u$ in $W^{2,2}(\Omega)$. Moreover, as the sets $A_1 = \{ u \in W^{1,2}(\Omega); C_\epsilon(u) \geq 0 \}$ and $A_2 = \{ u \in W^{2,2}(\Omega); C_\epsilon(u) \geq 0 \}$ are weakly closed for $W^{1,2}$-topology and $W^{2,2}$-topology (see [27]), respectively, we get that $u \in A_1$ and $u \in A_2$. Then, the limit $u$ belongs to the intersection $A = A_1 \cap A_2$ and thus $A$ is weakly closed. Therefore, the minimizer $u$ belongs to the set $A$, i.e., $C(u) \geq \epsilon > 0$, which finishes the proof.

### 3. Augmented Lagrangian method (ALM)

The energies $J(\cdot)$ are highly non-linear, and their numerical resolution is a non-trivial task. Thus, we propose an Augmented Lagrangian Method (ALM) which is often used for solving constrained minimization problems by replacing the original problem by an unconstrained problem. The method is similar to the penalty method where the constraints are incorporated in the objective functional and the problem is solved using alternating minimization of the sub-problems; see [1, 29, 32, 35, 43, 44] for various successful applications.
3.1. ALM iterations. Introducing three intermediate variables \( K, p \) and \( n \) to reformulate (2.1), we solve the following constrained minimization problem:

\[
\begin{align*}
\min_{u,K,p,n} \{ S(u) + \frac{\lambda}{2} \int_{\Omega} (n - \nabla_n R)^2 dx + \frac{\lambda}{2} \int_{\Omega} (|p| + |\nabla R| - |m|)^2 dx \}, \\
\text{w.r.t } K = T(u), \ p = \nabla K, \ |p|n = p, \ m = p + \nabla R, \ C_s(u) \geq 0.
\end{align*}
\]

Then, the augmented Lagrangian functional corresponding to the constrained optimization problem (3.1) is defined as follows:

\[
\mathcal{L}_1(u, K, p, n, m, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = S(u) + \frac{\lambda}{2} \int_{\Omega} (n - \nabla_n R)^2 dx + \frac{\lambda}{2} \int_{\Omega} (|p| + |\nabla R| - |m|)^2 dx + \frac{r_2}{2} \int_{\Omega} (p - \nabla K)^2 dx + \frac{r_3}{2} \int_{\Omega} (p - |p|n)^2 dx + \frac{r_4}{2} \int_{\Omega} (p + \nabla R - m)^2 dx + \int_{\Omega} (T(u) - K) \lambda_1 dx + \int_{\Omega} (p - \nabla K) \cdot \lambda_2 dx + \int_{\Omega} (p - |p|n) \cdot \lambda_3 dx + \int_{\Omega} (p + \nabla R - m) \cdot \lambda_4 dx + \frac{r_1}{2} \int_{\Omega} (T(u) - K)^2 dx + \frac{1}{2\sigma} \int_{\Omega} C_s(u, \lambda_5) dx,
\]

where

\[
C_s(u, \lambda_5) = [\min\{0, \sigma C_s(u) - \lambda_5\}]^2 - \lambda_5^2,
\]

\( \sigma > 0 \) and \( \lambda_i, (i = 1, \ldots, 5) \) are the Lagrange multipliers. Since the optimisation problem (2.1) admits a minimizer, the previous augmented Lagrangian admits a saddle point \((u^*, K^*, p^*, n^*, m^*, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_5^*)\).

3.2. Discretization and sub-problems. The images and the displacement fields are discretized on a uniform mesh using vertex centred discretization. We assume that the discrete solution \( u_{i,j} = u(x_i, y_j), i = 1, \ldots, l, j = 1, \ldots, c \) have \( l \times c \) pixels, where \( l \) and \( c \) are the numbers of rows and columns in the image, respectively. Other quantities are set up similarly.

For sake of simplicity, we use a generic notation \( u \) for discussing discretization. For the discrete differential operators, we assume periodic boundary conditions for \( u \). By choosing periodic boundary conditions, the action of each of the discrete differential operators can be regarded as a circular convolution of \( u \) and allows the use of fast Fourier transform (see [25, 34, 38] for more details). The discrete gradient is an operator from \( \mathbb{R}^{l \times c} \) to \( \mathbb{R} \), given by

\[
\nabla u = (\partial_x u, \partial_y u) \text{ where } \partial_x \text{ and } \partial_y \text{ are forward difference operators defined as follows:}
\]

\[
\partial_x u = \begin{cases} 
    u(i+1,j) - u(i,j) & 1 \leq i < l, 1 \leq j \leq c, \\
    u(1,j) - u(i,j) & i = l, 1 \leq j \leq c,
\end{cases}
\]

\[
\partial_y u = \begin{cases} 
    u(i,j+1) - u(i,j) & 1 \leq i \leq l, 1 \leq j < c, \\
    u(i,1) - u(i,j) & 1 \leq i \leq l, j = c.
\end{cases}
\]

The discrete divergence is an operator from \( \mathbb{R}^{l \times c} \) to \( \mathbb{R} \) and, for \( n = (n_1, n_2) \), given by \( \nabla \cdot n = \)
Then, the discrete Laplace operator is given by $\Delta u = \text{div}(\nabla u)$. Similarly, we define the following (forward and backward) second-order discrete differential operators:

$$
\partial_{xx} u = \frac{1}{\Delta x^2} \begin{cases}
  u(l, j) - 2u(i, j) + u(i + 1, j) & i = 1, 1 \leq j \leq c, \\
  u(i - 1, j) - 2u(i, j) + u(i + 1, j) & 1 < i < l, 1 \leq j \leq c, \\
  u(i - 1, j) - 2u(i, j) + u(1, i) & i = l, 1 \leq j \leq c,
\end{cases}
$$

$$
\partial_{yy} u = \frac{1}{\Delta y^2} \begin{cases}
  u(i, c) - 2u(i, j) + u(i, j + 1) & 1 \leq i \leq l, j = 1, \\
  u(i, j - 1) - 2u(i, j) + u(i, j + 1) & 1 \leq i \leq l, 1 < j \leq c, \\
  u(i, j - 1) - 2u(i, j) + u(i, 1) & 1 \leq i \leq l, j = c,
\end{cases}
$$

$$
\partial_{xy} u = \frac{1}{\Delta x \Delta y} \begin{cases}
  u(i, j) - u(i + 1, j) - u(i, j + 1) + u(i + 1, j + 1) & 1 \leq i \leq l, 1 \leq j \leq c, \\
  u(i, j) - u(i, j + 1) - u(l, j + 1) + u(l, j + 1) & i = l, 1 \leq j \leq c, \\
  u(i, j) - u(i + 1, j) - u(i, 1) + u(i + 1, 1) & 1 \leq i \leq l, j = c, \\
  u(i, j) - u(l, i) - u(l, j) + u(l, c) & i = l, j = 1,
\end{cases}
$$

Based on the above operators, we define the following fourth-order differential operator:

$$
\nabla^2 u = \frac{1}{\Delta x^2} \partial_{xx} u + \frac{1}{\Delta y^2} \partial_{yy} u + \frac{1}{\Delta x \Delta y} \partial_{xy} u + \frac{1}{\Delta x \Delta y} \partial_{yx} u.
$$

Thus the first version of an ALM algorithm is shown in Algorithm 3.1.

In order to solve the optimisation problem (3.4) more efficiently, we now consider a decoupled version of all main variables for the solution. The minimization problem is decomposed into a number of sub-problems, each of which can be solved quickly. In particular, we split the problem into four (main) sub-problems. Then, an alternating minimization and iterative procedure is obtained and shown in Algorithm 3.2. We discuss next how to solve these sub-problems.

**The u-subproblem.** Fixing $K^k$, $p^k$, $m^k$, and $\lambda^k_i$ ($i = 1, \ldots, 5$), the $u$-subproblem consists in finding $u^{k+1}$ from solving the following minimization problem:

$$
\min_u \{ S(u) + \frac{\gamma_1}{2} \int_{\Omega} (T(u) - K^k)^2 dx + \int_{\Omega} (T(u) - K^k) \lambda^k_i dx + \frac{1}{2\sigma} \int_{\Omega} C_u(u, \lambda^k) dx \}.
$$
Algorithm 3.1 Augmented Lagrangian method

1. Initialization: \( u^0, K^0, p^0, n^0, m^0 \) and \( \lambda_1^0, \lambda_2^0, \lambda_3^0, \lambda_4^0 \) and \( \lambda_5^0 \).
2. Iterate for \( k = 1, 2, \ldots \) until a required tolerance:
   - compute an approximate minimizers \( u^{k+1}, K^{k+1}, p^{k+1}, n^{k+1} \) and \( m^{k+1} \) of the augmented Lagrangian functional with the fixed Lagrange multipliers \( \lambda_1^k, \lambda_2^k, \lambda_3^k, \lambda_4^k \) and \( \lambda_5^k \):
     \[
     \begin{bmatrix}
     u^{k+1}, K^{k+1}, p^{k+1}, n^{k+1}, m^{k+1}
     \end{bmatrix} = \arg\min_{u, K, p, n, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5} \mathcal{L}(u, K, p, n, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5).
     \]
   - Update Lagrange multipliers
     \[
     \begin{align*}
     \lambda_1^{k+1} &= \lambda_1^k + r_1(T(u^{k+1}) - K^{k+1}), \\
     \lambda_2^{k+1} &= \lambda_2^k + r_2(p^{k+1} - \nabla K^{k+1}), \\
     \lambda_3^{k+1} &= \lambda_3^k + r_3(p^{k+1} - |p^{k+1}|n^{k+1}), \\
     \lambda_4^{k+1} &= \lambda_4^k + r_4(m^{k+1} - p^{k+1} - \nabla R), \\
     \lambda_5^{k+1} &= \max\{0, \lambda_5^k - \sigma C_c(u^{k+1})\},
     \end{align*}
     \]

Algorithm 3.2 An more efficient solution procedure for alternating iterations

1. Initialization: \( \tilde{u}^0 = u^k, K^0 = K^k, \tilde{p}^0 = p^k, n^0 = n^k \) and \( \tilde{m}^0 = m^k \).
2. Iterate for \( k = 1, 2, \ldots \) until a required tolerance:
   - Set the Lagrange multipliers \( \lambda_1 = \lambda_1^k, \lambda_2 = \lambda_2^k, \lambda_3 = \lambda_3^k, \lambda_4 = \lambda_4^k \) and \( \lambda_5 = \lambda_5^k \);
   - Solve for \( l = 1, \ldots, \ell \) the following problems:
     \[
     \begin{align*}
     \tilde{u}^{l+1} &= \arg\min_{u} \mathcal{L}_1(u, \tilde{K}^l, \tilde{p}^l, \tilde{n}^l, m^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5), \\
     \tilde{K}^{l+1} &= \arg\min_{K} \mathcal{L}_1(\tilde{u}^{l+1}, K, \tilde{p}^l, \tilde{n}^l, m^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5), \\
     \tilde{p}^{l+1} &= \arg\min_{p} \mathcal{L}_1(\tilde{u}^{l+1}, \tilde{K}^{l+1}, p, \tilde{n}^l, m^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5), \\
     \tilde{n}^{l+1} &= \arg\min_{n} \mathcal{L}_1(\tilde{u}^{l+1}, \tilde{K}^{l+1}, \tilde{p}^{l+1}, n, m^k, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5), \\
     \tilde{m}^{l+1} &= \arg\min_{m} \mathcal{L}_1(\tilde{u}^{l+1}, \tilde{K}^{l+1}, \tilde{p}^{l+1}, \tilde{n}^{l+1}, m, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5).
     \end{align*}
     \]
   - Prepare for the next iteration by setting
     \[
     [u^{k+1}, K^{k+1}, p^{k+1}, n^{k+1}, m^{k+1}] = [\tilde{u}^{l+1}, \tilde{K}^{l+1}, \tilde{p}^{l+1}, \tilde{n}^{l+1}, \tilde{m}^{l+1}].
     \]

It is clear that the above minimization problem admits at least a solution \( u = (u_1, u_2) \) by solving the following system of PDEs in \( \Omega \):

\[
\begin{align*}
-\alpha \Delta u_1^{k+1} + \alpha_1 \text{div}\, u_1^{k+1} + r_1(T(u^{k+1}) - K^{k})\partial_x T(u^{k+1}) + \lambda_1^{k} \partial_x u_1^{k+1} + \partial_{u_1} C_s(u^{k+1}, \lambda_1^{k}) &= 0, \\
-\alpha \Delta u_2^{k+1} + \alpha_1 \text{div}\, u_2^{k+1} + r_1(T(u^{k+1}) - K^{k})\partial_y T(u^{k+1}) + \lambda_1^{k} \partial_y u_2^{k+1} + \partial_{u_2} C_s(u^{k+1}, \lambda_1^{k}) &= 0.
\end{align*}
\]
with the periodic boundary conditions on \( \partial \Omega \). To solve the previous non-linear PDEs, we use a fast time marching method, i.e., find \( \mathbf{u}^{k+1} = (u_1^{k+1}, u_2^{k+1}) \) which solves

\[
\begin{align*}
\begin{cases}
u_1^{k+1} - dt[\alpha \Delta u_1^{k+1} + \alpha_1 \text{div}^2 \nabla^2 u_1^{k+1}] = F_1(u_{\text{odd}}^{k+1}), & \text{in } \Omega, \\
u_2^{k+1} - dt[\alpha \Delta u_2^{k+1} + \alpha_1 \text{div}^2 \nabla^2 u_2^{k+1}] = F_2(u_{\text{odd}}^{k+1}), & \text{in } \Omega,
\end{cases}
\end{align*}
\]

where \( dt \) is the time step, \( u_{\text{odd}}^{k+1} \) is the solution at the previous iteration for the time marching method and

\[
F_1(u) = -dt[r_1(T(u) - R) \partial_x T(u) - \lambda_2^k \partial_x T(u) - \partial_{u_1} C_s(u, \lambda_2^k)] + u_1,
\]

\[
F_2(u) = -dt[r_1(T(u) - R) \partial_y T(u) - \lambda_2^k \partial_y T(u) - \partial_{u_2} C_s(u, \lambda_2^k)] + u_2.
\]

To solve the above fourth-order equations in each time step iteration, we use the 2-dimensional discrete Fourier transforms. In fact, we have:

\[
L \odot F(u_1^{k+1}) = F(F_1(u_{\text{odd}}^{k+1})), \quad \text{and} \quad L \odot F(u_2^{k+1}) = F(F_2(u_{\text{odd}}^{k+1})),
\]

where \( L = I + dt F(\Delta \cdot) + \alpha_1 dt F(\text{div}^2 \nabla^2 \cdot) \). The operator \( F(\cdot) \) is the Fourier transform and “\( \odot \)” means point-wise multiplication of matrices. Therefore, the discrete solutions \( u_1 \) and \( u_2 \) can be obtained by applying the inverse of the discrete two-dimensional Fourier transform to the previous equation and we have:

\[
(3.13) \quad u_1^{k+1} = F^{-1}\left(F(F_1(u_{\text{odd}}^{k+1})) \odot L\right) \quad \text{and} \quad u_2^{k+1} = F^{-1}\left(F(F_2(u_{\text{odd}}^{k+1})) \odot L\right),
\]

where “\( \odot \)” means point-wise division of matrices.

**Remark 1.** We emphasize that computing the determinant is a non-trivial task. A discretization which well ensures that the map is diffeomorphic is discussed in [2, 13] and is based on finite element method. In our case, we are not using this discretization in the numerical computation as we are solving a system of PDEs defined only on the nodal points. However, the discretization is used for computing the determinant after getting the solution to check if the obtained map is diffeomorphic.

**The \( K \)-subproblem.** Fixing \( \mathbf{u}^{k+1}, \mathbf{p}^k, \mathbf{n}^k, \mathbf{m}^k \) and \( \lambda_i^k \) \((i = 1, \cdots, 5)\), the \( K \)-problem involves the minimization of the following energy:

\[
\min_{K} \left\{ \frac{r_1}{2} \int_{\Omega} (T(u^{k+1}) - K)^2 dx + \frac{r_2}{2} \int_{\Omega} (\mathbf{p}^k - \nabla K)^2 dx \\
+ \int_{\Omega} (T(u^{k+1}) - K) \lambda_k^{-1} dx + \int_{\Omega} (\mathbf{p}^k - \nabla K) \cdot \lambda_2^k dx \right\}.
\]

This minimization problem is solved through its optimality condition:

\[
(3.14) \quad -r_2 \Delta K^{k+1} + r_1 K^{k+1} = r_1 T(u^{k+1}) - r_2 \text{div } \mathbf{p}^k - \text{div } \lambda_2^k + \lambda_k.
\]

We take advantage from the use of the 2-dimensional discrete Fourier transforms to compute \( K \). In fact, applying the Fourier transforms to

\[
LS \odot F(K) = F(RS),
\]
where “⊙” means point-wise multiplication of matrices, \( RS \) is the right side of (3.14) and
\[
LS = -r_2 F(\Delta \cdot) + r_1 I.
\]
Therefore, the discrete solution is given by:
\[
(3.15) \quad K = F^{-1} (F(RS) \odot LS),
\]
where \( F^{-1} (\cdot) \) is the inverse of the discrete two-dimensional Fourier transform.

The p-subproblem. Fixing \( u^{k+1}, K^{k+1}, n^k, m^k \) and \( \lambda^k \) (\( i = 1, \cdots, 5 \)), the p-subproblem consists in minimizing, w.r.t., \( p \), the following energy:
\[
(3.16) \quad \frac{r_2}{2} \int_{\Omega} (p - \nabla K^{k+1})^2 dx + \frac{r_3}{2} \int_{\Omega} (p - |p| n^k)^2 dx + \frac{r_4}{2} \int_{\Omega} (p + \nabla R - m^k)^2 dx
\]
\[
+ \int_{\Omega} (p - \nabla K^{k+1}) \cdot \lambda^k_2 dx + \int_{\Omega} (p - |p| n^k) \cdot \lambda^k_3 + \int_{\Omega} (p + \nabla R - m^k) \cdot \lambda^k_4 dx
\]
\[
+ \frac{\lambda}{2} \int_{\Omega} (|p| + |\nabla R| - |m|)^2 dx.
\]

It is challenging to solve the above p-minimization problem due to the non-differentiability of \( |p| \) in the quadratic term. To alleviate this situation, we consider a fixed-point formulation by lagging \( |p^k| n^k \) in the \( k^{th} \) iteration instead of the constraint \( p = |p| n^k \). Thus, a simple reformulation rewrites the above problem as an equivalent minimization problem:
\[
(3.17) \quad \min_{p} \int_{\Omega} \beta |p| dx + \frac{r_2 + r_3 + r_4 + \lambda}{2} \int_{\Omega} (p - C)^2 dx + Res,
\]
where the quantity \( Res \) does not depend on \( p \), \( \beta = -\lambda^k_3 \cdot n^k - \lambda(|m^k| - |\nabla R|) \) and
\[
(3.18) \quad C = \frac{r_2 \nabla K^{k+1} + r_3 |p^k| n^k + r_4 (m^k - \nabla R) - \lambda^k_2 - \lambda^k_3 - \lambda^k_4}{r_2 + r_3 + r_4 + \lambda}.
\]
The minimization problem (3.17) has a closed from solution which is explicitly given by the following shrinkage-like formula:
\[
(3.19) \quad p^{k+1} = \max \left\{ 1 - \frac{\beta}{(r_2 + r_3 + r_4 + \lambda)|C|}, 0 \right\} C.
\]

The n-subproblem. Fixing \( u^{k+1}, K^{k+1} \) and \( p^{k+1} \) and \( \lambda^k \) (\( i = 1, \cdots, 5 \)), the n-problem consists in solving the following minimization problem:
\[
\min_{n} \frac{\lambda}{2} \int_{\Omega} (n - \nabla_n R)^2 dx + \frac{r_2}{2} \int_{\Omega} (p^{k+1} - |p^{k+1}| n)^2 dx + \int_{\Omega} (p^{k+1} - |p^{k+1}| n) \cdot \lambda^k_3 dx.
\]
The above problem has a closed from solution which is explicitly given by:
\[
(3.20) \quad n = \frac{\lambda \nabla_n R + r_3 |p^{k+1}| p^{k+1} + |p^{k+1}| \lambda^k_3}{\lambda + r_3}.
\]
The $m$-subproblem. To find the optimal value of $m^{k+1}$, we solve the following optimisation sub-problem:

$$
\min_m \frac{\lambda}{2} \int_\Omega (|p^{k+1}| + |\nabla R| - |m|)^2 \, dx + \frac{r_1}{2} \int_\Omega (p^{k+1} + \nabla R - m)^2 \, dx 
+ \int_\Omega (p^{k+1} + \nabla R - m) \cdot \lambda_t^i \, dx.
$$

(3.21)

The above problem is equivalent to minimizing the following energy:

$$
\min_m -\lambda \int_\Omega (|p^{k+1}| + |\nabla R|)|m| \, dx + \frac{\lambda + r_1}{2} \int_\Omega (m - C)^2 \, dx + Res,
$$

where both $Res$ and $C$ do not depend on $p$, with $C$ given by:

$$
C = \frac{r_1(p^{k+1} + \nabla R)}{\lambda + r_1}.
$$

The solution is explicitly given by:

$$
m^{k+1} = \max \left\{ 1 + \frac{\lambda(|p^{k+1}| + |\nabla R|)}{(\lambda + r_1)|C|}, 0 \right\} C.
$$

(3.22)

Lemma 3.1 ([37]). Let $f : \mathbb{R} \to \mathbb{R}$ be a closed, proper and convex function. Let $(w_n)_{n \in \mathbb{N}}$ be a sequence of distinct functions in $\text{dom} f$ converging to $w^* \in \text{int}(\text{dom} f)$ and let $S_n \in \partial f(w_n)$. Then there exists a subsequence $(S_{n_k})_{k \in \mathbb{N}}$ that converges to the point $S^*$, where $S^* \in \partial f(w^*)$.

In the sequel, we give a partial result about the limit behaviour of the solutions generated by the ALM method. Let us consider the space:

$$
\mathcal{X} = \tilde{W} \times W_0^{1,2}(\Omega) \times \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega) \times \mathcal{L}^2(\Omega),
$$

where $\tilde{W} = \{u \in W, \text{div}^2 u, \nabla^2 u \in L^2(\Omega) \}$ and

$$
\mathcal{L}^2(\Omega) = \{w \in (L^2(\Omega))^2, \text{div} w \in L^2(\Omega) \}.
$$

Proposition 3.2. If the sequence $(u^k, K^k, p^k, n^k, \lambda_t^k, \lambda_2^k, \lambda_3^k, \lambda_4^k, \lambda_t^k) \in \mathcal{X}$, generated by the ALM method, converges to a point $(u^*, K^*, p^*, n^*, \lambda_t^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_t^*) \in \mathcal{X}$, then the limit point satisfies the following first-order optimality conditions:

$$
\begin{cases}
-\alpha \Delta u_t^* + \alpha_1 \text{div}^2 \nabla u_t^* + \lambda_1 \partial_x T(u^*) + \partial_u C_s(u^*, \lambda_t^*) = 0, & \text{in } \Omega, \\
-\alpha \Delta u_2^* + \alpha_1 \text{div}^2 \nabla u_2^* + \lambda_1 \partial_x T(u^*) + \partial_u C_s(u^*, \lambda_t^*) = 0, & \text{in } \Omega, \\
\text{div} \lambda_2^* - \lambda_1^* = 0, & \\
-\beta^* S_p^* + \sum_{i=2}^4 \lambda_t^i = 0, & \\
\lambda m^* - \lambda \nabla R - |p^*| \lambda_t^3 = 0, & -\lambda(|p^*| + |\nabla R|) S_m^* + \lambda m^* - \lambda_1^* = 0 \\
\min(\lambda_t^*, \sigma F(u^*)) = 0, & T(u^*) = K^*, \quad p^* = \nabla K^*, \\
m^* = p^* + \nabla R, & p^* = |p^*| n^*,
\end{cases}
$$

where $\beta^* = -\lambda_3^* \cdot n^* - \lambda(|m^*| - |\nabla R|)$. Consequently $u^* = (u^*_1, u^*_2)$ is a stationary point of model (2.1).
Proof. By (3.5), (3.6), (3.7) and (3.8), we have:

\[(3.23) \quad \lim \frac{1}{r_{1}}(\lambda_{h+1}^{k} - \lambda_{h}^{k}) = \lim (T(u_{h+1}^{k}) - K_{h+1}^{k}) = T(u^{*}) - K^{*} = 0,\]

\[(3.24) \quad \lim \frac{1}{r_{2}}(\lambda_{h+1}^{k} - \lambda_{h}^{k}) = \lim (p_{h+1}^{k} - \nabla K_{h+1}^{k}) = p^{*} - \nabla K^{*} = 0,\]

\[(3.25) \quad \lim \frac{1}{r_{3}}(\lambda_{h+1}^{k} - \lambda_{h}^{k}) = \lim (n_{h+1}^{k} - |p_{h+1}^{k}|n_{h+1}^{k}) = n^{*} - |p^{*}|n^{*} = 0.\]

From (3.8), we get:

\[(3.27) \quad 0 = \lim (\lambda_{h+1}^{k+1} - \lambda_{h}^{k}) = \lim \max \{-\lambda_{h}^{k}, -\sigma F_{e}(u_{h+1}^{k})\} = \min \{\lambda_{h}^{*},\sigma F_{e}(u^{*})\}.\]

Back to the optimality condition for the $u$-subproblem in (3.11), taking the limit in (3.11) over $k$ and considering equalities (3.24) and (3.26), we get:

\[
\begin{cases}
-\alpha \Delta u_{1}^{*} + \alpha_{1} \text{div}^{2} u_{1}^{*} + \lambda_{1} \partial_{x} T(u^{*}) + \partial_{u_{1}} C_{s}(u^{*}, \lambda_{h}^{*}) = 0, & \text{in } \Omega, \\
-\alpha \Delta u_{1}^{*} + \alpha_{1} \text{div}^{2} u_{1}^{*} + \lambda_{1} \partial_{x} T(u^{*}) + \partial_{u_{2}} C_{s}(u^{*}, \lambda_{h}^{*}) = 0, & \text{in } \Omega.
\end{cases}
\]

Now, we consider the optimality conditions for the $K$-subproblem and take the limit over $k$:

\[
\begin{align*}
- r_{2}(\Delta K_{h+1}^{k} - \text{div} p^{k}) + r_{1}(K_{h+1}^{k} - T(u_{h+1}^{k})) + \text{div} \lambda_{0}^{h} - \lambda_{1}^{h} &= 0, \text{ i.e.} \\
- r_{2}\text{div}(\nabla K^{*} - p^{*}) + r_{1}(K^{*} - T(u^{*})) + \text{div} \lambda^{*} - \lambda_{1}^{*} &= 0
\end{align*}
\]

where we used $\text{div} \nabla K^{*} = \Delta K^{*}$. Using the equalities (3.23) and (3.24), $\nabla K^{*} - p^{*} = 0$ and $K^{*} - T(u^{*}) = 0$. Then $\text{div} \lambda^{*} - \lambda_{1}^{*} = 0$. The optimality condition for the modified $p$-subproblem (3.17) leads to:

\[-\beta S_{p}^{k+1} + r_{2}(p_{h+1}^{k} - \nabla K_{h+1}^{k}) + r_{3}(p_{h+1}^{k} - |p_{h+1}^{k}|n_{h+1}^{k}) + r_{4}(p + \nabla R - m_{h+1}^{k}) + \sum_{i=2}^{4} \lambda_{i}^{k},\]

where $S_{p}^{k+1} \in \partial \|p_{h+1}^{k}\|$ and $\beta$ is given in (3.18). By Lemma 3.1, there exists a subsequence, still denoted by $S_{p}^{k} \in \partial \|p_{h}^{k}\|$, converging to $S_{p}^{*} \in \partial \|p^{*}\|$. Taking the limit over $k$ and taking into account equalities (3.24) and (3.25), we obtain:

\[-\beta^{*} S_{p}^{*} + \sum_{i=2}^{4} \lambda_{i}^{*} = 0.\]

For the $n$-subproblem, the optimality conditions give:

\[\lambda(n_{h+1}^{k} - \nabla n R) + r_{3}(p_{h+1}^{k} - |p_{h+1}^{k}|n_{h+1}^{k})^{2} + \lambda_{3}^{k} d \mathbf{x} = 0\]

Considering the limit over $k$ (3.25), we get:

\[\lambda n^{*} - \lambda \nabla n R - |p^{*}| \lambda_{3}^{*} = 0.\]

The same analysis applied to the optimality condition for the $m$-subproblem (3.21) leads to the equality:

\[-\lambda(|p^{*}| + |\nabla R|) S_{m}^{*} + \lambda m^{*} - \lambda_{4}^{*} = 0, \quad S_{m}^{*} \in \partial \|m^{*}\|\].

Finally we remark on getting the initializations by a multiresolution technique, also to
avoid local minima and to speed up registration. We use a scale space approach by resizing
the original images to a sequence of coarser ones where computations are cheap and register
these smaller images (see Fig. 3). Then starting from the coarsest level, we interpolate the
obtained transformation fields to get a starting guess on finer (next) levels until the original
resolution on the finest level is reached.

![Figure 3. Example of a multilevel representation of images.](image)

4. Numerical experiments. In this section, we assess the performance of the proposed
model (denoted by “New Model” below) and its algorithm. We compare the proposed model
with two other multimodality models:

- A MI model (denoted by MI below) that combines the regulariser (2.2) and the MI
  similarity measure (1.3);
- A NGF model (denoted by NGF below) that combines the regulariser (2.2) and the
  standard NGF similarity measure (1.4).

To measure the quality of the registered images, the following quantity

\[
GF_{cr} = \frac{F(\nabla T(u), \nabla R)}{F_0}
\]

is used as the relative reduction of the dissimilarity, where for two vectors \(x = (x_1, x_2)\) and
\(y = (y_1, y_2)\), we have

\[
F(x, y) = \frac{x_t}{\|x_t\|} - \frac{y_t}{\|y_t\|}_1, \quad x_t = (x_1, x_2), \quad y_t = (y_1, y_2).
\]
Here, $F_0 = F(\nabla T(u), \nabla R)$ if $u = 0$. For additional criteria to measure the goodness of registration, we also use the relative normalized gradient fields

$$
\text{NGF}_{er} = \frac{D^{NGF}(T(u), R)}{NGF_0}
$$

where $NGF_0 = D^{NGF}(T(u), R)$ if $u = 0$, and non-negative mutual information measure

$$
\text{MI}_{er} = -D^{MI}(T(u), R).
$$

For all the numerical experiments presented here, we summarise the comparative results in a table where we give the error computed using formulas (4.1)-(4.2). To measure mesh validity, we compute $C(u) = \det(I + \nabla u)$ from (2.1) and monitor if it is positive.

In order to reduce the number of parameters to tune, we set $r_1 = 5$, $r_2 = 10$ and $r_3 = r_4 = 100$ in all numerical experiments unless stated otherwise. We consider $N_{max} = 70$ the maximum number of iterations for New Model from Algorithm 3.2 and we stop the iterations before reaching $N_{max} = 70$ if the following stopping criterion

$$
\|p^k + \nabla R - m^k\|_L^1 \leq \tau
$$

is satisfied for a given tolerance $\tau = 10^{-3}$, where $l$ and $c$ are the numbers of rows and columns in the image. Though we can use all equations from Algorithm 3.1 to stop iterations, we find that the above stopping criterion based on its 4th equation is sufficient as it includes information about the gradients of both images. Thus, it can control the ALM iterations and the quality of registration at the same time. For each variable $u_1$ and $u_2$, we computed the residual via finite differences approximation and the global residual is taken as the sum. The residual is given by the quantity

$$
S_{er} = \frac{1}{l \times c} \sum_{i=1}^{l} \sum_{j=1}^{c} \left| \frac{\partial J_1(u^k)}{\partial u_1^k} \right|_{i,j} + \frac{1}{l \times c} \sum_{i=1}^{l} \sum_{j=1}^{c} \left| \frac{\partial J_1(u^k)}{\partial u_2^k} \right|_{i,j}
$$

where $(\frac{\partial J_1(u^k)}{\partial u_1^k})_{i,j} = J(u^k) - J(u^{kij})$, $u^{kij} = (u_1^{kij}, u_2^{kij})$ and $u_1^{kij}$ takes the same values of $u_1^k$ on each point of the discrete domain, except on the position $(i, j)$ where it takes the values of the old ALM solution $u_1^{k-1}(i, j)$. The term $(\frac{\partial J_1(u^k)}{\partial u_2^k})_{i,j}$ is defined in a similar way. We also plot the curve of the quantity

$$
D_m = \frac{DG^F(T(u), R)}{DG^F(T, R)} + \frac{DT^M(T(u), R)}{DT^M(T, R)}
$$

which represents the relative errors for the new similarity measures as function of the ALM iterations.

For the NGF and MI similarity measures, the numerical experiments are performed using the publicly available image registration toolbox flexible algorithms for image registration (FAIR)\(^1\), where the implementation is based on the Gauss-Newton method. The constraint

\(^1\)http://www.siam.org/books/fa06/
on the determinant $\det(I + \nabla u) > 0$ is explicitly included in FAIR’s models; in fact, a line
search method is used in FAIR and the new descent direction is chosen such that the constraint
$\det(I + \nabla u) > 0$ is verified.

As we shall see, in almost all experiments, the New Model outperforms the standard
NGF and the New Model also outperforms MI in examples where dominating gradients
represent main image features or they correspond to each other, while the New Model
performs similarly to MI for other examples (e.g. Example 6).

Example 1. In the first example, we consider a synthetic image to illustrate the type of
images where mutual information (MI) and the normalized gradient field (NGF) models are
at disadvantages. We obtain a good result using New Model as seen in Fig. 4. Here, the
NGF and MI models were tested for different regularization parameters. The optimal choices
are considered by making different tests where we set $\alpha_1 = 1$, $\alpha = 0.01\alpha_1$ and we vary $\lambda$ such
that $\frac{\alpha}{\lambda} \in \{10^{-5}, 5 \times 10^{-5}, 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 10^{-2}\}$ for MI, and $\frac{\alpha}{\lambda} \in \{2.5, 2, 1.5, 1, 0.5, 0.1\}$
for NGF. The optimal parameters were $\frac{\alpha}{\lambda} = 10^{-4}$ and $\frac{\alpha}{\lambda} = 10^{-4} = 0.5$ for MI and NGF,
respectively. They were chosen such that the registered image is very close to the reference
and the transformations does not suffer from mesh folding. For comparison, we used the
Jaccard similarity coefficient (JSC) which is defined as follows:

$$JSC = \frac{|S_T \cap S_R|}{|S_T \cup S_R|},$$

where $S_T$ and $S_R$ represent, respectively, the segmented regions of interest (with red contour)
in the deformed template (after registration) and the reference.

Examples 2 and 3. In Fig 5, we consider a reference image from photon density weighted
MRI and a template image which represent MRI-T2, both of size $256 \times 256$. A second set of
examples is shown in Fig 6. We compare with the different multi-modal registration models.
For each model, we display registered templates. We can see that all models perform well for
both examples and give satisfactory results. The results of the NGF and MI are broadly
comparable. In both examples with all models, the results for the registration look visually
identical. We display an overlay in alternating squared patches of the registered and the
reference image (to possibly see major discontinuities of features). We quantify the quality of
registration using the GF$_{er}$, MI$_{er}$ and NGF$_{er}$ errors which confirm that New Model; e.g.
at the top left (second box down) of Fig 5, gives better alignments than compared models.
For the run runtime comparison with the MI and NGF models, we tested all models for
the pair images in 5 for different resolutions. The FAIR’s models are always slightly faster
because they are optimized (based on Gauss-Newton method)

Example 4. In Fig. 8, we present the result of registering two diffusion-MRI images of
size $256 \times 256$ with respectively high and low b-value diffusion. Since the intensity values for
different b-values are not comparable, conventional non-modality registration models (that
rely on matching the images based on the intensity values) will fail. We show the registration
results by our compared 3 models in Fig. 8. We notice that NGF and MI models give
comparable results. However, our New Model gives the best result comparing to the other
two and visually, the reference and the transformed template are well aligned in all regions.
Since $C(u) > 0$, all transformed grids have no mesh folding.
Example 5. In the next experiment in Fig. 9, our aim is to investigate capabilities of the proposed models for registration of MRI-T1 and MRI-T2 images in higher resolution $512 \times 512$. We can observe from overlaying of the registered and the reference images that all models work fine in producing acceptable registration results, however the registered result by New Model produces the best alignment in all parts and gives the better similarity value than NGF (here identical to MI). We also show the resulting transformed grids for all models where there is no mesh folding due to $C(u) > 0$. For the above 4 examples (Ex.2–Ex.5), in Fig 10, we display the evolution of the error versus the ALM iteration to the final solution. We also plot the evolution of the residual for the energy (2.1) as a function of ALM iterations. Here we see that our ALM algorithm converges though the convergence is not monotone.
Example 6. Example 6 tests the registration of a MRI image to a PET with much noise. In Fig. 12, we present the results obtained using the **New Model**, **NGF** and **MI**. Clearly, **New Model** and **MI** perform better than **NGF** in this case and in particular the **New Model** performs the best (even though it is slightly better than **MI Model**). We display an overlaying of the registered and the reference images which shows that the registered result by **New Model** produces the best alignment.

Regularisation parameters dependence test. In Table 3, we compare the sensitivity of the proposed model with respect to varying the ratio $\frac{\alpha}{\lambda}$. The model was tested on Example 2 where we set $\alpha_1 = 1$, $\alpha = 0.01\alpha_1$ and we vary $\lambda$ for all experiments. We can see a clear process of the changes of the relative error where the best error is obtained for $\frac{\alpha}{\lambda} = 0.017$ and the error increases as the ratio decreases more than 0.017.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>64 x 64</th>
<th>128 x 128</th>
<th>256 x 256</th>
<th>512 x 512</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s) for New Model</td>
<td>29.836</td>
<td>49.931</td>
<td>117.342</td>
<td>272.578</td>
</tr>
<tr>
<td>Time (s) for MI Model</td>
<td>14.794</td>
<td>21.437</td>
<td>48.881</td>
<td>76.398</td>
</tr>
<tr>
<td>Time (s) for NGF Model</td>
<td>22.003</td>
<td>42.845</td>
<td>100.961</td>
<td>264.388</td>
</tr>
</tbody>
</table>

Table 1

Run time comparison for all models for the pair of MRI images in Fig. 6

5. Conclusions. Image registration is an increasingly important and often challenging image processing task 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 regularisation term which combines first- and second-order derivatives of the displacement. After showing the solution existence, we present a fast ALM for its numerical implementation. Experimental tests confirm that our proposed model performs better in multi-modality images registration than compared models. It is pleasing to see much improved results over established models within the same modelling framework. Future work will consider generalizations to 3 dimensions and registration of images that do not have dominant gradients.

REFERENCES


Compared Models | NGF | New Model | MI
---|---|---|---
| \#G | \#N | GF_{er} | NGF_{er} | MI_{er} | GF_{er} | NGF_{er} | MI_{er} | GF_{er} | NGF_{er} | MI_{er} |
Ex 1 | 0.2% | 0.02% | 0.540 | 0.964 | 0.446 | 0.032 | 0.932 | 0.993 | 0.370 | 0.970 | 0.381 |
Ex 2 | 49% | 24% | 0.636 | 0.640 | 1.170 | 0.247 | 0.756 | 1.206 | 0.490 | 0.879 | 1.193 |
Ex 3 | 43% | 23% | 0.336 | 0.491 | 1.265 | 0.674 | 0.800 | 1.184 | 0.765 | 0.849 | 1.154 |
Ex 4 | 49% | 20% | 0.901 | 0.856 | 1.150 | 0.674 | 0.800 | 1.184 | 0.765 | 0.849 | 1.154 |
Ex 5 | 43% | 37% | 0.741 | 0.656 | 1.265 | 0.454 | 0.623 | 1.178 | 0.454 | 0.631 | 1.163 |
Ex 6 | 48% | 23% | 0.952 | 0.957 | 1.187 | 0.801 | 0.920 | 1.341 | 0.836 | 0.970 | 1.254 |

Table 2
Registration results of the different models for processing Examples 1-5 shown respectively in Fig. 5, 6, 7 and 8. The errors are computed using formula (4.1), (4.3) and (4.2). Here, \#N is the ratio of the number of pixels where \(\nabla_n T \cdot \nabla_n R \neq 0\) over the total number of pixels, whereas \#G is the ratio of number pixels where \(GF(T,R) + TM(T,R) \neq 0\) over the total number of pixels.

<table>
<thead>
<tr>
<th>(\alpha_1/\lambda)</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>0.017</th>
<th>0.0125</th>
<th>0.01</th>
<th>0.0075</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>0.238</td>
<td>0.237</td>
<td>0.237</td>
<td>0.236</td>
<td>0.237</td>
<td>0.237</td>
<td>0.238</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 3
Registration results for \(\alpha_1/\lambda\)-dependence tests of New Model for processing Example 3. The relative errors are computed using the normalized gradient fitting formula (4.1). In all cases, we set \(\alpha = 0.01\alpha_1\).


Figure 5. Example 2: Comparison of different models to register T-1 and T2-MRI images. **New Model** performs the best.

(a) The reference $R$

(b) The template $T$

(c) Overlay of $R$ and $T$

(d) $T(u)$ by **New Model**, $GF_{er}=0.247, MI_{er}=1.206, NGF_{er}=0.756$

(e) $T(u)$ using **NGF**, $GF_{er}=0.636, MI_{er}=1.17, NGF_{er}=0.640$

(f) $T(u)$ using **MI**, $GF_{er}=0.490, MI_{er}=1.193, NGF_{er}=0.879$

(g) Overlay of $R,T(u)$ for **New Model**

(h) Overlay of $R$ and $T(u)$ for **NGF**

(i) Overlay of $R$ and $T(u)$ for **MI**

(j) $x + u(x)$ by **New Model**, $\min \mathcal{C} = 0.22$

(k) $x + u(x)$ using **NGF**, $\min \mathcal{C} = 0.19$

(l) $x + u(x)$ using **MI**, $\min \mathcal{C} = 0.45$
Figure 6. Example 3: Registration of a second pair of MRI images (T1 and T2). New Model performs the best.
Figure 7. Comparison of 3 different models to register the MRI images from Fig. 6. Example 3 zoomed in the red squares (see Fig. 6): From left to right; Zooms in the reference $R$ and the registered $T(u)$ using New model, NGF and MI, respectively.
Figure 8. Example 4: High-b- and Low-b-value Diffusion-weighted MRIs (of 256 × 256) using different models. New Model performs the best.
Figure 9. Example 5: a pair of MRI images of higher resolution $512 \times 512$ by 3 different models. New Model and MI perform identically, both better than NGF.
Figure 10. Left: Log scale plot of the residual errors for $u$ versus ALM iteration numbers for examples 2-5. Right: Plot of the error $S_{	ext{er}}$ values versus ALM iteration numbers for examples 2-5.

Figure 11. Left: Log scale plot of the distance $D_m$ versus ALM iteration numbers for examples 2-5.
Figure 12. Example 6: Registering a PET image to an MRI image. New model performs better than others in this example.