Registration of Edge Sets for Mapping a Purkinje Fiber Network onto an Endocardium

Stefan Fürtinger*, Stephen Keeling†, Gernot Plank‡ and Anton J. Prassl§

December 28, 2010

Abstract. In this work, edge sets are mapped one to the other by representing these zero measure sets as *diffuse images* which have positive measure supports that can be registered elastically. The driving application for this work is to map a Purkinje fiber network in the epicardium of one heart to the epicardium of another heart. The approach is to register sufficiently accurate diffuse surface representations of two epicardia and then to apply the resulting transformation to the points of the Purkinje fiber network. To create a diffuse image from a given edge set, a region growing method is used to approximate diffusion of brightness from an edge set to a given point. To be minimized is the sum of squared differences of the registered diffuse images along with an elastic penalty for the registration. A Newton iteration is employed to solve the optimality system, and the degree of diffusion is larger in initial iterations while smaller in later iterations so that a desired local minimum is selected by means of vanishing diffusion. Global existence of minimizers is proved for each fixed degree of diffusion. Furthermore, convergence of global minimizers as diffusion decreases is proved. Nevertheless, it is demonstrated that the local minimizer introduced by vanishing diffusion is preferable to the trivial global minimizer corresponding to the absence of diffusion. Favorable results are shown for registering highly detailed rabbit heart models.

Keywords: elastic image registration, edge sets, diffuse surfaces, Purkinje Fibers, Endocardium.

*Institut für Mathematik, Karl–Franzens-Universität Graz, Heinrichstraße 36, 8010 Graz, Austria; email: stefan.fuertinger@uni-graz.at. Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 03, "Optimierung und Kontrolle".

†Institut für Mathematik, Karl–Franzens-Universität Graz, Heinrichstraße 36, 8010 Graz, Austria; email: stephen.keeling@uni-graz.at. Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 03, "Optimierung und Kontrolle".

‡Institut für Biophysik, Medizinische Universität Graz, Harrachgasse 21/IV, 8010 Graz, Austria; email: gernot.plank@medunigraz.at. Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 03, "Optimierung und Kontrolle".

§Institut für Biophysik, Medizinische Universität Graz, Harrachgasse 21/IV, 8010 Graz, Austria; email: anton.prassl@medunigraz.at. Supported by the Medical University of Graz under grant "Startfunding 2009".

1. Introduction

The heart is an electrically controlled mechanical pump. It's main function is to drive blood through the circulatory system, thus providing oxygen and metabolites to the organs. A well coordinated electrical activation sequence is of vital importance for allowing an energy-efficient mechanical contraction. In the ventricles, the main pumping chambers of the heart, the electrical impulse is conducted first via the specialized conduction system, referred to as the Purkinje system (PS). The PS is a highly ramified network of thin cable-like 1D structures which synchronizes ventricular activation by quickly distributing the electrical impulse to the endocardium, i.e., the surfaces of the ventricular cavities. The PS consists of two components, the endocardial PS which lines the ventricular endocardium, and the free running PS which traverses the cavities. The PS is electrically isolated from the ventricles except at discrete endpoints (Purkinje-myocardial junctions) [33]. Transmission of the electrical signals at these discrete junctional sites, referred to as Purkinje ventricular junctions (PVJ) is essential to excite the ventricular mass [20].

Under pathological conditions cardiac arrhythmias, i.e., disturbances of the healthy activation sequence, may arise which may even degenerate into highly irregular activation patterns, referred to as fibrillation. This loss of electrical synchronicity entails an impairment of the heart's ability to pump blood, which, ultimately, may even lead to sudden cardiac death. Despite the vital role of the PS in shaping the activation sequence under healthy conditions, it is assumed that the PS may be implicated in the formation and maintenance of such arrhythmias, under various conditions including shock-induced arrhythmogenesis, failure of defibrillation shocks [6, 12] and arrhythmias induced by focal activation [8].

Surprisingly, in most studies, both experimental as well as computational, PS effects are quite often neglected. Despite recent advances, PS activity at the organ level cannot be observed directly with currently available experimental modalities. Indirect observations are possible [15], but major advanced are hampered by the inability of current experimental techniques to resolve, with sufficient accuracy, electrical behavior confined to the depth of the ventricles or in the PS. Computer models quite naturally suggest themselves as a surrogate technique to bridge the gap between experimental observations, typically recorded at the epicardial surface of the heart, and electrical events occurring within the PS, at the ventricular epicardium or within the depth of ventricular walls. Despite major recent advancements in modeling technology [29, 28], integrating topologically realistic models of the PS with anatomically and functionally realistic models of the ventricles remains challenging.

Owing to the physiological importance of the PS it is highly desirable to include the PS in computational models. Hence, fast techniques to incorporate PS topologies with 3D finite element grids of the ventricles are sought after. The purpose of the present work was to develop a mathematical framework which enables the mapping of the endocardial PS between different ventricular surface geometries. In the absence of experimental data, a literature-based PS [35], constructed for the San Diego rabbit heart model [34], served as reference topology. A recent anatomically highly realistic model of rabbit ventricles [5] served as a target for developing and testing the mapping technique. Both models are shown in Figure 6.1.

The exact location of the PS (and especially the PVJ) with respect to the ventricular surface geometry is crucial in modeling the role of PS effects in the excitation of the cardiac muscle.

Hence it is not sufficient to just project a given reference PS into a heart model. The different surface geometries of the modeled ventricles have to be taken into account as well. Thus we first seek a geometric transformation to match a template heart model with a given reference model. In the context of mathematical image processing this can be seen as a 3D registration problem. Once the transformation is found it can be applied to map structures (like the PS) within the template model onto the reference model. This approach guarantees that not only topological features of the PS but also its relative position to the ventricle are preserved and projected onto the reference heart.

The heart models considered here are 3D finite element grids of very high complexity. Given the sheer size of the models it would require a massive computational effort to calculate just simple transformations. Hence we developed a method that is capable of computing even highly non-linear transformations in a reasonable time while requiring only moderate computational resources. We perform a dimensional reduction of the problem by treating the 3D models as 2D image sequences. This strategy reduces memory consumption considerably while simultaneously allowing us to use very efficient techniques to solve the resulting 2D registration problems.

2. Edges as Binary Images

We consider slices arising from cuts through (a) a realistic model of rabbit ventricles [5] and (b) the San Diego rabbit heart [34]. Thus we have two-dimensional edge-sets Γ_0 and Γ_1 and we assume that both edge-sets are closed and have finite Hausdorff-measure $\mathcal{H}^1(\Gamma_i) < \infty$ for $i = 0, 1$. Let $\Omega := [1, N]^2 \subset \mathbb{R}^2$ with $N \in \mathbb{N}$ and let I_0 and I_1 be the characteristic functions of Γ_0 and Γ_1 respectively. Then I_0 and I_1 can be interpreted as binary images on Ω. The goal now is to find a displacement $\boldsymbol{w} : \mathbb{R}^2 \to \mathbb{R}^2$ such that $I_0(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})) \approx I_1(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$. For the sake of brevity the argument of $\boldsymbol{w}$ is omitted in the following.

One approach to the computation of the desired displacement is to treat points on Γ_0 as if connected to one another by elastic springs which are perturbed minimally to meet the target set Γ_1. However, because of the potentially very high computational complexity of such a formulation, the approach used here to match the edge sets is to embed them into images which are then registered elastically. Elastic potential energy has been used by many authors to regularize image registration; see, e.g., [27], [21] and particularly the review in [23]. Such regularization is particularly natural when used to register images of tissues having undergone relatively small displacements. However, in the present context, the required displacement field is highly nonlinear, owing partly to the complex geometry of the heart and partly to the great difference in regularity of the two given edge sets. Nevertheless elastic registration is employed here, but with considerable precautions.

In addition to regularization, a notion of image similarity must be selected for the image registration problem at hand. Assuming that $\{I_i\}_{i=0,1} \subset L^2(\Omega)$ one of the simplest distance measures (see for instance [18]) is the sum of squared intensity differences (SSID) which in this case is given by

$$\frac{1}{2} \int_\Omega |I_0(\boldsymbol{x} + \boldsymbol{w}) - I_1(\boldsymbol{x})|^2 \, d\boldsymbol{x}. \tag{2.1}$$

However, in this form the approach is not feasible for the problem: both Γ_0 and Γ_1 are sets of measure zero in Ω and $\text{supp}(I_i) = \Gamma_i$ for $i = 0, 1$. Hence the trivial deformation $\boldsymbol{w} \equiv 0$ minimizes the SSID measure

$$\frac{1}{2}\int_{\Omega} |I_0(\boldsymbol{x}) - I_1(\boldsymbol{x})|^2 \, d\boldsymbol{x} = \frac{1}{2}\int_{\Gamma_0 \cup \Gamma_1} |I_0(\boldsymbol{x}) - I_1(\boldsymbol{x})|^2 \, d\boldsymbol{x} = 0. \tag{2.2}$$

Since we are dealing with edge-sets another perhaps more natural approach to measure difference is employing the Hausdorff-distance

$$d_H(\Gamma_0, \Gamma_1) := \max\left(\sup_{\boldsymbol{x} \in \Gamma_0} d_{\Gamma_1}(\boldsymbol{x}), \sup_{\boldsymbol{x} \in \Gamma_1} d_{\Gamma_0}(\boldsymbol{x}) \right), \tag{2.3}$$

where

$$d_{\Gamma_i}(\boldsymbol{x}) := \inf_{\boldsymbol{y} \in \Gamma_i} |\boldsymbol{x} - \boldsymbol{y}|_2 \,, \; i = 0, 1,$$

and $|\cdot|_2$ denotes the standard Euclidean norm in $\mathbb{R}^2$. The Hausdorff-distance is a popular tool in computer graphics and image processing it is and mainly used for shape recognition problems. For instance, Knauer et al. [22] developed a method for minimizing the Hausdorff-distance under translations and rigid motions to determine a registration in the context of neurosurgical operations. Though their proposed algorithm is efficient, it is limited to rigid transformations. Fuchs et al. [19] introduced an elastic deformation distance in a shape space; however, calculating the shortest path between shapes proved to be computationally expensive. Droske and Ring [16] developed a regularized shape gradient descent algorithm within a level-set framework for simultaneous registration and segmentation. However, due to (2.2) for instance, the present problem still lacks sufficient structure to be treated directly by such approaches.

We present here a technique that combines the simplicity of the SSID-measure (2.1) with the accuracy of the Hausdorff-distance (2.3). Instead of looking at I_0 and I_1 directly, we consider approximations of edge-sets by diffuse regions in images. Note that this strategy is also employed by Ambrosio and Tortorelli [2] in their approximation of the Mumford-Shah functional [24]. Here, for $\varepsilon > 0$ define $\mathcal{I}_i^\varepsilon$ for $i = 0, 1$ by

$$\mathcal{I}_i^\varepsilon(\boldsymbol{x}) := \begin{cases} 1 - d_{\Gamma_i}(\boldsymbol{x})/\varepsilon, & \text{if } d_{\Gamma_i}(\boldsymbol{x}) \le \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

The nonzero regions of $\mathcal{I}_i^\varepsilon$ are diffuse extensions of the edges Γ_i. Since the distance function $d_{\Gamma_i}(\boldsymbol{x})$ can be expensive to compute, $\mathcal{I}_i^\varepsilon$ is calculated in practice by a marching procedure in which the distance $d_{\Gamma_i}(\boldsymbol{x})$ is approximated in terms of the number of marching steps from the edge set Γ_i to $\boldsymbol{x}$. The extent of this "region-growing" depends on the magnitude of ε: the effect looks like a silk painting of Γ_i where ε affects the duration of the brush touching the fabric. From an image processing point of view this technique can be seen as a distance transform. (A comprehensive review on the use of distance transforms in digital image processing is given in [32].) Now we may use (2.1) on the region-grown versions $\mathcal{I}_i^\varepsilon$ of I_i to obtain an adapted distance

measure

$$S^\varepsilon(\boldsymbol{w}) := \frac{1}{2}\int_\Omega |\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 \, d\boldsymbol{x}. \tag{2.5}$$

As indicated above, we assume that $\boldsymbol{w}$ is an elastic deformation; thus, we define the following linear elastic potential (see for instance [23, Chap. 9])

$$E(\boldsymbol{w}) := \frac{\lambda}{2}\int_\Omega (\nabla\cdot\boldsymbol{w})^2 \, d\boldsymbol{x} + \frac{\mu}{4}\int_\Omega \left|\nabla\boldsymbol{w}^\top + \nabla\boldsymbol{w}\right|_F^2 \, d\boldsymbol{x}, \tag{2.6}$$

where $|\cdot|_F$ denotes the Frobenius-norm and μ and λ are positive constants describing the elastic properties of the body, the so-called Navier–Lamé constants. Additively extending S^ε by soft constraints, namely the linear elastic potential E, ensures that among all solutions elastic deformations are favored. Hence we end up with the following cost functional

$$\begin{aligned} J^\varepsilon(\boldsymbol{w}) :=& S^\varepsilon(\boldsymbol{w}) + E(\boldsymbol{w}) \\ =& \frac{1}{2}\left\|\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})\right\|_{L^2(\Omega)}^2 + \frac{\lambda}{2}\left\|\nabla\cdot\boldsymbol{w}\right\|_{L^2(\Omega)}^2 + \frac{\mu}{4}\left\|\nabla\boldsymbol{w}^\top + \nabla\boldsymbol{w}\right\|_{L^2(\Omega)}^2, \end{aligned}$$

for a fixed $\varepsilon > 0$. Assuming that $\boldsymbol{w} \in H^1(\Omega)$ the cost J^ε is well defined. Summarized we compute a registration by solving the following minimization problem

$$\min_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}). \tag{2.7}$$

Note that the images $\mathcal{I}_i^\varepsilon$ vary with the value of ε. Hence for each ε, (2.7) forms a stand-alone minimization problem. A rigorous proof of existence of a solution to (2.7) for ε fixed is given in the appendix. For a better understanding of the behavior of the minimization problem (2.7) as $\varepsilon \to 0$, consider the following remark. (A further investigation is presented in the appendix.)

Remark 1. Note that there is a close relation between the Navier–Lamé constants λ, μ and ε. As indicated above for $\varepsilon = 0$, i.e., $\mathcal{I}_i^\varepsilon = I_i$, a solution to the registration problem (2.7) is $\boldsymbol{w} = 0$. However, it should be mentioned that for small values of ε ("small" compared to the Navier–Lamé constants) J^ε is minimized by $\boldsymbol{w} = 0$ as well. To clarify this, consider a simplified 1D-example.

Let $\Omega := (0,1) \subset \mathbb{R}$ and consider the "binary images" I_0 and I_1 defined by

$$I_0(x) := \begin{cases} 1, & x = \frac{1}{4}, \\ 0, & \text{otherwise}, \end{cases} \qquad I_1(x) := \begin{cases} 1, & x = \frac{1}{2}, \\ 0, & \text{otherwise}. \end{cases}$$

We want to register I_0 and I_1, i.e., we want to find a deformation w such that $I_0(x + w(x)) \approx I_1(x)$. Hence we introduce the following cost functional

$$\begin{aligned} J(w) :=& \int_0^1 |I_0(x+w(x)) - I_1(x)|^2 \, dx + \mu\int_0^1 |w'(x)|^2 \, dx \\ =& \mathcal{S}(I_0, I_1; w) + \mathcal{P}(w), \end{aligned}$$

with a regularization parameter $\mu > 0$. Since $\operatorname{supp}(I_i)$ consists only of discrete points we

obviously have (similar to the 2D case) $\mathcal{S}(I_0, I_1; 0) = 0$. Hence we introduce the "blurred" approximations

$$\mathcal{I}_0^\varepsilon(x) := \begin{cases} 1, & \frac{1}{4} \le x \le \frac{1}{4} + \varepsilon, \\ 0, & \text{otherwise}, \end{cases} \qquad \mathcal{I}_1^\varepsilon(x) := \begin{cases} 1, & \frac{1}{2} \le x \le \frac{1}{2} + 2\varepsilon, \\ 0, & \text{otherwise}, \end{cases} \qquad 0 < \varepsilon \le \frac{1}{4},$$

i.e., the "images" $\mathcal{I}_i^\varepsilon$ are step functions, see Fig. 2.1 (a) for a sketch. Thus we modify the cost J by defining

$$J^\varepsilon(w) := \mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w) + \mathcal{P}(w).$$

With a view to Fig. 2.1 (a) we look for deformations of the form

$$w_\varepsilon(x) := bx,$$

with some scalar $b \in \mathbb{R}^+$ and $0 < \varepsilon \le \frac{1}{4}$ fixed. The "optimal" deformation (from an image processing point of view) is obviously given by

$$w_\varepsilon^\star(x) := x,$$

i.e., $b^\star = 1$. Due to the simple structure of the problem we are able to compute $\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon)$ explicitly. We have to consider four cases depending on the magnitude of b (compare Fig. 2.1 (a)):

$(1+b)(\frac{1}{4}+\varepsilon) \le \frac{1}{2}$, i.e., the right "edge" of $\mathcal{I}_0^\varepsilon$ is not shifted further than to the left "edge" of $\mathcal{I}_1^\varepsilon$. Then we obtain

$$\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon) = (1+b)\left(\left(\frac{1}{4}+\varepsilon\right) - \frac{1}{4}\right) + \left(\frac{1}{2} + 2\varepsilon - \frac{1}{2}\right) = (b+3)\varepsilon.$$

$\frac{1}{2} \le (1+b)(\frac{1}{4}+\varepsilon) \le \frac{1}{2} + 2\varepsilon$, i.e., the right edge of $\mathcal{I}_0^\varepsilon$ is shifted into the interior of $\mathcal{I}_1^\varepsilon$. We compute

$$\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon) = \left(\frac{1}{2} - (1+b)\frac{1}{4}\right) + \left(\frac{1}{2} + 2\varepsilon - (1+b)\left(\frac{1}{4}+\varepsilon\right)\right) = \frac{1}{2}(1-b)(1+2\varepsilon).$$

$\frac{1}{4}(1+b) \le \frac{1}{2} + 2\varepsilon \le (1+b)(\frac{1}{4}+\varepsilon)$, i.e., the shifted edges of $\mathcal{I}_0^\varepsilon$ are left (left edge) and right (right edge) of the right edge of $\mathcal{I}_1^\varepsilon$. Hence we get

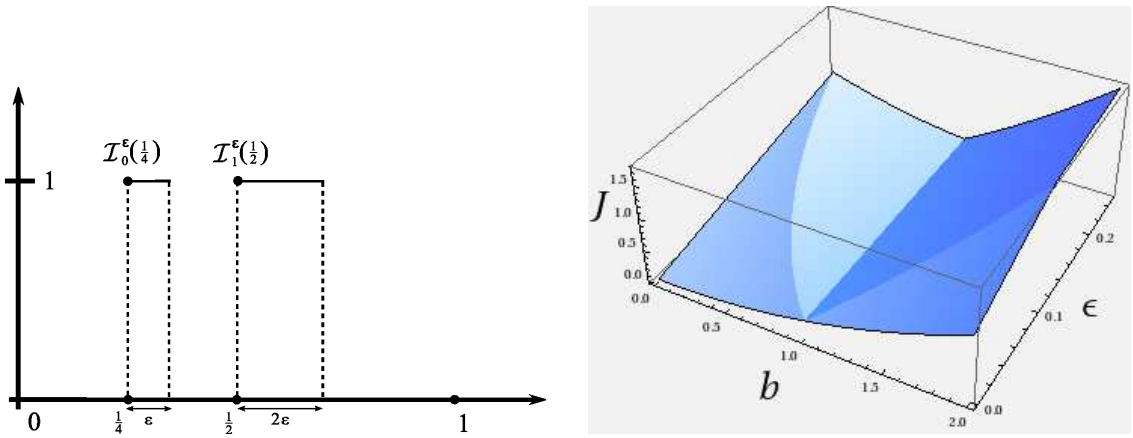
$$\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon) = \left(\frac{1}{4}(1+b) - \frac{1}{2}\right) + \left((1+b)\left(\frac{1}{4}+\varepsilon\right) - \frac{1}{2} - 2\varepsilon\right) = \frac{1}{2}(b-1)(1+2\varepsilon).$$

$\frac{1}{2} + 2\varepsilon \le \frac{1}{4}(1+b) \le 1$, i.e., the left edge of $\mathcal{I}_0^\varepsilon$ is shifted to the right of the right edge of $\mathcal{I}_1^\varepsilon$. This yields

$$\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon) = \left(\frac{1}{2} + 2\varepsilon - \frac{1}{2}\right) + (1+b)\varepsilon = (b+3)\varepsilon.$$

The penalizer $\mathcal{P}$ for $w_\varepsilon = bx$ is given by

$$\mathcal{P}(w_\varepsilon) = \mu \int_0^1 |b|^2 \, dx = \mu b^2.$$



(a) (b)

Figure 2.1: (a) Sketch of the simplified 1D-example discussed in Remark 1. (b) Graph of $G(\varepsilon, b)$ on $[0, \frac{1}{4}] \times [0, 2]$ for $\mu = \frac{1}{4}$.

Now consider the case $b = 0$, i.e., $w_\varepsilon = 0$, then we have $\mathcal{P}(0) = 0$ and the similarity measure is given by (the first case applies)

$$\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; 0) = 3\varepsilon.$$

On the other hand for the heuristic "optimal" deformation $w_\varepsilon^\star(x) := x$ with $b^\star = 1$ we have $\mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon^\star) = 0$ and $\mathcal{P}(w_\varepsilon^\star) = \mu$. With $0 < \varepsilon < \min(\frac{1}{4}, \frac{\mu}{3})$ the previous considerations imply

$$J^\varepsilon(0) = \mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; 0) = 3\varepsilon < 3 \cdot \frac{\mu}{3} = \mathcal{P}(w_\varepsilon^\star) + \mathcal{S}(\mathcal{I}_0^\varepsilon, \mathcal{I}_1^\varepsilon; w_\varepsilon^\star) = J^\varepsilon(w_\varepsilon^\star).$$

We see that if ε becomes "too small" compared to the regularization parameter μ, the trivial deformation $w \equiv 0$ minimizes the cost J^ε. To illustrate this phenomenon Fig. 1 (b) shows the graph of $G(\varepsilon, b) := J^\varepsilon(bx)$ for $\varepsilon \in [0, \frac{1}{4}]$, $b \in [0, 2]$ and $\mu = \frac{1}{4}$. Obviously the global minimizer of J^ε is $b = 0$ (and hence $w = 0$) for small values of ε but $b = 1$ for sufficiently large ε. Note that this behavior is not a specialty of the one dimensional case as it can be observed in 2D as well. Therefore we developed an iterative solution strategy. We compute an initial deformation for a "large" ε. This rough solution is used as initial guess to solve the registration problem for a smaller value of ε. We repeat this process to iteratively refine the transformation by simultaneously avoiding that the computed transformations get close to zero as $\varepsilon \to 0$ (details are given in Sec. 5.2). Although this means that we only compute local minima of the cost (as ε gets smaller) the example given above shows that global minima are not necessarily better suited to obtain a good registration.

3. Optimality Conditions

To obtain the optimality conditions for the optimization problem (2.7) we compute the variational derivative of J^ε in an arbitrary direction $\boldsymbol{v} \in C^\infty(\bar{\Omega})$

$$\frac{\delta J^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w}; \boldsymbol{v}) := \frac{d}{ds} J^\varepsilon(\boldsymbol{w} + s\boldsymbol{v})\Big|_{s=0}.$$

To keep things clear we employ the linearity of the Gâteaux derivative and split the computation into two parts

$$\frac{\delta J^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) = \frac{\delta S^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) + \frac{\delta E}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}).$$

Starting with S^ε we get

$$\frac{\delta S^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) = \int_\Omega (\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))\nabla \mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\cdot \boldsymbol{v}\, d\boldsymbol{x},$$

where here ∇ denotes the distributional derivative of $\mathcal{I}_0^\varepsilon$. Next we write $E(\boldsymbol{w})$ in a more explicit form by using

$$\left|\nabla \boldsymbol{w}^\top + \nabla \boldsymbol{w}\right|_F^2 = \sum_{j,\ell=1}^{2} \left(\frac{\partial}{\partial x_j} w_\ell + \frac{\partial}{\partial x_\ell} w_j\right)^2,$$

where $\boldsymbol{w}(\boldsymbol{x}) := (w_1(\boldsymbol{x}), w_2(\boldsymbol{x}))^\top$and compute with $\boldsymbol{v}(\boldsymbol{x}) := (v_1(\boldsymbol{x}), v_2(\boldsymbol{x}))^\top$ and $\boldsymbol{x} := (x_1, x_2)^\top$

$$\begin{aligned}
\frac{\delta E}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) &= \frac{d}{ds}\int_\Omega \frac{\mu}{4}\sum_{j,\ell=1}^{2} \left(\partial_{x_j}(w_\ell + s v_\ell) + \partial_{x_\ell}(w_j + s v_j)\right)^2 d\boldsymbol{x}\Big|_{s=0} \\
&\quad + \frac{d}{ds}\int_\Omega \frac{\lambda}{2}\left(\nabla\cdot(\boldsymbol{w}+s\boldsymbol{v})\right)^2 d\boldsymbol{x}\Big|_{s=0} \\
&= \int_\Omega \frac{\mu}{2}\sum_{j,\ell=1}^{2} (\partial_{x_j} w_\ell + \partial_{x_\ell} w_j)(\partial_{x_j} v_\ell + \partial_{x_\ell} v_j)\, d\boldsymbol{x}\Big|_{s=0} \\
&\quad + \frac{\lambda}{2}\frac{d}{ds}\int_\Omega (\nabla\cdot\boldsymbol{w})^2 + 2s\,(\nabla\cdot\boldsymbol{w})\,(\nabla\cdot\boldsymbol{v}) + s^2\,(\nabla\cdot\boldsymbol{v})^2\, d\boldsymbol{x}\Big|_{s=0} \\
&= \int_\Omega \mu \sum_{\ell=1}^{2} (\nabla w_\ell + \partial_{x_\ell}\boldsymbol{w})\cdot\nabla v_\ell + \lambda\,(\nabla\cdot\boldsymbol{w})\,(\nabla\cdot\boldsymbol{v})\, d\boldsymbol{x}. \qquad (3.1)
\end{aligned}$$

The weak necessary condition for the minimization of J^ε is

$$\frac{\delta J^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) = \frac{\delta S^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) + \frac{\delta E}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) = 0, \quad \forall \boldsymbol{v} \in C^\infty(\bar\Omega).$$

To obtain a strong optimality formulation, we assume first that $\boldsymbol{w}$ is sufficiently regular. For employing the fundamental Lemma of calculus of variations we have to shift the derivatives from $\boldsymbol{v}$ to $\boldsymbol{w}$; thus, by using partial integration we obtain

$$\begin{aligned}
\frac{\delta E}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) =& \mu\int_{\partial\Omega}\sum_{\ell=1}^{2} v_\ell\,(\nabla w_\ell + \partial_{x_\ell}\boldsymbol{w})\cdot\boldsymbol{n}\, dS - \mu\int_\Omega (\Delta\boldsymbol{w} + \nabla(\nabla\cdot\boldsymbol{w}))\cdot\boldsymbol{v}\, d\boldsymbol{x} \\
&+ \lambda\int_{\partial\Omega}(\nabla\cdot\boldsymbol{w})\,(\boldsymbol{v}\cdot\boldsymbol{n})\, dS - \lambda\int_\Omega \boldsymbol{v}\cdot\nabla(\nabla\cdot\boldsymbol{w})\, d\boldsymbol{x}.
\end{aligned}$$

where $\boldsymbol{n}$ denotes the outer unit normal vector on $\partial\Omega$. The boundary conditions result from letting $\boldsymbol{v}$ be concentrated on $\partial\Omega$ which implies

$$\lambda n_\ell \nabla\cdot\boldsymbol{w} + \mu\,(\nabla w_\ell + \partial_{x_\ell}\boldsymbol{w})\cdot\boldsymbol{n} = 0, \quad \text{on } \partial\Omega. \qquad (3.2)$$

Hence by setting the variational derivative of J^ε to zero we obtain

$$
\begin{aligned}
0 = \frac{\delta J^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w};\boldsymbol{v}) = &-\mu \int_\Omega \left(\Delta \boldsymbol{w} + \nabla\left(\nabla\cdot\boldsymbol{w}\right)\right)\cdot \boldsymbol{v}\, d\boldsymbol{x} - \lambda \int_\Omega \boldsymbol{v}\cdot\nabla\left(\nabla\cdot\boldsymbol{w}\right)\, d\boldsymbol{x} \\
&+ \int_\Omega (\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\cdot\boldsymbol{v}\, d\boldsymbol{x} \\
&+ \mu\int_{\partial\Omega}\sum_{\ell=1}^{2} v_\ell\left(\nabla w_\ell + \partial_{x_\ell}\boldsymbol{w}\right)\cdot\boldsymbol{n}\, dS + \lambda\int_{\partial\Omega}\left(\nabla\cdot\boldsymbol{w}\right)\left(\boldsymbol{v}\cdot\boldsymbol{n}\right)\, dS;
\end{aligned}
\tag{3.3}
$$

Since this holds for any variation $\boldsymbol{v}\in C^\infty(\bar\Omega)$ we may apply the fundamental Lemma of calculus of variations to obtain the Euler–Lagrange equations of the minimization problem (2.7)

$$
\begin{cases}
\left(\mu\Delta\boldsymbol{w} + (\mu+\lambda)\nabla\left(\nabla\cdot\boldsymbol{w}\right)\right) = (\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}), & \text{in } \Omega,\\
\lambda n_\ell \nabla\cdot\boldsymbol{w} + \mu\left(\nabla w_\ell + \partial_{x_\ell}\boldsymbol{w}\right)\cdot\boldsymbol{n} = 0, & \text{on } \partial\Omega.
\end{cases}
\tag{3.4}
$$

We define an operator $\mathcal{E}$ by

$$
\mathcal{E}\boldsymbol{w} := \mu\Delta\boldsymbol{w} + (\mu+\lambda)\nabla\left(\nabla\cdot\boldsymbol{w}\right),
\tag{3.5}
$$

which is the so-called elasticity operator. It gives the inner stress of a given material (determined by the Navier–Lamé constants) under the deformation $\boldsymbol{w}$ (see for instance [23, Chap. 9]). Similarly, we set

$$
\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w};\mathcal{I}_0^\varepsilon,\mathcal{I}_1^\varepsilon) := (\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}),
\tag{3.6}
$$

which is the driving force of the registration. Thus the Euler–Lagrange equations (3.4) become

$$
\mathcal{E}\boldsymbol{w} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{w};\mathcal{I}_0^\varepsilon,\mathcal{I}_1^\varepsilon).
\tag{3.7}
$$

Note that the driving force of the registration is the Gâteaux derivative of the employed distance measure. This gives rise to some important observations. If the force field $\boldsymbol{f}$ is small, a slight deformation $\boldsymbol{w}$ is enough to satisfy (3.7). In the special case of edge-sets that means that the applied distance measure has to be sensitive enough to "capture" differences in the pathway of the edges Γ_i. Here it becomes obvious again that a naive application of the SSID-measure on the original binary images I_i is not feasible for our problem. The driving force generated by $\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w};I_0,I_1)$ is too small to yield a sufficiently large deformation $\boldsymbol{w}$ such that the shape of Γ_0 changes visibly.

However, at the same time the applied distance measure has to be robust so that it is able to "ignore" minor differences in the details of the edges Γ_i in the presence of noise. Though the Hausdorff–distance generates a sufficiently large driving force for an elastic registration it is very sensitive to noise. Only modifications of the Hausdorff–distance-measure (2.3) are capable of matching noisy image data. (see for instance [30] or [36].)

The technique of "region-growing" proposed here is able to satisfy both requirements. Depending on the choice of $\varepsilon > 0$ the broadening of the blurred edges $\tilde\Gamma_i^\varepsilon := \mathrm{supp}(\mathcal{I}_i^\varepsilon)$ for $i = 0, 1$ can be arbitrarily extended; thus provoking a large driving force $\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w};\mathcal{I}_0^\varepsilon,\mathcal{I}_1^\varepsilon)$ if the original edges

Γ_i differ significantly. On the other hand the blurring effect or our "region-growing" approach automatically smooths noisy edges by simultaneously preserving characteristic features. However, an inadequate choice of ε can lead to either excessive blurring and hence loss of important features or insufficient enhancement of the edges and thus a too small driving force $\boldsymbol{f}$. Though the value of ε is therefore critical, the process of finding a suitable ε is usually not too difficult. Nevertheless, in the following we present a solution strategy which is robust against the choice of ε.

4. Solution Strategy

Since the Euler–Lagrange equations (3.4) are a system of nonlinear partial differential equations (PDEs) in $\boldsymbol{w}$ we employ Newton's method on the functional J^ε which takes the form

$$\left\{\begin{aligned} \frac{\delta^2 J^\varepsilon}{\delta \boldsymbol{w}^2}(\boldsymbol{w}_k;\boldsymbol{v},\delta\boldsymbol{w}_k) &= -\frac{\delta J^\varepsilon}{\delta \boldsymbol{w}}(\boldsymbol{w}_k;\boldsymbol{v}), \quad \forall \boldsymbol{v}\in C^\infty(\bar\Omega), \qquad k=1,2,\dots,\\ \boldsymbol{w}_{k+1} &= \boldsymbol{w}_k + \tau\delta\boldsymbol{w}_k, \end{aligned}\right. \tag{4.1}$$

where $\tau > 0$ denotes a given step-size and k is the iteration index. By employing again the linearity of the of the Gâteaux derivative we first compute

$$\frac{\delta^2 S^\varepsilon}{\delta \boldsymbol{w}^2}(\boldsymbol{w}_k;\boldsymbol{v},\delta\boldsymbol{w}_k) = \int_\Omega \boldsymbol{v}\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}_k)\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}_k)^\top\delta\boldsymbol{w}_k\,d\boldsymbol{x}.$$

By using the first variational derivative (3.1) of the linear elastic potential we obtain further

$$\begin{aligned} \frac{\delta^2 E}{\delta \boldsymbol{w}^2}(\boldsymbol{w}_k;\boldsymbol{v},\delta\boldsymbol{w}_k) &= \frac{d}{ds}\int_\Omega \mu\sum_{j=1}^{2}\left(\nabla(w_{k_j}+s\delta w_{k_j}) + \partial_{x_j}(\boldsymbol{w}_k+s\delta\boldsymbol{w}_k)\right)\cdot\nabla v_j\,d\boldsymbol{x}\Big|_{s=0}\\ &\quad + \frac{d}{ds}\int_\Omega \lambda\left(\nabla\cdot(\boldsymbol{w}_k+s\delta\boldsymbol{w}_k)\right)(\nabla\cdot\boldsymbol{v})\;d\boldsymbol{x}\Big|_{s=0}\\ &= \int_\Omega \mu\sum_{j=1}^{2}\left(\nabla\delta w_{k_j} + \partial_{x_j}\delta\boldsymbol{w}_k\right)\cdot\nabla v_j + \lambda\left(\nabla\cdot\delta\boldsymbol{w}_k\right)(\nabla\cdot\boldsymbol{v})\;d\boldsymbol{x}. \end{aligned}$$

Hence, by using the force field $\boldsymbol{f}$ (3.6) we get the following weak form of the Newton step:

$$\begin{aligned} &\int_\Omega \mu\sum_{j=1}^{2}\left(\nabla\delta w_{k_j} + \partial_{x_j}\delta\boldsymbol{w}_k\right)\cdot\nabla v_j\,d\boldsymbol{x}\\ &\qquad + \int_\Omega \lambda\left(\nabla\cdot\delta\boldsymbol{w}_k\right)(\nabla\cdot\boldsymbol{v})\;d\boldsymbol{x}\\ &+\int_\Omega \boldsymbol{v}\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}_k)\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}_k)^\top\delta\boldsymbol{w}_k\,d\boldsymbol{x} = -\int_\Omega \mu\sum_{j=1}^{2}\left(\nabla w_{k_j} + \partial_{x_j}\boldsymbol{w}_k\right)\cdot\nabla v_j\,d\boldsymbol{x}\\ &\qquad\qquad - \int_\Omega \lambda\left(\nabla\cdot\boldsymbol{w}_k\right)(\nabla\cdot\boldsymbol{v})\;d\boldsymbol{x}\\ &\qquad\qquad - \int_\Omega \boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}_k;\mathcal{I}_0^\varepsilon,\mathcal{I}_1^\varepsilon)\cdot\boldsymbol{v}\,d\boldsymbol{x}, \end{aligned} \tag{4.2}$$

for all $\boldsymbol{v} \in C^\infty(\bar{\Omega})$. Using the operator $\mathcal{E}$ and again partial integration and the boundary conditions (3.2) together with the fundamental Lemma of calculus of variations we arrive (under stronger assumptions on the regularity of $\boldsymbol{w}$) at the following strong formulation:

$$\begin{cases} \left(-\mathcal{E} + \nabla \mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}_k)\nabla \mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}_k)^\top\right)\delta\boldsymbol{w}_k = \mathcal{E}\boldsymbol{w}_k - \boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}_k;\mathcal{I}_0^\varepsilon,\mathcal{I}_1^\varepsilon)), & \text{in } \Omega, \\ \lambda n_\ell \nabla\cdot\boldsymbol{w}_k + \mu\left(\nabla w_{k_\ell} + \partial_{x_\ell}\boldsymbol{w}_k\right)\cdot\boldsymbol{n} = 0, & \text{on} \partial\Omega. \end{cases} \tag{4.3}$$

which is a linear PDE-system for the unknown function $\delta\boldsymbol{w}_k$.

4.1. Solvability of the Newton Step

For the sake of brevity we drop the iteration index k for the moment and set $\boldsymbol{u} = \delta\boldsymbol{w}$. We reconsider the weak formulation (4.2) of the Newton step and define the bilinear form

$$\begin{aligned} B(\boldsymbol{u},\boldsymbol{v}) := &\int_\Omega \mu\left(\nabla\boldsymbol{u}^\top + \nabla\boldsymbol{u}\right) : \left(\nabla\boldsymbol{v}^\top + \nabla\boldsymbol{v}\right) + \lambda\left(\nabla\cdot\boldsymbol{u}\right)\left(\nabla\cdot\boldsymbol{v}\right)\, d\boldsymbol{x} \\ &+ \int_\Omega \boldsymbol{u}\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})^\top\boldsymbol{v}\, d\boldsymbol{x}, \end{aligned}$$

and the linear functional

$$\begin{aligned} F(\boldsymbol{v}) := &-\int_\Omega \mu\left(\nabla\boldsymbol{w}^\top + \nabla\boldsymbol{w}\right) : \left(\nabla\boldsymbol{v}^\top + \nabla\boldsymbol{v}\right) + \lambda\left(\nabla\cdot\boldsymbol{w}\right)\left(\nabla\cdot\boldsymbol{v}\right)\, d\boldsymbol{x} \\ &- \int_\Omega (\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\cdot\boldsymbol{v}\, d\boldsymbol{x}, \end{aligned}$$

where : denotes a component-wise matrix scalar product. Then the variational formulation (4.2) of the Newton step takes the form

$$B(\boldsymbol{u},\boldsymbol{v}) = F(\boldsymbol{v}), \quad \forall \boldsymbol{v} \in C^\infty(\bar{\Omega}). \tag{4.4}$$

Now we show well-posedness of the variational problem (4.4).

Theorem 2. *Let $\boldsymbol{v} \in C^\infty(\bar{\Omega})$, $\mathcal{I}_0^\varepsilon \in W^{1,\infty}(\Omega)$, $\mathcal{I}_1^\varepsilon \in L^\infty(\Omega)$ and $\lambda, \mu > 0$.*
If for given $\boldsymbol{w} \in H^1(\Omega)$

$$\int_\Omega \left|\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\cdot(\boldsymbol{a}+M\boldsymbol{x})\right|^2\, d\boldsymbol{x} = 0 \quad \textit{implies} \quad \boldsymbol{a}+M\boldsymbol{x} = 0, \tag{4.5}$$

for every skew-symmetric matrix $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ and every vector $\boldsymbol{a} \in \mathbb{R}^N$ then there exists a unique $\boldsymbol{u} \in H^1(\Omega)$ satisfying (4.4).

Proof. We will show that $B(\boldsymbol{u},\boldsymbol{v})$ is bounded and coercive on $H^1(\Omega)$ and that $F(\boldsymbol{v})$ is bounded on $H^1(\Omega)$. Then due to the fact that $C^\infty(\bar{\Omega})$ is dense in $H^1(\Omega)$ we may apply the Lax–Milgram

Lemma [9] to prove the claim. We start by showing boundedness of B and F:

$$\begin{aligned} |B(\boldsymbol{u},\boldsymbol{v})| \leq & 4\mu \left\|\nabla \boldsymbol{u}\right\|_{L^2(\Omega)} \left\|\nabla \boldsymbol{v}\right\|_{L^2(\Omega)} + \lambda \left\|\nabla \boldsymbol{u}\right\|_{L^2(\Omega)} \left\|\nabla \boldsymbol{v}\right\|_{L^2(\Omega)} \\ & + \left\|\boldsymbol{u}\right\|_{L^2(\Omega)} \left\|\mathcal{I}_0^\varepsilon\right\|_{W^{1,\infty}(\Omega)}^2 \left\|\boldsymbol{v}\right\|_{L^2(\Omega)} \\ \leq & \left(4\mu + \lambda + \left\|\mathcal{I}_0^\varepsilon\right\|_{W^{1,\infty}(\Omega)}^2\right) \left\|\boldsymbol{u}\right\|_{H^1(\Omega)} \left\|\boldsymbol{v}\right\|_{H^1(\Omega)}, \end{aligned} \tag{4.6}$$

and analogously

$$\begin{aligned} |F(\boldsymbol{v})| \leq & \left(\left\|\mathcal{I}_0^\varepsilon\right\|_{L^\infty(\Omega)} + \left\|\mathcal{I}_1^\varepsilon\right\|_{L^\infty(\Omega)}\right) \left\|\mathcal{I}_0^\varepsilon\right\|_{W^{1,\infty}(\Omega)} \left\|\boldsymbol{v}\right\|_{L^2(\Omega)} \\ & + (4\mu + \lambda) \left\|\boldsymbol{w}\right\|_{H^1(\Omega)} \left\|\boldsymbol{v}\right\|_{H^1(\Omega)}. \end{aligned}$$

For showing coercivity of B we define the energy norm

$$\left\|\boldsymbol{u}\right\|_E := \mu \left\|\nabla \boldsymbol{u}^\top + \nabla \boldsymbol{u}\right\|_{L^2(\Omega)}^2 + \lambda \left\|\nabla \cdot \boldsymbol{u}\right\|_{L^2(\Omega)}^2,$$

and obtain

$$B(\boldsymbol{u},\boldsymbol{u}) = \left\|\boldsymbol{u}\right\|_E + \int_\Omega \left|\nabla \mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w}) \cdot \boldsymbol{u}\right|^2 d\boldsymbol{x} \geq \left\|\boldsymbol{u}\right\|_E. \tag{4.7}$$

By using Korn's inequality [31] we have

$$\begin{aligned} \left\|\boldsymbol{u}\right\|_{H^1(\Omega)}^2 \leq & c_1 \left\|\boldsymbol{u}\right\|_{L^2(\Omega)}^2 + K \left\|\nabla \boldsymbol{u}^\top + \nabla \boldsymbol{u}\right\|_{L^2(\Omega)}^2 \\ \leq & c_1 \left\|\boldsymbol{u}\right\|_{L^2(\Omega)}^2 + c_2 \left\|\boldsymbol{u}\right\|_E^2, \end{aligned} \tag{4.8}$$

for some positive constants c_1, K and $c_2 = K/\mu$. Now assume that there exists a sequence $\{\boldsymbol{u}_n\}_{n\geq 1} \subset H^1(\Omega)$ such that

$$\forall n \in \mathbb{N} : \left\|\boldsymbol{u}_n\right\|_{H^1(\Omega)} = 1, \quad \text{and} \quad B(\boldsymbol{u}_n,\boldsymbol{u}_n) \overset{n\to\infty}{\longrightarrow} 0, \tag{4.9}$$

which as a consequence of (4.7) immediately implies

$$\left\|\boldsymbol{u}_n\right\|_E \overset{n\to\infty}{\longrightarrow} 0. \tag{4.10}$$

Due to the compact embedding of $H^1(\Omega)$ in $L^2(\Omega)$ [1] there exists a subsequence $\{\boldsymbol{u}_{n_k}\}_{k\geq 1} \subset \{\boldsymbol{u}_n\}_{n\geq 1}$ such that $\boldsymbol{u}_{n_k} \overset{k\to\infty}{\underset{L^2(\Omega)}{\longrightarrow}} \boldsymbol{u}^\star \in L^2(\Omega)$. But (4.8) even implies

$$\left\|\boldsymbol{u}_{n_k} - \boldsymbol{u}_{n_l}\right\|_{H^1(\Omega)}^2 \leq c_1 \left\|\boldsymbol{u}_{n_k} - \boldsymbol{u}_{n_l}\right\|_{L^2(\Omega)}^2 + c_2 \left\|\boldsymbol{u}_{n_k} - \boldsymbol{u}_{n_l}\right\|_E^2.$$

Since $\{\boldsymbol{u}_{n_k}\}_{k\geq 1}$ converges in $L^2(\Omega)$ and due to (4.10) both terms on the right hand side vanish; thus, $\{\boldsymbol{u}_{n_k}\}_{k\geq 1}$ is in fact a Cauchy–sequence in $H^1(\Omega)$ and hence $\boldsymbol{u}^\star \in H^1(\Omega)$. Further we

observe that due to (4.7) and (4.6)

$$\begin{aligned}\|\boldsymbol{u}^\star\|_E^2 \leq & B(\boldsymbol{u}^\star,\boldsymbol{u}^\star) = B(\boldsymbol{u}^\star-\boldsymbol{u}_{n_k},\boldsymbol{u}^\star)+B(\boldsymbol{u}_{n_k},\boldsymbol{u}^\star-\boldsymbol{u}_{n_k})+B(\boldsymbol{u}_{n_k},\boldsymbol{u}_{n_k})\\ \leq & c_3\,\|\boldsymbol{u}^\star-\boldsymbol{u}_{n_k}\|_{H^1(\Omega)}\left(\|\boldsymbol{u}^\star\|_{H^1(\Omega)}+\|\boldsymbol{u}_{n_k}\|_{H^1(\Omega)}\right)+|B(\boldsymbol{u}_{n_k},\boldsymbol{u}_{n_k})|\,,\end{aligned}$$

where $c_3 := 4\mu+\lambda+\|\mathcal{I}_0^\varepsilon\|^2_{W^{1,\infty}(\Omega)}$. Since $\{\boldsymbol{u}_{n_k}\}_{k\geq 1}$ is a converging sequence in $H^1(\Omega)$ and due to assumption (4.9) the right hand side tends to zero as $k\to\infty$ which gives $B(\boldsymbol{u}^\star,\boldsymbol{u}^\star)=0$. This implies

$$0 = \left(\nabla\boldsymbol{u}^{\star\top}+\nabla\boldsymbol{u}^\star\right)_{i,j} =: e^\star_{i,j},$$

and thus for each component function $u_1^\star$ and $u_2^\star$ of $\boldsymbol{u}^\star$ we have [10]

$$\partial_{x_ix_j}u_k^\star = \partial_{x_i}e^\star_{j,k}+\partial_{x_j}e^\star_{k,i}-\partial_{x_k}e^\star_{i,j}=0,$$

which implies that $\boldsymbol{u}^\star$ is a linear function of $\boldsymbol{x}$. Hence we set $\boldsymbol{u}^\star := \boldsymbol{a}+\boldsymbol{M}\boldsymbol{x}$ with $\boldsymbol{a}\in\mathbb{R}^N$ and $\boldsymbol{M}\in\mathbb{R}^{N\times N}$. Since $\nabla\boldsymbol{u}^{\star\top}+\nabla\boldsymbol{u}^\star = \boldsymbol{M}^\top+\boldsymbol{M}=0$ the matrix $\boldsymbol{M}$ is skew-symmetric. Thus by employing (4.7) we obtain

$$0 = B(\boldsymbol{u}^\star,\boldsymbol{u}^\star)\geq\int_\Omega|\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\cdot\boldsymbol{u}^\star|^2\,d\boldsymbol{x} = \int_\Omega|\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\cdot(\boldsymbol{a}+\boldsymbol{M}\boldsymbol{x})|^2\,d\boldsymbol{x},$$

which according to the initial assumption (4.5) implies $\boldsymbol{u}^\star=\boldsymbol{a}+\boldsymbol{M}\boldsymbol{x}=0$. On the other hand according to (4.9) we have for all $n\in\mathbb{N}$ that $\|\boldsymbol{u}_n\|_{H^1(\Omega)}=1$ which is a contradiction. Hence (4.9) is invalid and B is coercive on $H^1(\Omega)$. □

Based on the proof given above the convergence of Newton's method (4.1) for a fixed $\varepsilon>0$ and a suitable initial guess $\boldsymbol{w}_1$ can be showed using the Newton–Kantorovich Theorem (see for instance [26] or [13]).

5. Numerical Approximation

We will first introduce a discretization scheme for the strong formulation (4.3) of the Newton step and then explain the discrete realization of Newton's method (4.1). We start by rewriting the Euler-Lagrange equations (3.7). Let from now on u and v denote the component functions of $\boldsymbol{w}$, i.e., $\boldsymbol{w}(\boldsymbol{x}) := (u(\boldsymbol{x}),v(\boldsymbol{x}))^\top$, and let $\boldsymbol{x}=(x,y)\in\Omega$. Then we obtain for the elasticity operator $\mathcal{E}$

$$\begin{aligned}\mathcal{E}\boldsymbol{w} =& \mu\Delta\boldsymbol{w}+(\mu+\lambda)\nabla\left(\nabla\cdot\boldsymbol{w}\right)\\ =& \mu\left(\frac{\partial^2\boldsymbol{w}}{\partial x^2}+\frac{\partial^2\boldsymbol{w}}{\partial y^2}\right)+(\mu+\lambda)\nabla\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)\\ =& \mu\begin{pmatrix}\frac{\partial^2}{\partial x^2}u+\frac{\partial^2}{\partial y^2}u\\ \frac{\partial^2}{\partial x^2}v+\frac{\partial^2}{\partial y^2}v\end{pmatrix}+(\mu+\lambda)\begin{pmatrix}\frac{\partial^2}{\partial x^2}u+\frac{\partial^2}{\partial x\partial y}v\\ \frac{\partial^2}{\partial x\partial y}u+\frac{\partial^2}{\partial x^2}v\end{pmatrix}\\ =& \begin{pmatrix}(\lambda+2\mu)\frac{\partial^2}{\partial x^2}+\mu\frac{\partial^2}{\partial y^2} & (\lambda+\mu)\frac{\partial^2}{\partial x\partial y}\\ (\lambda+\mu)\frac{\partial^2}{\partial x\partial y} & \mu\frac{\partial^2}{\partial x^2}+(\lambda+2\mu)\frac{\partial^2}{\partial y^2}\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}.\end{aligned}$$

We define the following operators which act on the component functions u and v of $\boldsymbol{w}$

$$\begin{aligned}\mathcal{E}_{11} :=&(\lambda+2\mu)\frac{\partial^2}{\partial x^2}+\mu\frac{\partial^2}{\partial y^2},\\ \mathcal{E}_{12}=\mathcal{E}_{21} =&(\lambda+\mu)\frac{\partial^2}{\partial x\partial y},\\ \mathcal{E}_{22} =&\mu\frac{\partial^2}{\partial x^2}+(\lambda+2\mu)\frac{\partial^2}{\partial y^2},\end{aligned}$$

and thus we can write

$$\mathcal{E}\boldsymbol{w}=\begin{pmatrix}\mathcal{E}_{11} & \mathcal{E}_{12}\\ \mathcal{E}_{21} & \mathcal{E}_{22}\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}. \tag{5.1}$$

Similarly we obtain the following expression for the force field $\boldsymbol{f}$

$$\begin{aligned}f(\boldsymbol{x},\boldsymbol{w}) :=&\begin{pmatrix}f_1(\boldsymbol{x},\boldsymbol{w})\\ f_2(\boldsymbol{x},\boldsymbol{w})\end{pmatrix}\\ =&\begin{pmatrix}(\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})-\mathcal{I}_1^\varepsilon(\boldsymbol{x}))\frac{\partial}{\partial x}\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\\ (\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})-\mathcal{I}_1^\varepsilon(\boldsymbol{x}))\frac{\partial}{\partial y}\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})\end{pmatrix}.\end{aligned} \tag{5.2}$$

5.1. Discretization of the Newton Step

For the sake of simplicity we drop again the iteration index k in this subsection. Since we are working with digital images we define a grid $\Omega_h := \{1,\dots,N\}^2$, where N denotes the resolution of the images. We use a unit step size $h := 1$, i.e., the width of a cell is one, and employ standard central finite differences to discretize the Newton step (4.3). In particular let $\boldsymbol{j} := (j_1,\dots,j_N) \in \mathbb{R}^N$ be an integer component multi index, $\mathbf{1} := (1,\dots,1)^\top \in \mathbb{R}^N$ and the cell centroids be given by $\boldsymbol{x_j} := \boldsymbol{j}$, $\mathbf{1} \le \boldsymbol{j} \le N\cdot\mathbf{1}$. We denote the array arising from evaluating e.g. u at each grid point by $u(\Omega_h) \in \mathbb{R}^{N\times N}$. Hence $U_{\boldsymbol{j}} \approx u(\boldsymbol{x_j})$ and $\vec{u} \in \mathbb{R}^{N^2}$ denotes the vector of values $\{U_{\boldsymbol{j}}\}$ corresponding to the lexicographic ordering in which j_1 increments first from 1 to N, then j_2 and so on. Further, let $\boldsymbol{D}(\vec{u}) \in \mathbb{R}^{N^2\times N^2}$ be the diagonal matrix arising from situating the values $\{U_{\boldsymbol{j}}\}$ along the diagonal according to lexicographic ordering.

We start by discretizing the elasticity operator $\mathcal{E}$. We show for instance the discretization of $\mathcal{E}_{11}$ near the lower left corner of Ω_h

$$\begin{array}{ccc}\vdots & & \vdots\\ \begin{pmatrix}0 & 0 & -2\mu\\ 0 & 8\mu+4\lambda & -4\mu-4\lambda\\ 0 & 0 & -2\mu\end{pmatrix} & \begin{pmatrix}-2\mu & 0 & -2\mu\\ -4\mu-4\lambda & 16\mu+8\lambda & -4\mu-4\lambda\\ -2\mu & 0 & -2\mu\end{pmatrix} & \cdots\\ \begin{pmatrix}0 & 0 & -2\mu\\ 0 & 2\lambda+4\mu & -2\mu-2\lambda\\ 0 & 0 & 0\end{pmatrix} & \begin{pmatrix}-2\mu & 0 & -2\mu\\ -2\mu-2\lambda & 8\mu+4\lambda & -2\mu-2\lambda\\ 0 & 0 & 0\end{pmatrix} & \cdots\end{array} \tag{5.3}$$

The upper right block represents the stencil weights for neighbors of a field cell. Similarly the other blocks show for boundary cells the stencil weights for their neighbors. With the same

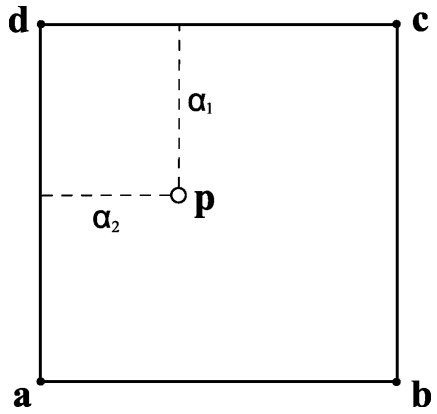


Figure 5.1: Sketch illustrating two dimensional bilinear interpolation.

format we represent the stencils of $\mathcal{E}_{12}$:

$$\begin{array}{cc} \vdots & \vdots \\ \begin{pmatrix} 0 & \mu-\lambda & -\mu-\lambda \\ 0 & 0 & 0 \\ 0 & -\mu-\lambda & \mu+\lambda \end{pmatrix} & \begin{pmatrix} \mu+\lambda & 0 & -\mu-\lambda \\ 0 & 0 & 0 \\ -\mu-\lambda & 0 & \mu+\lambda \end{pmatrix} \quad \cdots \\ \begin{pmatrix} 0 & \mu-\lambda & -\mu-\lambda \\ 0 & \mu+\lambda & -\mu+\lambda \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} \mu+\lambda & 0 & -\mu-\lambda \\ \mu-\lambda & 0 & -\mu-\lambda \\ 0 & 0 & 0 \end{pmatrix} \quad \cdots \end{array} \tag{5.4}$$

The stencils for $\mathcal{E}_{22}$ and $\mathcal{E}_{21}$ are constructed by adequate copying and mirroring of (5.3) and (5.4) respectively. This gives rise to matrices $\boldsymbol{E}_{k.\ell} \in \mathbb{R}^{N^2 \times N^2}$ with $1 \leq k, \ell \leq 2$ which form the discrete version of the operator $\mathcal{E}$ under lexicographic ordering, i.e.,

$$\mathcal{E}(u(\Omega_h), v(\Omega_h)) \approx \begin{pmatrix} \boldsymbol{E}_{11}\vec{u} + \boldsymbol{E}_{12}\vec{v} \\ \boldsymbol{E}_{21}\vec{u} + \boldsymbol{E}_{22}\vec{v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix} \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix},$$

which we abbreviate by setting

$$\boldsymbol{E} := \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix} \in \mathbb{R}^{2N^2 \times 2N^2} \quad \text{and} \quad \vec{w} := \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix} \in \mathbb{R}^{2N^2},$$

so that we may write

$$\boldsymbol{E}\vec{w} = \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix} \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix}.$$

Now we develop the discretization of the images $\mathcal{I}_i^\varepsilon$ and the gradient $\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})$. Note that for non-integer values of $\boldsymbol{w}$ a (multi)linear interpolation scheme has to be applied to compute $\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})$ and $\nabla\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})$. (This partially justifies the assumption on $\mathcal{I}_0^\varepsilon$ being a $W^{1,\infty}(\Omega)$-function.) Let $\boldsymbol{p} := \boldsymbol{x_j} + \boldsymbol{w_j}$ and the nodes of the cell enclosing $\boldsymbol{p}$ be denoted by $\boldsymbol{a}$ (lower left), $\boldsymbol{b}$ (lower right), $\boldsymbol{c}$ (upper right) and $\boldsymbol{d}$ (upper left); a sketch is given in Figure 5.1. Then the

value of $\mathcal{I}_0^\varepsilon$ at $\boldsymbol{p}$ can be computed by employing the following (bi)linear interpolation scheme

$$\begin{aligned}\mathcal{I}_0^\varepsilon(\boldsymbol{p}) = \mathcal{I}_0^\varepsilon(\boldsymbol{x_j} + \boldsymbol{w_j}) \approx &\mathcal{I}_0^\varepsilon(\boldsymbol{d})(1-\alpha_1)(1-\alpha_2) + \mathcal{I}_0^\varepsilon(\boldsymbol{a})\alpha_1(1-\alpha_2) \\ &+ \mathcal{I}_0^\varepsilon(\boldsymbol{c})(1-\alpha_1)\alpha_2 + \mathcal{I}_0^\varepsilon(\boldsymbol{b})\alpha_1\alpha_2 =: \tilde{\mathcal{I}}_{0,\boldsymbol{j}},\end{aligned} \tag{5.5}$$

where $\alpha_i := p_i - d_i \in [0,1]$ for $i = 1,2$. If $\boldsymbol{p}$ lies outside of Ω we assign the extrapolation value zero. Using this we introduce the matrix $\tilde{\mathcal{I}}_0 \in \mathbb{R}^{N\times N}$ of values $\left\{\tilde{\mathcal{I}}_{0,\boldsymbol{j}}\right\}$. The discrete version $\tilde{\mathcal{I}}_1 \in \mathbb{R}^{N\times N}$ of $\mathcal{I}_1^\varepsilon$ is readily established by setting $\tilde{\mathcal{I}}_{1,\boldsymbol{j}} := \mathcal{I}_1^\varepsilon(\boldsymbol{x_j})$.

Finally we approximate the gradient $\nabla\mathcal{I}_0^\varepsilon$. Let δh denote a fraction of the cell width h, i.e., $\delta h := h/2 = 1/2$. Employing this we get the increment of $\mathcal{I}_0^\varepsilon(\boldsymbol{x_j}+\boldsymbol{w_j})$ in x-direction to compute the following discrete derivative

$$\begin{aligned}\frac{\partial}{\partial x}\mathcal{I}_0^\varepsilon(\boldsymbol{x_j} + \boldsymbol{w_j}) \approx &\frac{1}{2\delta h}\left(\mathcal{I}_0^\varepsilon(\boldsymbol{x_j} + e_1\delta h + \boldsymbol{w_j})\right. \\ &\left.-\mathcal{I}_0^\varepsilon(\boldsymbol{x_j} - e_1\delta h + \boldsymbol{w_j})\right) =: \delta_x\tilde{\mathcal{I}}_{0,\boldsymbol{j}},\end{aligned}$$

where e_1 is the first unit vector in $\mathbb{R}^2$. Note that we make again use of the introduced interpolation scheme (5.5) if $\mathcal{I}_0^\varepsilon$ is evaluated at non-grid points. Analogously we compute $\delta_y\tilde{\mathcal{I}}_{0,\boldsymbol{j}}$ and are finally ready to set up the discrete Newton step. With : denoting again a componentwise matrix scalar product we define the discrete symmetrized gradient

$$\boldsymbol{\mathcal{D}}(\tilde{\mathcal{I}}_0) := \begin{pmatrix} \boldsymbol{D}\left(\overrightarrow{\delta_x\tilde{\mathcal{I}}_0 : \delta_x\tilde{\mathcal{I}}_0}\right) & \boldsymbol{D}\left(\overrightarrow{\delta_x\tilde{\mathcal{I}}_0 : \delta_y\tilde{\mathcal{I}}_0}\right) \\ \boldsymbol{D}\left(\overrightarrow{\delta_x\tilde{\mathcal{I}}_0 : \delta_y\tilde{\mathcal{I}}_0}\right) & \boldsymbol{D}\left(\overrightarrow{\delta_y\tilde{\mathcal{I}}_0 : \delta_y\tilde{\mathcal{I}}_0}\right) \end{pmatrix} \in \mathbb{R}^{2N^2\times 2N^2},$$

and use the discretized images in the force field (5.2) to define

$$\vec{f} := \begin{pmatrix} \overrightarrow{(\tilde{\mathcal{I}}_0 - \tilde{\mathcal{I}}_1) : \delta_x\tilde{\mathcal{I}}_0} \\ \overrightarrow{(\tilde{\mathcal{I}}_0 - \tilde{\mathcal{I}}_1) : \delta_y\tilde{\mathcal{I}}_0} \end{pmatrix} \in \mathbb{R}^{2N^2}.$$

Then the Newton step (4.3) is discretized by

$$\left(-\boldsymbol{E} + \boldsymbol{\mathcal{D}}(\tilde{\mathcal{I}}_0)\right)\delta\vec{w} = -\boldsymbol{E}\vec{w} - \vec{f}, \tag{5.6}$$

which is a linear equation system in the unknown $\delta\vec{w} \in \mathbb{R}^{2N^2}$.

5.2. The Discrete Newton Iteration

Now we want to establish a discrete version of Newton's method (4.1) for the functional J^ε. We use $\vec{w}_1 := 0 \in \mathbb{R}^{2N^2}$as initial guess and employ the discrete Newton step (5.6) to obtain the following iteration

$$\begin{cases} \left(-\boldsymbol{E} + \boldsymbol{\mathcal{D}}(\tilde{\mathcal{I}}_{0,k})\right)\delta\vec{w}_k = -\boldsymbol{E}\vec{w}_k - \vec{f}_k, & \\ \vec{w}_{k+1} = \vec{w}_k + \tau_k\delta\vec{w}_k, & k = 1,2,\dots \end{cases} \tag{5.7}$$

Note that the discrete gradient $\boldsymbol{\mathcal{D}}(\tilde{\mathcal{I}}_{0,k})$ and the discrete force $\vec{f}_k$ depend on the current value of $\vec{w}_k$ hence they have to be updated at each iteration.

We use a backtracking-like line search to determine the step size τ_k. We establish a discrete approximation J_h^ε of the cost functional J^ε and determine τ_k as follows:

$$\begin{cases} \tau_k = \min\limits_{\tau\in\mathcal{T}} J_h^\varepsilon(\vec{w}_k + \tau\delta\vec{w}_k), \\ \mathcal{T} := \left\{\tau = \dfrac{2\ell}{L} | \ell = 1,\dots,L\right\}, \end{cases} \tag{5.8}$$

where $\vec{w}_k$ denotes the current iterate and $\delta\vec{w}_k$ the currently computed Newton direction. Note that this line-search algorithm determines the optimal step-size on the interval $[2/L, 2]$ hence contrary to classic backtracking approaches [11] step sizes larger than one can be chosen. This method has proven to provide good performance and less total computational cost than standard Armijo–Goldstein or Wolfe–Powell techniques [25]. Since $L > 0$ is not chosen to be particularly large, only a few evaluations of J_h^ε are computed to obtain τ_k and the computation of J_h^ε does not involve the expensive calculation of the discrete gradient $\boldsymbol{\mathcal{D}}(\tilde{\mathcal{I}}_{0,k})$.

We use a relative stopping criterion in the Newton iteration (5.7). Note that the right hand side of the Newton step (5.6) corresponds to the discretized Euler–Lagrange equations (3.4) of the original minimization problem (2.7). Hence we want the residual $r_k := -E\vec{w}_k - \vec{f}_k$ to be "small" and thus define the relative residual

$$r_b := \frac{|r_k|}{|r_1|}.$$

On the other hand the algorithm should terminate if the minimization comes to a standstill, i.e., if the distance between two successive iterates is "small". Hence we define

$$r_e := \frac{|\vec{w}_k - \vec{w}_{k-1}|}{|\vec{w}_k|}, \quad \text{and } r_e := 0, \quad \text{if } |\vec{w}_k| = 0,$$

and combine these two notions in a stopping criterion. Let $\mathrm{tol} > 0$ be a predefined tolerance. We want the algorithm to terminate if either

$$\min(r_b, r_e) < \mathrm{tol}, \tag{5.9}$$

or $k > k_{\max}$, where $k_{\max}$ denotes the maximal number of iterations.

As mentioned above the choice of ε in the definition of $\mathcal{I}_i^\varepsilon$ (2.4) can have negative effects on the quality of the diffuse approximations $\mathcal{I}_i^\varepsilon$ of the original images I_i. Whereas it is usually not too difficult to avoid too small values of ε it is very hard to give an upper bound for ε beforehand. However, a value of ε too large leads to excessively blurred approximations $\mathcal{I}_i^\varepsilon$ of I_i and hence a loss of potentially important features of the original edges Γ_i. Hence we want the solution strategy of the registration problem (2.7) to be robust against the choice of ε. To account for the introduced approximation error we augment Newton's method (5.7) with an outer iteration in which we reduce blurring. We compute a rough solution $\vec{w}^\star$ of (5.7) by using a low number of maximal iterations $k_{\max}$. To reduce blurring in the images we compute the element-wise-square of $\tilde{\mathcal{I}}_i^\varepsilon$ which accentuates the original edge sets Γ_i in contrast to their

Algorithm 5.1 Iterative method to solve the elastic registration problem (2.7).

Choose $\varepsilon > 0$, $k_{\text{inc}} \in \mathbb{N} : k_{\text{inc}} \geq 2$ and $\vec{w} \in \mathbb{R}^{2N^2}$.

1. Given the edge-sets Γ_0 and Γ_1 embed them in the center of images I_0 and I_1.
2. Compute diffuse versions $\mathcal{I}_i^\varepsilon$ of I_i.
3. **for** $\kappa = 1$ **to** K
   a) Set $k_{\max} = k_{\text{inc}} \cdot \kappa$
   b) Compute $\vec{w}^\star$ according to (5.7) using the line search strategy (5.8) until either the stopping criterion (5.9) is satisfied or $k > k_{\max}$.
   c) Set $\tilde{\mathcal{I}}_i^\varepsilon(\boldsymbol{x_j}) := \tilde{\mathcal{I}}_i^\varepsilon(\boldsymbol{x_j})^2$ and $\vec{w}_1 = \vec{w}^\star$.
4. **end**
5. Set $\vec{w} = \vec{w}^\star$.

blurred surroundings. Then we restart Newton's method with the images $\tilde{\mathcal{I}}_i^\varepsilon(\boldsymbol{x_j})^2$, $\vec{w}_1 = \vec{w}^\star$ and increase $k_{\max}$. We repeat this procedure K times such that the squared images are sufficiently close approximations to the original images, i.e., $\tilde{\mathcal{I}}_i^\varepsilon(\boldsymbol{x_j})^{2^K} \approx I_i(\boldsymbol{x_j})$. To fix ideas we summarize this method in Algorithm 5.1.

The centering of the edges as stated in Algorithm 5.1 is an important pre-registration step: if the edges are not centered within the images I_i the diffuse extensions of Γ_i created by (2.4) may intersect image boundaries. This would greatly impair the computation of the deformation $\vec{w}$ at the boundary since mass having diffused over the boundary cannot be matched with its natural counterpart in the other image. But even if the edges Γ_i did not "diffuse" out of the image, the centering prior to the approximation has beneficial effects on the performance of the proposed solution strategy. Specifically since both edges Γ_0 and Γ_1 share roughly the same area of the images I_0 and I_1, respectively, the deformation field does not have to account for large translations within the image domain.

Altogether this approach has proven itself in practice. Since the minimization problem (2.7) is only approximately solved for blurred versions $\tilde{\mathcal{I}}_i^\varepsilon$ of I_i in the beginning, a rough initial deformation field is obtained that captures only global deformations in I_0. By successively deblurring the images $\tilde{\mathcal{I}}_0^\varepsilon$ and $\tilde{\mathcal{I}}_1^\varepsilon$ the pathway of the original edge sets Γ_i takes shape again, and the deformation $\vec{w}$ is updated to capture local features of the edges.

6. Computational Results

All computations were carried out in MATLAB$^{\text{TM}}$ 2009b running on a Dell Optiplex 745 equipped with 8 GB of RAM. The operating system used was openSUSE 11.2 (64bit, kernel 2.6.31.12-0.2).

We want to apply the solution strategy developed in the previous section to register the San Diego rabbit heart [34] (referred to as "TBunnyC" in the following) to an anatomically highly realistic model [5] (called "Oxford-heart" from now on). Both models are shown in Figure 6.1. The basic idea is to slice up the 3D models which gives rise to 2D edges. Then each slice is registered consecutively by employing Algorithm 5.1. The registered slices are assembled again

(a) (b)

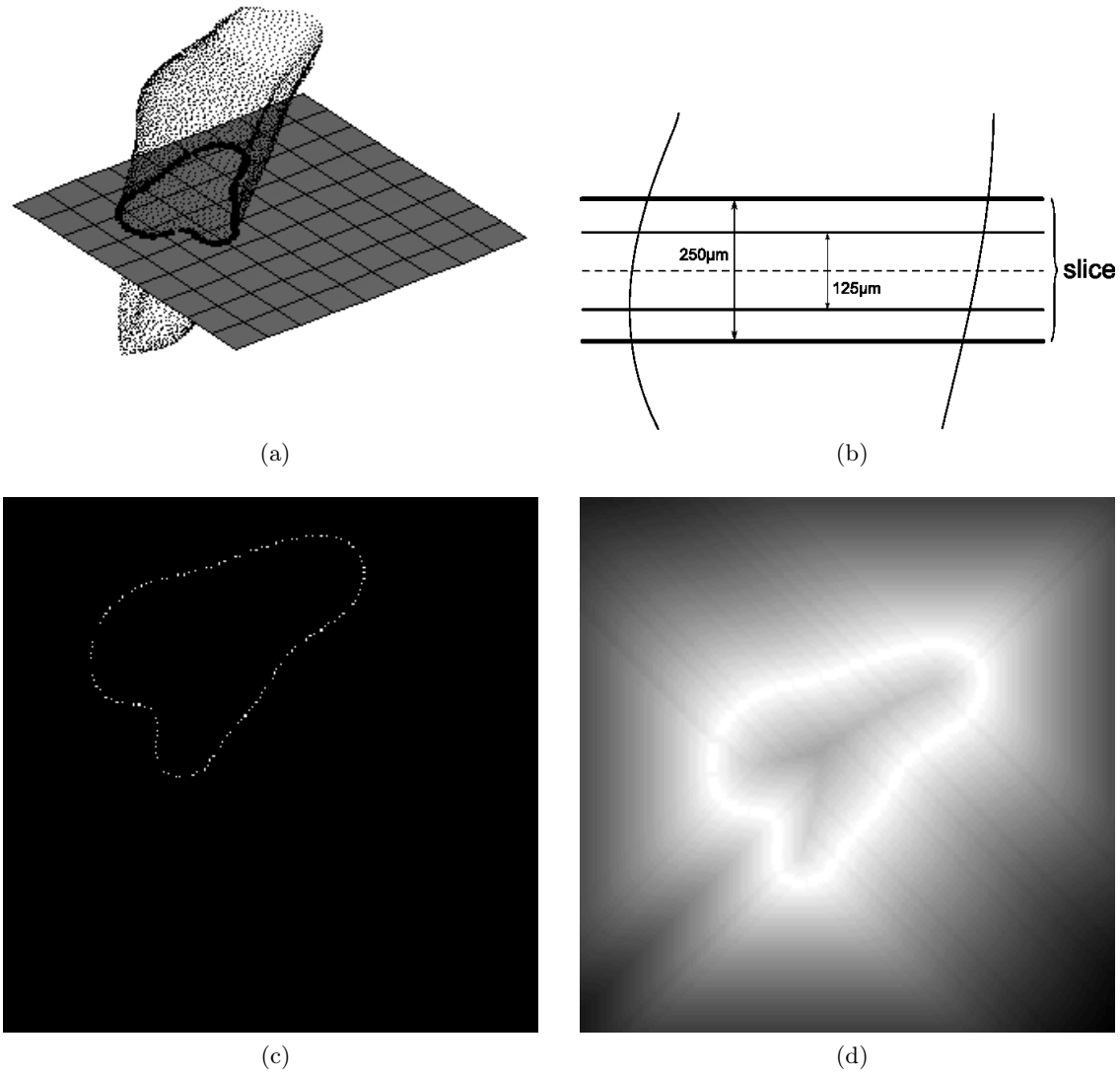
Figure 6.1: The given 3D heart models: (a) The San Diego rabbit heart [34] "TBunnyC" (29 394 points) and (b) the anatomically highly realistic model [5] "Oxford Heart" (258 178 points)

to obtain a 3D representation of the results. Finally the deformation fields computed for each slice are used to map a literature-based Purkinje fiber network [35] onto the endocardium of the Oxford-heart.

Both models are given as ASCII-files containing the Cartesian coordinates of the grid points constituting the surface meshes seen in Fig. 6.1. The measuring unit of the models is μm; hence, we rescaled the large values in the models to reduce the size of the arising linear systems in the registration. Further, to obtain a maximal spatial "overlap" we translated the models. Since we want to apply the proposed registration scheme for edge sets we "cut" the 3D models to obtain 2D edges. To ensure a clear differentiation between left and right cavity we separated the heart models accordingly. We chose the z-axis as cutting direction which seems to be a natural choice once left and right cavities are considered separately.

Due to the fact that the grid points are not equally distributed on regular xy-planes in the vertical direction we have to divide the models into layers of finite thickness, particularly with a thickness of 250μm. The desired 2D edges arise from "squeezing" the thickness until the slice becomes an xy-plane.

Having generated a series of slices we are ready to apply Algorithm 5.1. Each slice is a sparse double MATLAB matrix of size $N \times N$. We apply two preprocessing steps. For each slice we first center the edges within the images and second we employ a "region-growing" algorithm to obtain blurred approximations $\mathcal{I}_i^\varepsilon$. We approximate the distance function d_{Γ_i} used in the definition (2.4) of $\mathcal{I}_i^\varepsilon$ by the following marching scheme. Each non-zero entry of I_i is multiplied by N and added to its 3×3 neighborhood. This procedure is repeated until either the edge set "outgrows" the image or a maximal number of iterations has been exceeded. An exemplary result of this region-growing is depicted in Figure 6.2(d). This way we generate two image series: the blurred slices of the 3D model TBunnyC $\{\mathcal{I}_0^{\varepsilon,m}\}_{m=1}^M$ which serve as template image stack and



(a) (b)

(c) (d)

Figure 6.2: Illustration of the dissection of the 3D heart models. Shown is the left cavity of the TBunnyC model. First the model is cut in vertical direction (a). Only points in the innermost layer of a slice are gathered (b) to generate a 2D edge set (c). We center the image and apply our region-growing algorithm (d).

the blurred cuts of the Oxford-heart $\{\mathcal{I}_1^{\varepsilon,m}\}_{m=1}^M$ which form the reference stack. Thus we want to find elastic deformations $\boldsymbol{w}_m$ so that $\mathcal{I}_0^{\varepsilon,m}(\boldsymbol{x}+\boldsymbol{w}_m) \approx \mathcal{I}_1^{\varepsilon,m}(\boldsymbol{x})$ for $m = 1,\ldots,M$. This is achieved by employing a MATLAB-implementation of the augmented Newton method depicted in Algorithm 5.1. A list of all used parameters and their values is given in Table 6.1. The discrete Newton step (5.6) is solved using the built-in MATLAB-function `mldivide` ("backslash"). To utilize memory efficiently the coefficient matrices in each Newton step are sparse double MATLAB matrices. Then we apply the computed deformations $\boldsymbol{w}_m$ to the original images I_0^m to register the edges Γ_0^m to Γ_1^m. Figure 6.3 sketches the procedure for a single slice. Figure 6.4 shows the 3D reconstruction of the registered TBunnyC slices in comparison to the 3D reference, i.e., the Oxford heart, for both left and right cavities.

Having registered each slice of the left and right cavities we want to utilize the computed

Parameter	Value	Description
λ	1e-2	Navier–Lamé constant, see (2.6)
μ	1e-2	Navier–Lamé constant, see (2.6)
L	10	Linesearch parameter, see (5.8)
tol	1e-3	Tolerance for stopping criterion, see (5.9)
K	4	Number of outer iterations in Algorithm 5.1
k_{inc}	5	Increment for $k_{\max}$ in Algorithm 5.1

Table 6.1: The used parameters.

deformation fields to map an artificial Purkinje fiber network given for the TBunnyC model onto the Oxford heart. The Purkinje fiber network was given as a list of Cartesian coordinates of 832 points (414 points for the left, 418 points for the right cavity respectively) representing the spatial locations of the nodes of a 3D graph. Figure 6.5 shows the Purkinje fiber network in the TBunnyC model in comparison to the registered network in the Oxford heart.

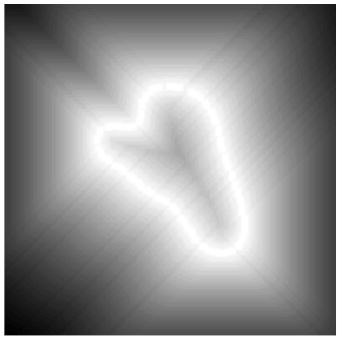
7. Discussion

The artificial PS considered here was modeled to be a subset of the TBunnyC model. Hence we may safely assume that the computed deformations constitute a suitable mapping for projecting the network nodes from the TBunnyC model onto the Oxford heart. The application required the registered network to be a subset of the Oxford heart. This requirement together with the sheer number of network nodes rendered it impossible to construct an affine linear mapping to roughly project the nodes into a proximity of the Oxford endocardium and successively correcting the spatial position of each individual point. Moreover, the complex geometry of the Oxford heart which differs significantly from the topology of the TBunnyC model proved to be captured adequately only by elastically deforming the TBunnyC endocardial walls. Hence the computed elastic deformations guarantee that despite even large differences in the endocardial geometries of both models, the artificial Purkinje fiber network is mapped sufficiently close to the Oxford endocardium.

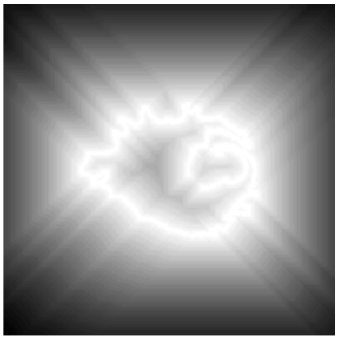
Our simulations confirmed that separating the cavities of the heart models is crucial. This can be seen rather easily by looking at Fig. 6.1. We see that cutting the hearts in the vertical direction gives rise to slices which contain edges from both the left and right cavities. This fact seriously impairs the outcome of the registration. Depending on the choice of the Navier–Lamé constants and the proximity of the cavities, the computed deformation fields "pull" edges corresponding to the left cavity to an edge arising from a cut through the right cavity and vice versa. Despite numerous tests using different parameter values and cutting directions this undesirable effect could not be completely eliminated. Thus we separated the 3D hearts into left and right cavities.

It should be noted that the diameter of the slices used to dissect the 3D heart models has a considerable impact on the performance of the registration. A large layer-diameter introduces a modeling error which may not be negligible. On the other hand thinner diameters increase the total number of slices necessary for a complete dissection of the models, which increases the computational effort for the registration considerably. Therefore we employed the following

Template image I_0 | Reference image I_1



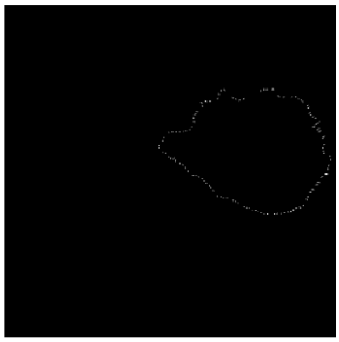
(a) $\mathcal{I}_0^\varepsilon$



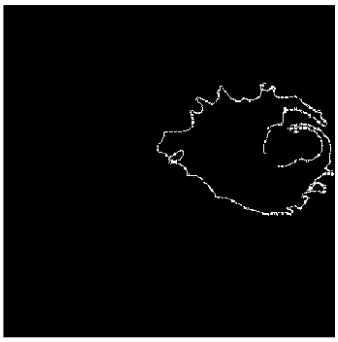
(b) $\mathcal{I}_1^\varepsilon$

(c) $\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})^{2^K}$

(d) $\mathcal{I}_1^\varepsilon(\boldsymbol{x})^{2^K}$



(e) $I_0(\boldsymbol{x}+\boldsymbol{w})$



(f) $I_1(\boldsymbol{x})$

Figure 6.3: The different stages of Algorithm 5.1 for the template image I_0 (left column) and the reference image I_1(right column). The top row shows the centered and region grown versions of the images. The middle row presents the resulting image $\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{w})^{2^K}$ after completion of Algorithm 5.1 versus the final reference $\mathcal{I}_1^{\varepsilon 2^K}$. The bottom row depicts the original image I_0 after application of the computed transformation $\boldsymbol{w}$ versus the original target I_1.

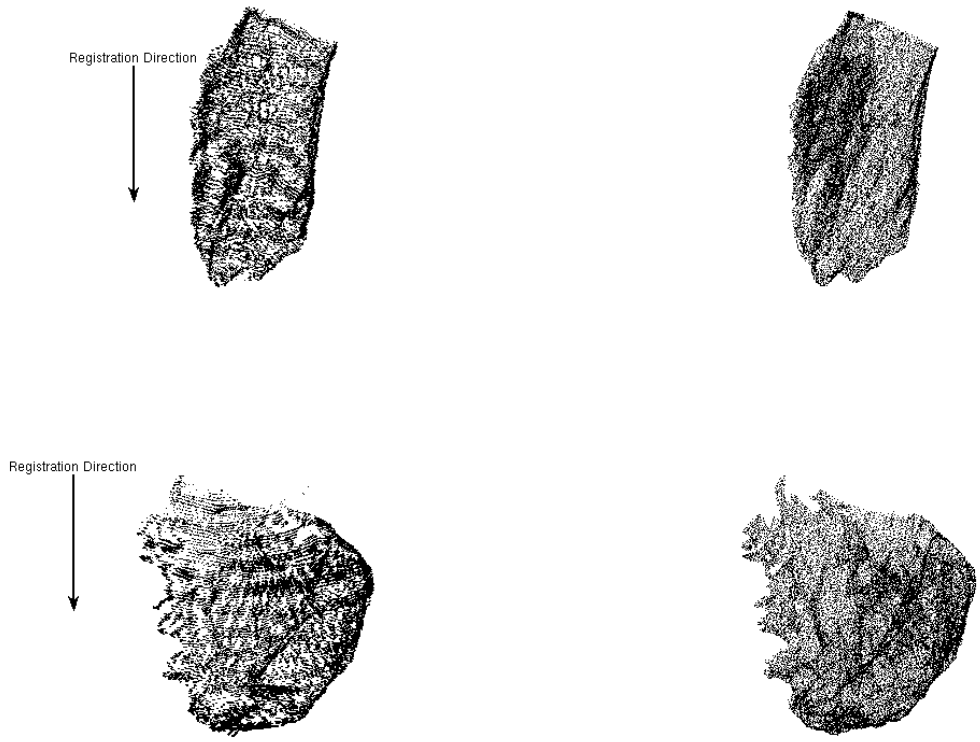
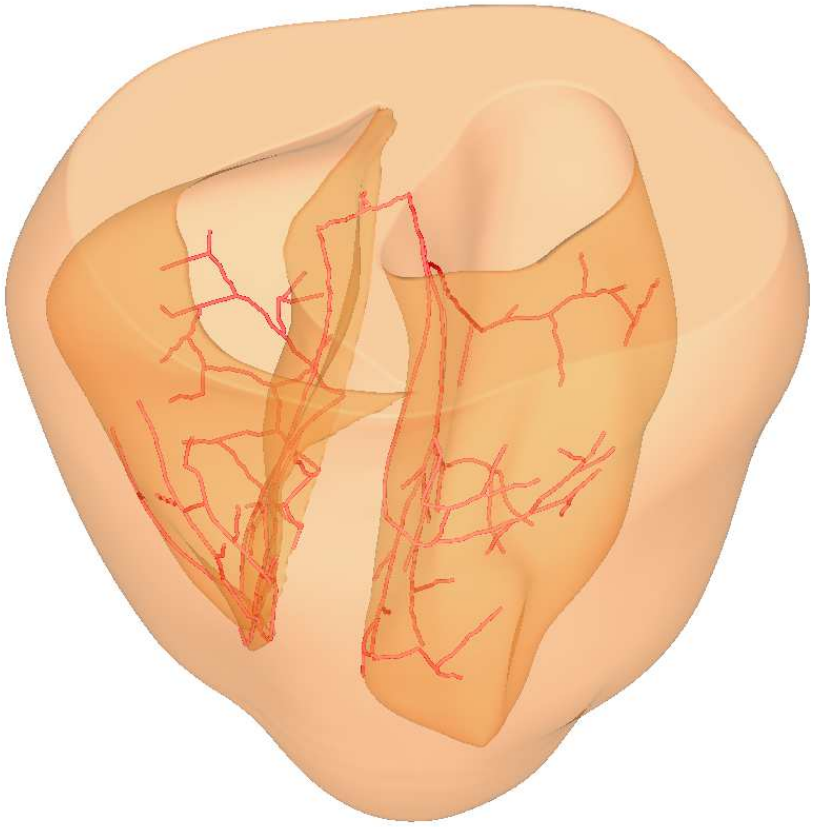


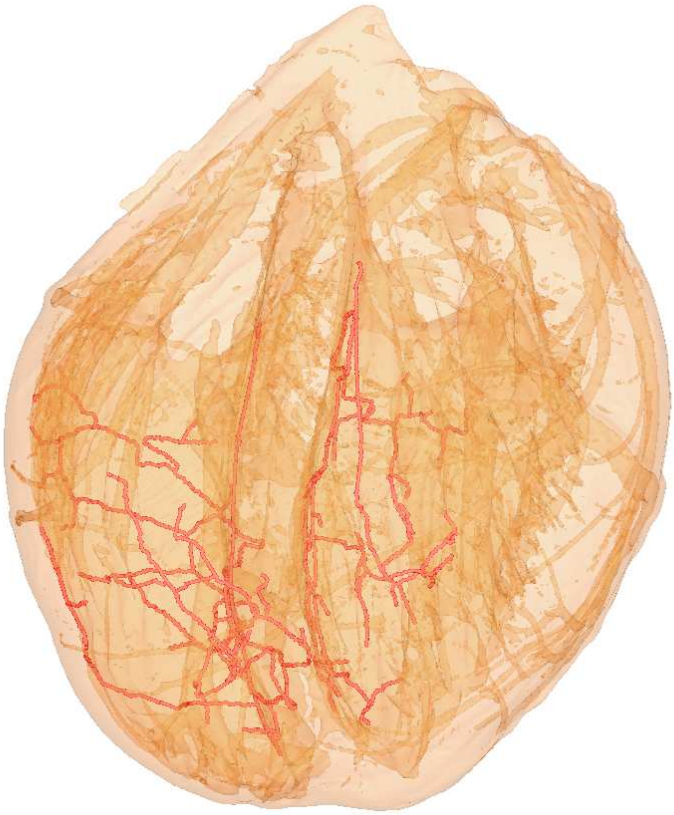
Figure 6.4: 3D Reconstruction of the registered TBunnyC slices (left column) vs the reference Oxford heart (right column). The top row shows the left, the bottom row the right cavity respectively.

scheme. We keep the diameter of 250μm and hence the total number of slices fixed, but to the midway of the slice we assign all points within 125μm. By shrinking the threshold of 125μm we come closer and closer to the selected planes. Including all points in the layer gives rise to excessive displacement of grid points. Figure 6.2 shows this procedure for a single slice.

Despite the application presented here, our method proved to be a highly efficient and reliable technique to register 2D edges. In contrast to previously developed techniques involving the Hausdorff-distance, our approach is computationally cheap and not limited to rigid transformations. The approximation of edges by diffuse surfaces allows us to use a standard SSID distance-measure which is simple to compute and can be easily extended by an elastic penalizer to account for nonlinear deformations. Due to the plain structure of the associated cost functional, the derivation of necessary optimality conditions by means of variational calculus is straightforward. The use of variational derivatives further enables us to employ fast and theoretically well-founded optimization routines such as Newton's method. Furthermore, the driving force of the registration can be quickly evaluated and is easy to interpret.



(a)



(b)

Figure 6.5: The artificial Purkinje fiber network in the TBunnyC model (a) and the registered network in the Oxford heart (b).

A. Appendix

Below, we use the notation $\boldsymbol{A} : \boldsymbol{B} = \sum_{ij} A_{ij}B_{ij}$ for matrices $\boldsymbol{A} = \{A_{ij}\}$ and $\boldsymbol{B} = \{B_{ij}\}$, Also, $|\cdot| = \|\cdot\|_{\ell_2}$ denotes the ℓ_2 vector norm, and the same notation is used below for the ℓ_2 matrix norm. For the sake of simplicity we set here $\Omega := (0,1)^d \subset \mathbb{R}^d$ and define $\Gamma_0, \Gamma_1 \subset \Omega$ to be closed sets with finite Hausdorf measure $\mathcal{H}^{d-1}(\Gamma_i) < \infty$, $i = 1, 2$. As a result, $|\Gamma_i| = \mathcal{H}^d(\Gamma_i) = 0$ [4].

It will be shown that J^ε is bounded but not coercive on $H^1(\Omega)$. Instead, J^ε is coercive on the closed linear subspace

$$\mathcal{H} = \left\{ \boldsymbol{w} \in H^1(\Omega) : \int_\Omega \boldsymbol{w}(\boldsymbol{x})d\boldsymbol{x} = 0, \int_\Omega \left[\boldsymbol{w}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}\boldsymbol{w}(\boldsymbol{x})^{\mathrm{T}}\right] d\boldsymbol{x} = 0 \right\}, \tag{A.1}$$

which is the orthogonal complement in $H^1(\Omega)$ of the subspace of infinitesimal rigid motions

$$\mathrm{RM} = \left\{ \boldsymbol{w} = \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x} : \boldsymbol{c} \in \mathbb{R}^d, \boldsymbol{W} \in \mathbb{S}^d \right\},$$

where

$$\mathbb{S}^d = \left\{ \boldsymbol{W} \in \mathbb{R}^{d\times d} : \boldsymbol{W} + \boldsymbol{W}^{\mathrm{T}} = 0 \right\}.$$

This direct sum decomposition of $H^1(\Omega)$ is established as follows.

Lemma 1. *$H^1(\Omega) = \mathrm{RM} \oplus \mathcal{H}$.*

Proof. To show that $\mathrm{RM} \perp \mathcal{H}$ holds in $H^1(\Omega)$, let $\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x} = \boldsymbol{u} \in \mathrm{RM}$ and $\boldsymbol{v} \in \mathcal{H}$ be arbitrary. Then

$$\begin{aligned}
\int_\Omega \boldsymbol{u}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{v}(\boldsymbol{x})d\boldsymbol{x} &= \boldsymbol{c}^{\mathrm{T}} \int_\Omega \boldsymbol{v}(\boldsymbol{x})d\boldsymbol{x} + \int_\Omega \boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{v}(\boldsymbol{x})d\boldsymbol{x} = \boldsymbol{W}^{\mathrm{T}} : \int_\Omega \boldsymbol{v}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} \\
&= \tfrac{1}{2}(\boldsymbol{W}^{\mathrm{T}} - \boldsymbol{W}) : \int_\Omega \boldsymbol{v}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} \\
&= \tfrac{1}{2}\boldsymbol{W} : \int_\Omega \left[\boldsymbol{v}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}\boldsymbol{v}(\boldsymbol{x})^{\mathrm{T}}\right] d\boldsymbol{x} = 0.
\end{aligned}$$

Also,

$$\begin{aligned}
\int_\Omega \nabla\boldsymbol{u}(\boldsymbol{x}) : \nabla\boldsymbol{v}(\boldsymbol{x})d\boldsymbol{x} &= \boldsymbol{W} : \int_\Omega \nabla\boldsymbol{v}(\boldsymbol{x})d\boldsymbol{x} \\
&= \tfrac{1}{2}(\boldsymbol{W} - \boldsymbol{W}^{\mathrm{T}}) : \int_\Omega \nabla\boldsymbol{v}(\boldsymbol{x})d\boldsymbol{x} \\
&= \tfrac{1}{2}\boldsymbol{W} : \int_\Omega \left[\nabla\boldsymbol{v}(\boldsymbol{x}) - \nabla\boldsymbol{v}(\boldsymbol{x})^{\mathrm{T}}\right] d\boldsymbol{x} = 0.
\end{aligned}$$

Thus, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{H^1(\Omega)} = 0$ holds.

Now let $\boldsymbol{w} \in H^1(\Omega)$ be arbitrary. It will be shown that there exists a unique $\boldsymbol{u} \in \mathrm{RM}$ and a unique $\boldsymbol{v} \in \mathcal{H}$ such that $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$. The $(d+1) \times d$ unknowns of $\boldsymbol{c}$ and $\boldsymbol{W}$ may be defined

uniquely after it is shown that the following system is uniquely solvable:

$$
\begin{aligned}
\int_\Omega \boldsymbol{w}(\boldsymbol{x})d\boldsymbol{x} &= \int_\Omega [\boldsymbol{c}+\boldsymbol{W}\boldsymbol{x}]d\boldsymbol{x} = \boldsymbol{c}\int_\Omega d\boldsymbol{x} + \boldsymbol{W}\int_\Omega \boldsymbol{x}d\boldsymbol{x},\\
\int_\Omega \left[\boldsymbol{w}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}\boldsymbol{w}(\boldsymbol{x})^{\mathrm{T}}\right] d\boldsymbol{x} &= \int_\Omega \left[(\boldsymbol{c}+\boldsymbol{W}\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}(\boldsymbol{c}+\boldsymbol{W}\boldsymbol{x})^{\mathrm{T}}\right] d\boldsymbol{x}\\
&= \boldsymbol{c}\left[\int_\Omega \boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}\right] - \left[\int_\Omega \boldsymbol{x}d\boldsymbol{x}\right]\boldsymbol{c}^{\mathrm{T}}\\
&+ \boldsymbol{W}\left[\int_\Omega \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}\right] - \left[\int_\Omega \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}\right]\boldsymbol{W}^{\mathrm{T}},\\
0 &= \boldsymbol{W}+\boldsymbol{W}^{\mathrm{T}}.
\end{aligned}
\tag{A.2}
$$

Given $\boldsymbol{c}$ and $\boldsymbol{W}$, define $\boldsymbol{u} = \boldsymbol{c}+\boldsymbol{W}\boldsymbol{x}$, which is clearly in RM. Then define $\boldsymbol{v} = \boldsymbol{w}-\boldsymbol{u}$, which is clearly in $\mathcal{H}$ because of the first two equations in (A.2). Should there exist $\tilde{\boldsymbol{u}} \in$ RM and $\tilde{\boldsymbol{v}} \in \mathcal{H}$ such that $\boldsymbol{w} = \tilde{\boldsymbol{u}}+\tilde{\boldsymbol{v}}$ holds, it follows from $\boldsymbol{u}-\tilde{\boldsymbol{u}} = \boldsymbol{v}-\tilde{\boldsymbol{v}}$ and $0 = \langle \boldsymbol{u}-\tilde{\boldsymbol{u}}, \boldsymbol{v}-\tilde{\boldsymbol{v}}\rangle_{H^1(\Omega)} = \langle \boldsymbol{v}-\tilde{\boldsymbol{v}}, \boldsymbol{v}-\tilde{\boldsymbol{v}}\rangle_{H^1(\Omega)}$ that $\boldsymbol{u}-\tilde{\boldsymbol{u}} = \boldsymbol{v}-\tilde{\boldsymbol{v}} = 0$. It remains to show that (A.2) can be solved.

The equation (A.3) for $\boldsymbol{W}$ below is obtained after multiplying the first equation in (A.2) by $-\int_\Omega \boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}$ from the right, the transposed first equation by $\int_\Omega \boldsymbol{x}d\boldsymbol{x}$ from the left, finally the second equation by $|\Omega|$ and then summing the results:

$$
\begin{aligned}
&|\Omega|\int_\Omega \left[\boldsymbol{w}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}} - \boldsymbol{x}\boldsymbol{w}(\boldsymbol{x})^{\mathrm{T}}\right] d\boldsymbol{x} - \int_\Omega \boldsymbol{w}(\boldsymbol{x})d\boldsymbol{x}\int_\Omega \boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} + \int_\Omega \boldsymbol{x}d\boldsymbol{x}\int_\Omega \boldsymbol{w}(\boldsymbol{x})^{\mathrm{T}}d\boldsymbol{x} =\\
&\boldsymbol{W}\left[|\Omega|\int_\Omega \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} - \int_\Omega \boldsymbol{x}d\boldsymbol{x}\int_\Omega \boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}\right] - \left[|\Omega|\int_\Omega \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} - \int_\Omega \boldsymbol{x}d\boldsymbol{x}\int_\Omega \boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}\right]\boldsymbol{W}^{\mathrm{T}}.
\end{aligned}
\tag{A.3}
$$

After replacing $-\boldsymbol{W}^{\mathrm{T}}$ with $\boldsymbol{W}$, according to the last equation in (A.2), (A.3) can be rewritten as the algebraic Riccati–equation $\boldsymbol{B}-\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{W}\boldsymbol{A}+\boldsymbol{A}\boldsymbol{W}$ in terms of the symmetric matrix $\boldsymbol{A}$

$$
\boldsymbol{A} = \left[|\Omega|\int_\Omega \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} - \int_\Omega \boldsymbol{x}d\boldsymbol{x}\int_\Omega \boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x}\right],
$$

and the skew symmetric matrix $\boldsymbol{B}$

$$
\boldsymbol{B} = |\Omega|\int_\Omega \boldsymbol{w}(\boldsymbol{x})\boldsymbol{x}^{\mathrm{T}}d\boldsymbol{x} + \int_\Omega \boldsymbol{x}d\boldsymbol{x}\int_\Omega \boldsymbol{w}(\boldsymbol{x})^{\mathrm{T}}d\boldsymbol{x}.
$$

For the solution of (A.3), the matrix $\boldsymbol{A}$ will be shown to be positive definite. For an arbitrary $\boldsymbol{b} \in \mathbb{R}^d$, note that the following holds for the linear function $f(\boldsymbol{x}) = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$:

$$
\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b} = |\Omega|^2\left[\frac{1}{|\Omega|}\int_\Omega f(\boldsymbol{x})^2 d\boldsymbol{x} - \left(\frac{1}{|\Omega|}\int_\Omega f(\boldsymbol{x})d\boldsymbol{x}\right)^2\right] \geq 0,
\tag{A.4}
$$

where the inequality in (A.4) follows from

$$\left(\int_\Omega f(\boldsymbol{x})\cdot 1 d\boldsymbol{x}\right)^2 \leq \|f\|^2_{L^2(\Omega)}\|1\|^2_{L^2(\Omega)} = |\Omega|\|f\|^2_{L^2(\Omega)}. \tag{A.5}$$

If equality were to hold in (A.4) and (A.5), this would require that f and 1 be linearly dependent. However, such dependence is impossible when f is linear. Hence, the diagonalization $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$ follows for an orthonormal matrix $\boldsymbol{Q}$ and a diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1,\ldots,\lambda_d\}$ containing the positive eigenvalues of $\boldsymbol{A}$. Then by defining $\{U_{ij}\} = \boldsymbol{U} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{Q}$, (A.3) can be rewritten as $\boldsymbol{U}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{U} = \boldsymbol{C}$ where $\{C_{ij}\} = \boldsymbol{C} = \boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{B}-\boldsymbol{B}^{\mathrm{T}})\boldsymbol{Q}$. The solution $\boldsymbol{U}$ is determined componentwise according to $U_{ij} = C_{ij}/(\lambda_i+\lambda_j)$, which is skew symmetric since $\boldsymbol{C}$ is. Finally, the solution to (A.3) is given by $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{U}\boldsymbol{Q}^{\mathrm{T}}$, which is skew symmetric since $\boldsymbol{U}$ is. Given $\boldsymbol{W}$, $\boldsymbol{c}$ is determined directly by the first equation in (A.2). □

Note that the whole of RM is in the kernel of the elastic penalty E in (2.6). Thus, J^ε is not coercive on $H^1(\Omega)$:

$$c\|\boldsymbol{u}\|^2_{H^1(\Omega)} \not\leq J^\varepsilon(\boldsymbol{u}) = S^\varepsilon(\boldsymbol{u}) \leq 2|\Omega|, \quad \forall \boldsymbol{u}\in \mathrm{RM}.$$

The cost J^ε is however bounded and coercive on $\mathcal{H}$, and it will be shown below that J^ε has a minimizer in $H^1(\Omega)$. For this existence, it will be useful to show that the energy norm

$$\begin{aligned}
\langle \boldsymbol{v}_1,\boldsymbol{v}_2\rangle_{\mathrm{E}} &= \int_\Omega \left[\tfrac{\lambda}{2}(\nabla\cdot\boldsymbol{v}_1(\boldsymbol{x}))(\nabla\cdot\boldsymbol{v}_1(\boldsymbol{x}))\right.\\
&+ \left.\tfrac{\mu}{4}(\nabla\boldsymbol{v}_1(\boldsymbol{x})+\nabla\boldsymbol{v}_1(\boldsymbol{x})^{\mathrm{T}}):(\nabla\boldsymbol{v}_1(\boldsymbol{x})+\nabla\boldsymbol{v}_1(\boldsymbol{x})^{\mathrm{T}})\right] d\boldsymbol{x},\\
\|\boldsymbol{v}\|^2_{\mathrm{E}} &= \langle \boldsymbol{v}_1,\boldsymbol{v}_2\rangle_{\mathrm{E}},
\end{aligned}$$

is equivalent to the $H^1(\Omega)$ norm on $\mathcal{H}$.

Lemma 2. *The $H^1(\Omega)$ norm is equivalent to the energy norm on $\mathcal{H}$.*

Proof. We seek constants $a_1, a_2 > 0$ such that

$$a_1\|\boldsymbol{v}\|_E \leq \|\boldsymbol{v}\|_{H^1(\Omega)} \leq a_2\|\boldsymbol{v}\|_E, \quad \forall \boldsymbol{v}\in\mathcal{H}. \tag{A.6}$$

The existence of the constant a_1 follows readily. To show the existence of the constant a_2, assume for the sake of contradiction that there exists a sequence $\{\boldsymbol{v}_n\}\subset\mathcal{H}$ satisfying:

$$\|\boldsymbol{v}_n\|_{H^1(\Omega)} = 1, \qquad \text{while} \qquad \|\boldsymbol{v}_n\|_E \to 0.$$

Since $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$ [1], there is a subsequence $\{\boldsymbol{v}_{n_l}\}$ which converges in $L^2(\Omega)$. From Korn's Inequality [7],

$$\begin{aligned}
\|\nabla\boldsymbol{w}\|^2_{L^2(\Omega)} &\leq k_1\|\boldsymbol{w}\|^2_{L^2(\Omega)} + k_2\|\nabla\boldsymbol{w}^{\mathrm{T}}+\nabla\boldsymbol{w}\|^2_{L^2(\Omega)} \qquad \forall \boldsymbol{w}\in H^1(\Omega),\\
&\leq k_1\|\boldsymbol{w}\|^2_{L^2(\Omega)} + k_2\tfrac{4}{\mu}\|\boldsymbol{w}\|^2_E,
\end{aligned}$$

it follows that

$$\|\boldsymbol{v}_{n_l} - \boldsymbol{v}_{n_k}\|_{H^1(\Omega)} \leq c_1\|\boldsymbol{v}_{n_l} - \boldsymbol{v}_{n_k}\|_{L^2(\Omega)} + c_2\|\boldsymbol{v}_{n_l} - \boldsymbol{v}_{n_k}\|_E.$$

Since both terms on the right side vanish, it follows that $\{\boldsymbol{v}_{n_l}\}$ is a Cauchy sequence in $H^1(\Omega)$ with some limit $\boldsymbol{v}^* \in H^1(\Omega)$ which satisfies:

$$\|\boldsymbol{v}^*\|_{H^1(\Omega)} = \lim_{n_l \to 0} \|\boldsymbol{v}_{n_l}\|_{H^1(\Omega)} = 1, \quad \|\boldsymbol{v}^*\|_E = \lim_{n_l \to 0} \|\boldsymbol{v}_{n_l}\|_E = 0. \tag{A.7}$$

Since $\mathcal{H}$ is a closed linear subspace of $H^1(\Omega)$, it follows that $\boldsymbol{v}^* \in \mathcal{H}$. From (A.7) and thus $\|\nabla \boldsymbol{v}^* + \nabla \boldsymbol{v}^{*\mathrm{T}}\|_{L^2(\Omega)} = 0$, it follows that $\boldsymbol{v}^* \in \mathrm{RM}$ [7]. According to Lemma 1, $\boldsymbol{v}^* \in \mathcal{H} \cap \mathrm{RM} = \{0\}$. However, $\boldsymbol{v}^*$ cannot vanish and satisfy (A.7) simultaneously. With this contradiction (A.6) is obtained. □

Lemma 3. *Let $\varepsilon > 0$ be fixed. If $\mathcal{I}_0^\varepsilon$ is Lipschitz continuous on $\mathbb{R}^d$, then S^ε is continuous on $L^2(\Omega)$.*

Proof. Define the Lipschitz constant L for $\mathcal{I}_0^\varepsilon$:

$$|\mathcal{I}_0^\varepsilon(\boldsymbol{x}) - \mathcal{I}_0^\varepsilon(\boldsymbol{y})| \leq L|\boldsymbol{x} - \boldsymbol{y}|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d.$$

For $\boldsymbol{w}_1, \boldsymbol{w}_2 \in L^2(\Omega)$,

$$\begin{aligned}
|S^\varepsilon(\boldsymbol{w}_1) - S^\varepsilon(\boldsymbol{w}_2)| &\leq \int_\Omega \left||\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}_1(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 - |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}_2(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2\right| d\boldsymbol{x} \\
&\leq \int_\Omega |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}_1(\boldsymbol{x})) - \mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}_2(\boldsymbol{x}))| \times \\
&\qquad [\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}_1(\boldsymbol{x})) + \mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}_2(\boldsymbol{x})) - 2\mathcal{I}_1^\varepsilon(\boldsymbol{x})] d\boldsymbol{x} \\
&\leq 2\int_\Omega L|\boldsymbol{w}_1(\boldsymbol{x}) - \boldsymbol{w}_2(\boldsymbol{x})| d\boldsymbol{x} \quad \leq \quad 2L\sqrt{|\Omega|}\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{L^2(\Omega)}.
\end{aligned} \tag{A.8}$$

So $|S^\varepsilon(\boldsymbol{w}_1) - S^\varepsilon(\boldsymbol{w}_2)| \to 0$ as $\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{L^2(\Omega)} \to 0$. □

Lemma 4. *Let $\varepsilon > 0$ be fixed. Assume that $\mathcal{I}_0^\varepsilon$ is Lipschitz continuous on $\mathbb{R}^d$. Suppose further that for a given $\boldsymbol{w} \in L^2(\Omega)$, there exists a $\tilde{\boldsymbol{c}} \in \mathbb{R}^d$ and a $\tilde{\boldsymbol{W}} \in \mathbb{S}^d$ such that*

$$\int_\Omega |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \tilde{\boldsymbol{c}} + \tilde{\boldsymbol{W}}\boldsymbol{x}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} < \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x}, \tag{A.9}$$

Then there exists a $\boldsymbol{c}^* \in \mathbb{R}^d$ and a $\boldsymbol{W}^* \in \mathbb{S}^d$ such that

$$S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c}^* + \boldsymbol{W}^*\boldsymbol{x}) = \inf_{\boldsymbol{c} \in \mathbb{R}^d, \boldsymbol{W} \in \mathbb{S}^d} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}), \tag{A.10}$$

Proof. Before it is proved that the infimum is achieved on a bounded set, certain estimates of the function $(\boldsymbol{c}, \boldsymbol{W}) \mapsto S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}^*\boldsymbol{x})$ will be derived. First note that for arbitrary $t > 0$,

$\boldsymbol{c} \in \mathbb{R}^d$ and $\boldsymbol{W} \in \mathbb{S}^d$

$$
\begin{aligned}
& S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \\
& = \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} \\
& + \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|\geq t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} \qquad \text{(A.11)} \\
& \geq \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} - |\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| \geq t\}| \\
& \geq \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} - \frac{1}{t^2}\int_\Omega |\boldsymbol{w}(\boldsymbol{x})|^2 d\boldsymbol{x},
\end{aligned}
$$

where the estimate of the measure of the set $\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| \geq t\}$ follows according to the Chebychev Inequality [3]. Now, for $|\boldsymbol{c}| \geq 2 + |\boldsymbol{W}| + t$ and $\boldsymbol{x} \in \{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t\}$, it follows from

$$
\begin{aligned}
|\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}| &\geq |\boldsymbol{c}| - |\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}| \\
&\geq |\boldsymbol{c}| - (|\boldsymbol{x}| + |\boldsymbol{w}(\boldsymbol{x})| + |\boldsymbol{W}||\boldsymbol{x}|) \\
&\geq |\boldsymbol{c}| - (1 + |\boldsymbol{W}| + t) \geq 1,
\end{aligned}
$$

that $\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x} \notin \Omega = (0,1)^d$ and therefore $\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) = 0$. Hence, for every $t > 0$,

$$
|\boldsymbol{c}| \geq 2 + |\boldsymbol{W}| + t \quad \Rightarrow \quad S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \geq \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{1}{t^2}\|\boldsymbol{w}\|^2_{L^2(\Omega)}.
$$

Again, according to the Chebychev Inequality,

$$
\begin{aligned}
\left| \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \right| &= \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|\geq t\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \\
&\leq |\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| \geq t\}| \leq \frac{1}{t^2}\int_\Omega |\boldsymbol{w}(\boldsymbol{x})|^2 d\boldsymbol{x}.
\end{aligned} \qquad \text{(A.12)}
$$

Combining the last two inequalities shows that for every $t > 0$,

$$
|\boldsymbol{c}| \geq 2 + |\boldsymbol{W}| + t \quad \Rightarrow \quad S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \geq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{2}{t^2}\|\boldsymbol{w}\|^2_{L^2(\Omega)}. \qquad \text{(A.13)}
$$

Now consider the case that $|\boldsymbol{c}| > 2 + |\boldsymbol{W}| + t$ may not hold for large $|\boldsymbol{c}| + |\boldsymbol{W}|$. Since $\boldsymbol{W} \in \mathbb{S}^d$, $\boldsymbol{W}$ is normal and therefore possesses an orthonormal set of eigenvectors $\{\boldsymbol{n}_i\}_{i=1}^d$. Suppose the eigenvalue magnitude $M = |\boldsymbol{W}|$ corresponds to $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ in the sense that the following hold:

$$
\boldsymbol{W}\boldsymbol{n}_1 = -M\boldsymbol{n}_2, \quad \boldsymbol{W}\boldsymbol{n}_2 = M\boldsymbol{n}_1.
$$

Define $\boldsymbol{P}$ as the projection onto $\text{span}\{\boldsymbol{n}_1, \boldsymbol{n}_2\}$ and define the seminorm $p(\boldsymbol{x}) = |\boldsymbol{P}\boldsymbol{x}|$. Then for

$\boldsymbol{x} \in \{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t\}$ with $p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) > (2 + t)$, it follows from

$$\begin{aligned} |\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}| &\geq |\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}| - |\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})| \\ &\geq [p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x})^2 + \cdots]^{\frac{1}{2}} - (|\boldsymbol{x}| + |\boldsymbol{w}(\boldsymbol{x})|) \\ &\geq p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) - (1 + t) > 1, \end{aligned}$$

that $\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x} \notin \Omega = (0, 1)^d$ and therefore $\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{W}\boldsymbol{x}) = 0$. So (A.11) can be written as:

$$\begin{aligned} & S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \\ & \geq \int_{\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \leq (2+t)\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{W}\boldsymbol{x}) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} \\ & + \int_{\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) > (2+t)\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{1}{t^2} \|\boldsymbol{w}\|_{L^2(\Omega)}^2. \end{aligned} \tag{A.14}$$

The above integration set, $\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \leq (2 + t)\}$, is estimated as follows using $\boldsymbol{c} = \sum_{i=1}^d \chi_i \boldsymbol{n}_i$ and $\boldsymbol{x} = \sum_{i=1}^d \xi_i \boldsymbol{n}_i$ for arbitrary $\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^d$. First note that

$$\Omega = (0, 1)^d \subset \left\{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}|^2 \leq d\right\} = \left\{\sum_{i=1}^d \xi_i \boldsymbol{n}_i : \sum_{i=1}^d \xi_i^2 \leq d\right\} \subset \left\{\sum_{i=1}^d \xi_i \boldsymbol{n}_i : |\xi_i| \leq \sqrt{d}\right\}.$$

Also note that:

$$\left\{\boldsymbol{x} \in \mathbb{R}^d : p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \leq (2 + t)\right\} = \left\{\sum_{i=1}^d \xi_i \boldsymbol{n}_i : (\chi_1 - M\xi_2)^2 + (\chi_2 + M\xi_1)^2 \leq (2 + t)^2\right\}.$$

Thus, the integration set, $\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \leq (2 + t)\}$, is a subset of the cylinder:

$$\left\{\sum_{i=1}^d \xi_i \boldsymbol{n}_i : (\chi_1/M - \xi_2)^2 + (\chi_2/M + \xi_1)^2 \leq [(2 + t)/M]^2; \quad |\xi_i| \leq \sqrt{d}, i > 2\right\},$$

which has the measure $\pi[(2 + t)/M]^2[2\sqrt{d}]^{d-2}$. So (A.14) can be written as:

$$\begin{aligned} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) &\geq \int_{\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) > (2+t)\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{1}{t^2} \|\boldsymbol{w}\|_{L^2(\Omega)}^2 \\ &- |\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \leq (2 + t)\}| \\ &\geq \int_{\{\boldsymbol{x} \in \Omega : |\boldsymbol{w}(\boldsymbol{x})| < t, p(\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) > (2+t)\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \\ &- \pi[2\sqrt{d}]^{d-2} \left[\frac{2 + t}{M}\right]^2 - \frac{1}{t^2} \|\boldsymbol{w}\|_{L^2(\Omega)}^2. \end{aligned} \tag{A.15}$$

Again, according to the Chebychev Inequality,

$$\left| \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t, p(\boldsymbol{c}+\boldsymbol{W}\boldsymbol{x})>(2+t)\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \right|$$

(A.16)

$$\leq \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{w}(\boldsymbol{x})|<t, p(\boldsymbol{c}+\boldsymbol{W}\boldsymbol{x})\leq(2+t)\}} |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \leq \pi[2\sqrt{d}]^{d-2} \left[\frac{2+t}{M}\right]^2.$$

So combining (A.12), (A.15) and (A.16) gives:

$$S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) \geq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - 2\pi[2\sqrt{d}]^{d-2} \left[\frac{2+t}{M}\right]^2 - \frac{2}{t^2}\|\boldsymbol{w}\|^2_{L^2(\Omega)}. \tag{A.17}$$

Now let $\{(\boldsymbol{c}_n, \boldsymbol{W}_n)\}$ be a minimizing sequence for $S^\varepsilon(\boldsymbol{w}(\boldsymbol{x})+\boldsymbol{c}+\boldsymbol{W}\boldsymbol{x})$. Suppose for the sake of contradiction that $\{|\boldsymbol{c}_n|+|\boldsymbol{W}_n|\}$ is not bounded. If $\{|\boldsymbol{W}_n|\}$ is bounded, then $|\boldsymbol{c}_n| \geq 2+|\boldsymbol{W}_n|+t$ holds for n large enough and for every fixed $t > 0$. So according to (A.13), the following holds for every $t > 0$:

$$\lim_{n\to\infty} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c}_n + \boldsymbol{W}_n\boldsymbol{x}) \geq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{2}{t^2}\|\boldsymbol{w}\|^2_{L^2(\Omega)},$$

and hence:

$$\inf_{\boldsymbol{c}\in\mathbb{R}^d, \boldsymbol{W}\in\mathbb{S}^d} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}) = \lim_{n\to\infty} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c}_n + \boldsymbol{W}_n\boldsymbol{x}) \geq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x}. \tag{A.18}$$

However, this inequality violates (A.9). The contradiction implies that $\{|\boldsymbol{c}_n|\}$ may not be unbounded while $\{|\boldsymbol{W}_n|\}$ is bounded. Now suppose that $\{M_n = |\boldsymbol{W}_n|\}$ is unbounded. Then whether $\{|\boldsymbol{c}_n|\}$ is bounded or not, (A.17) implies for every $t > 0$ that:

$$\begin{aligned}
\lim_{n\to\infty} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c}_n + \boldsymbol{W}_n\boldsymbol{x}) &\geq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{2}{t^2}\|\boldsymbol{w}\|^2_{L^2(\Omega)} \\
&- \lim_{n\to\infty} 2\pi[2\sqrt{d}]^{d-2} \left[\frac{2+t}{M_n}\right]^2 \\
&= \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} - \frac{2}{t^2}\|\boldsymbol{w}\|^2_{L^2(\Omega)},
\end{aligned}$$

and hence (A.18) is obtained again. The contradiction implies that $\{|\boldsymbol{c}_n|\}$ and $\{|\boldsymbol{W}_n|\}$ are bounded. Thus, there are subsequences $\{\boldsymbol{c}_{n_l}\}$ and $\{\boldsymbol{W}_{n_l}\}$ which converge to some $\boldsymbol{c}^* \in \mathbb{R}^d$ and $\boldsymbol{W}^* \in \mathbb{S}^d$ respectively. According to Lemma 3, S^ε is continuous on $L^2(\Omega)$, and hence

$$S^\varepsilon(\boldsymbol{w} + \boldsymbol{c}^* + \boldsymbol{W}^*\boldsymbol{x}) = \lim_{n_l\to\infty} S^\varepsilon(\boldsymbol{w} + \boldsymbol{c}_{n_l} + \boldsymbol{W}_{n_l}\boldsymbol{x}) = \inf_{\boldsymbol{c}\in\mathbb{R}^d, \boldsymbol{W}\in\mathbb{S}^d} S^\varepsilon(\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}),$$

and the desired result (A.10) is obtained. □

Lemma 5. *Let $\varepsilon > 0$ be fixed. Assume that $\mathcal{I}_0^\varepsilon$ is Lipschitz continuous on $\mathbb{R}^d$. Then for every*

$\boldsymbol{u} \in \mathrm{RM}$ *there exists a* $\boldsymbol{v}^* \in \mathcal{H}$ *such that*

$$J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}^*) = \inf_{\boldsymbol{v}\in\mathcal{H}} J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}). \tag{A.19}$$

Proof. Let $\{\boldsymbol{v}_n\} \subset \mathcal{H}$ be a minimizing sequence. In particular, assume that

$$J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_1) \geq J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_2) \geq \cdots J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n) \overset{n\to\infty}{\longrightarrow} \inf_{\boldsymbol{v}\in\mathcal{H}} J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}).$$

Since the subspace RM is in the kernel of the penalty E, it follows that:

$$J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n) = S^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n) + E(\boldsymbol{v}_n)$$

From $E(\boldsymbol{v}_n) \leq J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n) \leq J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_1)$, it follows with Lemma 2 that the sequence $\{\boldsymbol{v}_n\}$ is bounded in $H^1(\Omega)$. Thus, there is a subsequence, again denoted by $\{\boldsymbol{v}_n\}$ for convenience, which converges weakly in $H^1(\Omega)$ and strongly in $L^2(\Omega)$ to some $\boldsymbol{v}^* \in H^1(\Omega)$ [14]. By the Hahn-Banach Theorem, $\boldsymbol{v}^*$ belongs to the closure in $H^1(\Omega)$ of $\mathrm{span}\{\boldsymbol{v}_n\}$, and hence $\boldsymbol{v}^* \in \mathcal{H}$. According to Lemma 3, the similarity measure is continous on $H^1(\Omega) \subset L^2(\Omega)$

$$S^\varepsilon(\boldsymbol{u}+\boldsymbol{v}^*) = \lim_{n\to\infty} S^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n),$$

and by the lower semicontinuity of the energy norm

$$E(\boldsymbol{v}^*) \leq \liminf_{n\to\infty} E(\boldsymbol{v}_n), \tag{A.20}$$

we get

$$\begin{aligned} J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}^*) = S^\varepsilon(\boldsymbol{u}+\boldsymbol{v}^*) + E(\boldsymbol{v}^*) &\leq \liminf_{n\to\infty} S^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n) + E(\boldsymbol{v}_n) \\ &= \lim_{n\to\infty} J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}_n) = \inf_{\boldsymbol{v}\in\mathcal{H}} J^\varepsilon(\boldsymbol{u}+\boldsymbol{v}), \end{aligned}$$

and the desired result (A.19) is obtained. □

We compute a registration by solving the minimization problem

$$\inf_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}). \tag{A.21}$$

Since the images $\mathcal{I}_i^\varepsilon$ vary with the value of ε, (A.21) forms a stand-alone minimization problem for each ε. Thus we first prove that (A.21) admits a solution $\boldsymbol{w} \in H^1(\Omega)$ for a fixed $\varepsilon > 0$.

Theorem 3. *Let* $\varepsilon > 0$ *be fixed,* $\mathcal{I}_0^\varepsilon \in W^{1,\infty}(\Omega)$, $\mathcal{I}_1^\varepsilon \in L^\infty(\Omega)$ *and* $\lambda, \mu > 0$. *Then there exists a* $\boldsymbol{w}^\star \in H^1(\Omega)$ *such that*

$$J^\varepsilon(\boldsymbol{w}^\star) = \min_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}). \tag{A.22}$$

Proof. Assume that $\{\tilde{\boldsymbol{w}}_n\} \subset H^1(\Omega)$ is a minimizing sequence. In particular, assume that

$$J^\varepsilon(\tilde{\boldsymbol{w}}_1) \geq J^\varepsilon(\tilde{\boldsymbol{w}}_2) \geq \cdots J^\varepsilon(\tilde{\boldsymbol{w}}_n) \overset{n\to\infty}{\longrightarrow} \inf_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}).$$

For the case that $\{\tilde{\boldsymbol{w}}_n\}$ is not bounded in $H^1(\Omega)$, a related minimizing sequence will be de-

rived which can be shown to be bounded. For this, Lemma 1 is used to define the orthogonal decomposition:

$$\tilde{\boldsymbol{w}}_n = \boldsymbol{u}_n + \tilde{\boldsymbol{v}}_n, \quad \boldsymbol{u}_n \in \mathrm{RM}, \quad \tilde{\boldsymbol{v}}_n \in \mathcal{H}.$$

By Lemma 5, there exists for each n a $\boldsymbol{v}_n \in \mathcal{H}$ such that

$$J^\varepsilon(\boldsymbol{u}_n + \boldsymbol{v}_n) = \inf_{\boldsymbol{v}\in\mathcal{H}} J^\varepsilon(\boldsymbol{u}_n + \boldsymbol{v}) \leq J^\varepsilon(\tilde{\boldsymbol{w}}_n).$$

Now define $\boldsymbol{w}_n = \boldsymbol{u}_n + \boldsymbol{v}_n$ and note that according to

$$J^\varepsilon(\boldsymbol{w}_n) = J^\varepsilon(\boldsymbol{u}_n + \boldsymbol{v}_n) \leq J^\varepsilon(\tilde{\boldsymbol{w}}_n) \overset{n\to\infty}{\longrightarrow} \inf_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}),$$

the sequence $\{\boldsymbol{w}_n\}$ is minimizing. From

$$E(\boldsymbol{v}_n) \leq J^\varepsilon(\boldsymbol{w}_n) \leq J^\varepsilon(\tilde{\boldsymbol{w}}_n) \overset{n\to\infty}{\longrightarrow} \inf_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}),$$

it follows with Lemma 2 that the sequence $\{\boldsymbol{v}_n\}$ is bounded in $H^1(\Omega)$. Thus, there is a subsequence, again denoted by $\{\boldsymbol{v}_n\}$ for convenience, which converges weakly in $H^1(\Omega)$ and strongly in $L^2(\Omega)$ to some $\boldsymbol{v}^* \in H^1(\Omega)$ [14]. By the Hahn-Banach Theorem, $\boldsymbol{v}^*$ belongs to the closure in $H^1(\Omega)$ of span$\{\boldsymbol{v}_n\}$, and hence $\boldsymbol{v}^* \in \mathcal{H}$. Now suppose that:

$$\int_\Omega |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{v}^*(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} \geq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x}, \quad \forall \boldsymbol{u} \in \mathrm{RM}. \tag{A.23}$$

Then using the lower semicontinuity of the energy norm, choose a subsequence $\{\boldsymbol{v}_{n_k}\} \subset \{\boldsymbol{v}_n\}$ satisfying:

$$E(\boldsymbol{v}^*) \leq \liminf_{n\to\infty} E(\boldsymbol{v}_n) = \lim_{k\to\infty} E(\boldsymbol{v}_{n_k}) \tag{A.24}$$

Also with (A.8) note that

$$\lim_{k\to\infty} |S^\varepsilon(\boldsymbol{v}^* + \boldsymbol{u}_{n_k}) - S^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k})| \leq 2L\sqrt{\Omega} \lim_{k\to\infty} \|\boldsymbol{v}^* - \boldsymbol{v}_{n_k}\|_{L^2(\Omega)} = 0.$$

Combining these results gives:

$$\int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \leq \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} + E(\boldsymbol{v}^*)$$

$$\leq \liminf_{k\to\infty} \int_\Omega |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{v}^*(\boldsymbol{x}) + \boldsymbol{u}_{n_k}(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} + E(\boldsymbol{v}^*)$$

$$= \liminf_{k\to\infty} \{J^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k}) + [S^\varepsilon(\boldsymbol{v}^* + \boldsymbol{u}_{n_k}) - S^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k})] + [E(\boldsymbol{v}^*) - E(\boldsymbol{v}_{n_k})]\}$$

$$= \lim_{k\to\infty} J^\varepsilon(\boldsymbol{w}_{n_k}) + \lim_{k\to\infty} [S^\varepsilon(\boldsymbol{v}^* + \boldsymbol{u}_{n_k}) - S^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k})] + [E(\boldsymbol{v}^*) - \lim_{k\to\infty} E(\boldsymbol{v}_{n_k})]$$

$$= \inf_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}) + [E(\boldsymbol{v}^*) - \lim_{k\to\infty} E(\boldsymbol{v}_{n_k})] \leq \inf_{\boldsymbol{w}\in H^1(\Omega)} J^\varepsilon(\boldsymbol{w})$$

Thus, if (A.23) holds, one can as well take a minimizer to be $\boldsymbol{w}^* = \boldsymbol{c}$ for $\boldsymbol{c}$ large enough that $(\Omega + \boldsymbol{c}) \cap \Omega = \emptyset$. Hence it may be assumed that

$$\int_\Omega |\mathcal{I}_0^\varepsilon(\boldsymbol{x} + \boldsymbol{v}^*(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x}))|^2 d\boldsymbol{x} < \int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x}, \quad \forall \boldsymbol{u} \in \mathrm{RM},$$

and due to the fact that $\mathcal{I}_0^\varepsilon$ as $W^{1,\infty}(\Omega)$-function is Lipschitz continuous [17] we may apply Lemma 4. There exists a $u^* \in \mathrm{RM}$ such that

$$S^\varepsilon(\boldsymbol{u}^* + \boldsymbol{v}^*) = \min_{\boldsymbol{u} \in \mathrm{RM}} S^\varepsilon(\boldsymbol{u} + \boldsymbol{v}^*). \tag{A.25}$$

Now define $\boldsymbol{w}^* = \boldsymbol{u}^* + \boldsymbol{v}^*$, which will be shown to minimize J^ε. Using (A.25) and the subseqence $\{n_k\}$ of (A.24),

$$\begin{aligned}
S^\varepsilon(\boldsymbol{u}^* + \boldsymbol{v}^*) &\leq \liminf_{k\to\infty} S^\varepsilon(\boldsymbol{u}_{n_k} + \boldsymbol{v}^*) \\
&= \liminf_{k\to\infty} \{S^\varepsilon(\boldsymbol{u}_{n_k} + \boldsymbol{v}^*) - S^\varepsilon(\boldsymbol{u}_{n_k} + \boldsymbol{v}_{n_k}) + S^\varepsilon(\boldsymbol{u}_{n_k} + \boldsymbol{v}_{n_k})\} \\
&\leq 2L\sqrt{\Omega} \lim_{k\to\infty} \|\boldsymbol{v}^* - \boldsymbol{v}_{n_k}\|_{L^2(\Omega)} + \liminf_{k\to\infty} S^\varepsilon(\boldsymbol{u}_{n_k} + \boldsymbol{v}_{n_k}) \\
&= \liminf_{k\to\infty} S^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k})
\end{aligned}$$

where (A.8) has been used. Thus,

$$\begin{aligned}
J^\varepsilon(\boldsymbol{w}^*) = S^\varepsilon(\boldsymbol{v}^\star + \boldsymbol{u}^\star) + E(\boldsymbol{v}^*) &\leq \liminf_{k\to\infty} S^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k}) + \lim_{k\to\infty} E(\boldsymbol{v}_{n_k}) \\
&= \liminf_{k\to\infty} S^\varepsilon(\boldsymbol{v}_{n_k} + \boldsymbol{u}_{n_k}) + E(\boldsymbol{v}_{n_k}) \\
&= \liminf_{k\to\infty} J^\varepsilon(\boldsymbol{w}_{n_k}) \\
&\leq \liminf_{k\to\infty} J^\varepsilon(\tilde{\boldsymbol{w}}_{n_k}) = \inf_{\boldsymbol{w} \in H^1(\Omega)} J^\varepsilon(\boldsymbol{w}),
\end{aligned}$$

and the desired result (A.22) is obtained. □

Now that we have guaranteed existence of a solution to (A.21) for a fixed $\varepsilon > 0$ we are able to investigate the behavior of solutions of (A.21) for various values of ε as $\varepsilon \to 0$. Now we show the convergence of the blurred images $\mathcal{I}_i^\varepsilon$ to I_i as $\varepsilon \to 0$.

Lemma 6. *As $\varepsilon \to 0$, $\mathcal{I}_0^\varepsilon$ and $\mathcal{I}_1^\varepsilon$ converge pointwise to I_0 and I_1, respectively.*

Proof. Let $\boldsymbol{x} \in \Omega$ be arbitrary. If $d_{\Gamma_0}(\boldsymbol{x}) = 0$ holds, then since Γ_0 is closed, $\boldsymbol{x} \in \Gamma_0$ and $I_0(\boldsymbol{x}) = 1$ hold. Also, $\forall \varepsilon > 0$, $0 = d_{\Gamma_0}(\boldsymbol{x}) \leq \varepsilon$ implies that $\mathcal{I}_0^\varepsilon(\boldsymbol{x}) = 1 - d_{\Gamma_0}(\boldsymbol{x}) = 1 = I_0(\boldsymbol{x})$. On the other hand, if $d_{\Gamma_0}(\boldsymbol{x}) > 0$ holds, then $\boldsymbol{x} \notin \Gamma_0$ and $I_0(\boldsymbol{x}) = 0$ hold. Also, according to (2.4), choosing $\varepsilon < d_{\Gamma_0}(\boldsymbol{x})$ gives $\mathcal{I}_0^\varepsilon(\boldsymbol{x}) = 0 = I_0(\boldsymbol{x})$. Thus, $\lim_{\varepsilon\to 0} \mathcal{I}_0^\varepsilon(\boldsymbol{x}) = I_0(\boldsymbol{x})$. In the same way, $\mathcal{I}_1^\varepsilon$ converge pointwise to I_1. □

Finally, we have the following convergence of minimizers as $\varepsilon \to 0$.

Theorem 4. *Assume the conditions of Theorem 3. For every $\varepsilon > 0$, let $\boldsymbol{w}_\varepsilon \in H^1(\Omega)$ denote the minimizer of $J^\varepsilon(\boldsymbol{w})$. Let $\boldsymbol{u}_\varepsilon$ and $\boldsymbol{v}_\varepsilon$ denote the projections of $\boldsymbol{w}_\varepsilon$ onto* RM *and $\mathcal{H}$, respectively. Then*

$$\lim_{\varepsilon\to 0} \boldsymbol{v}_\varepsilon = 0. \tag{A.26}$$

Also, there exists a $\boldsymbol{u}_0 \in \mathrm{RM}$ *such that*

$$\lim_{\varepsilon\to 0} S^\varepsilon(\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon) = \lim_{\varepsilon\to 0} S^\varepsilon(\boldsymbol{u}_\varepsilon+\boldsymbol{v}_\varepsilon) = 0. \tag{A.27}$$

Proof. For every $\varepsilon > 0$, it must be that

$$\int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \geq J^\varepsilon(\boldsymbol{w}_\varepsilon) = S^\varepsilon(\boldsymbol{u}_\varepsilon+\boldsymbol{v}_\varepsilon) + E(\boldsymbol{v}_\varepsilon). \tag{A.28}$$

Otherwise, $\boldsymbol{w}_\varepsilon$ could as well be chosen to be a constant $\boldsymbol{c}$ large enough that $(\Omega+\boldsymbol{c})\cap\Omega = \emptyset$. By Lemma 6, $\mathcal{I}_1^\varepsilon$ converges a.e. to zero. Since $|\mathcal{I}_1^\varepsilon| \leq 1$, it follows with the Dominated Convergence Theorem [17] that

$$\lim_{\varepsilon\to 0}\int_\Omega |\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} = 0. \tag{A.29}$$

Hence, by (A.28), $E(\boldsymbol{v}_\varepsilon) \to 0$, $\varepsilon \to 0$. It follows with Lemma 2 that

$$\lim_{\varepsilon\to 0} \|\boldsymbol{v}_\varepsilon\|_{H^1(\Omega)} = 0 \tag{A.30}$$

and (A.26) is established. Similarly, it follows from (A.28) that

$$\lim_{\varepsilon\to 0} S^\varepsilon(\boldsymbol{u}_\varepsilon+\boldsymbol{v}_\varepsilon) = 0. \tag{A.31}$$

It remains to construct $\boldsymbol{u}_0$ and to prove (A.27). For this, choose some $\varepsilon_0 > 0$ and let $\boldsymbol{c} = \boldsymbol{u}_0(\boldsymbol{x}) \in \mathrm{RM}$ be chosen large enough that

$$|\boldsymbol{c}| > 2 + \|\boldsymbol{v}_\varepsilon\|_{L^2(\Omega)}^{1/2}, \quad \forall \varepsilon \in (0,\varepsilon_0) \tag{A.32}$$

Now, for every $t > 0$,

(A.33)

$$\begin{aligned}
S^\varepsilon(\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon) &= \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|<t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{u}_0(\boldsymbol{x})+\boldsymbol{v}_\varepsilon(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \\
&+ \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|\geq t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{u}_0(\boldsymbol{x})+\boldsymbol{v}_\varepsilon(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} \\
&\leq \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|<t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{u}_0(\boldsymbol{x})+\boldsymbol{v}_\varepsilon(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} + |\{\boldsymbol{x}\in\Omega : |\boldsymbol{v}_\varepsilon(\boldsymbol{x})| \geq t\}| \\
&\leq \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|<t\}} |\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{u}_0(\boldsymbol{x})+\boldsymbol{v}_\varepsilon(\boldsymbol{x})) - \mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x} + \frac{1}{t^2}\int_\Omega |\boldsymbol{v}_\varepsilon(\boldsymbol{x})|^2 d\boldsymbol{x}
\end{aligned}$$

where the last estimate follows from the Chebychev Inequality [3]. For $\varepsilon \in (0,\varepsilon_0)$ and for $t = \|\boldsymbol{v}_\varepsilon\|_{L^2(\Omega)}^{1/2}$ it follows with (A.32) that $|\boldsymbol{c}| > 2 + t$. Then in the integration set, $\{x \in \Omega : |\boldsymbol{v}_\varepsilon(\boldsymbol{x})| < t\}$, it follows from

$$|\boldsymbol{x}+\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon(\boldsymbol{x})| = |\boldsymbol{x}+\boldsymbol{c}+\boldsymbol{v}_\varepsilon(\boldsymbol{x})| \geq |\boldsymbol{c}| - (|\boldsymbol{x}| + |\boldsymbol{v}_\varepsilon(\boldsymbol{x})|) \geq |\boldsymbol{c}| - (1+t) > 1 \tag{A.34}$$

that $\boldsymbol{x}+\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon(\boldsymbol{x})\notin\Omega$ and therefore $\mathcal{I}_0^\varepsilon(\boldsymbol{x}+\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon(\boldsymbol{x}))=0$. Thus, (A.33) can be simplified as follows:

$$S^\varepsilon(\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon) \;\leq\; \int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|<t\}}|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}+\frac{1}{t^2}\int_\Omega|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x} \tag{A.35}$$

The first integral on the right in (A.35) can be estimated further acccording to the Chebychev Inequality,

$$\begin{aligned}&\left|\int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}(\boldsymbol{x})|<t\}}|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}-\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}\right|+\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}\\ &=\int_{\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}(\boldsymbol{x})|\geq t\}}|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}+\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x} \qquad\text{(A.36)}\\ &\leq|\{\boldsymbol{x}\in\Omega:|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|\geq t\}|+\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}\leq\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}+\frac{1}{t^2}\int_\Omega|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}.\end{aligned}$$

Combining (A.35) and (A.36) gives:

$$S^\varepsilon(\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon)\leq\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}+\frac{2}{t^2}\int_\Omega|\boldsymbol{v}_\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}$$

Setting $t=\|\boldsymbol{v}_\varepsilon\|_{L^2(\Omega)}^{1/2}$ gives finally:

$$S^\varepsilon(\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon)\leq\int_\Omega|\mathcal{I}_1^\varepsilon(\boldsymbol{x})|^2d\boldsymbol{x}+2\|\boldsymbol{v}_\varepsilon\|_{L^2(\Omega)}. \tag{A.37}$$

Using (A.30) and (A.29) in (A.37) shows that $S^\varepsilon(\boldsymbol{u}_0+\boldsymbol{v}_\varepsilon)\to0$ as $\varepsilon\to0$. Combining this result with (A.31) gives (A.27). □

References

[1] R.A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.

[2] L. Ambrosio and V.M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via γ-convergence. *Communications on Pure and Applied Mathematics*, 43:999–1036, 1990.

[3] R. B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.

[4] Gilles Aubert and Pierre Kornprobst. *Mathematical Problems In Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer, 2nd edition, 2006.

[5] M. J. Bishop, G. Plank, R. A. Burton, J. E. Schneider, D. J. Gavaghan, V. Grau, and P. Kohl. Development of an anatomically detailed MRI-derived rabbit ventricular model and assessment of its impact on simulations of electrophysiological function. *Am J Physiol Heart Circ Physiol*, 298(2):H699–718, 2010.

[6] Patrick M Boyle, Makarand Deo, and Edward J Vigmond. Behaviour of the purkinje system during defibrillation-strength shocks. *Conf Proc IEEE Eng Med Biol Soc*, 2007:419–422, 2007.

[7] S. C. Brenner. Korn's inequalities for piecewise h^1 vector fields. *Math. Comp.*, 73:1067–1087, 2004.

[8] Marina Cerrone, Sami F Noujaim, Elena G Tolkacheva, Arkadzi Talkachou, Ryan O'Connell, Omer Berenfeld, Justus Anumonwo, Sandeep V Pandit, Karen Vikstrom, Carlo Napolitano, Silvia G Priori, and José Jalife. Arrhythmogenic mechanisms in a mouse model of catecholaminergic polymorphic ventricular tachycardia. *Circ Res*, 101(10):1039–1048, Nov 2007.

[9] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[10] P.G. Ciarlet. *Mathematical Elasticity. Volume II: Theory of Plates*, volume 27 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, 1997.

[11] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice–Hall, Englewood Cliffs, NJ,, 1983.

[12] Makarand Deo, Patrick Boyle, Gernot Plank, and Edward Vigmond. Role of purkinje system in cardiac arrhythmias. *Conf Proc IEEE Eng Med Biol Soc*, 2008:149–152, 2008.

[13] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*. Series Computational Mathematics 35. Springer, 2004.

[14] Manfred Dobrowolski. *Angewandte Funktionalanalysis*. Springer, 2006.

[15] D. J. Dosdall, K. A. Cheng, J. Huang, J. S. Allison, J. D. Allred, W. M. Smith, and R. E. Ideker. Transmural and endocardial Purkinje activation in pigs before local myocardial activation after defibrillation shocks. *Heart Rhythm*, 4(6):758–65, 2007.

[16] M. Droske and W. Ring. A mumford-shah level-set approach for geometric image registration. *SIAM Journal of Applied Mathematics*, 66(6):2127–2148, 2006.

[17] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. AMS, 2008.

[18] J.M. Fitzpatrick, D.L.G. Hill, and C.R. Maurer. *Image Registration, Medical Image Processing*, volume 2, chapter 8 of the Handbook of Medical Imaging. SPIE Press, July 2000.

[19] Matthias Fuchs, Bert Jüttler, Otmar Scherzer, and Huaiping Yang. Shape metrics based on elastic deformations. *J. Math. Imaging Vis.*, 35(1):86–102, 2009.

[20] D. J. Huelsing, K. W. Spitzer, J. M. Cordeiro, and A. E. Pollard. Conduction between isolated rabbit purkinje and ventricular myocytes coupled by a variable resistance. *Am J Physiol*, 274(4 Pt 2):H1163–H1173, Apr 1998.

[21] S.L. Keeling and W. Ring. Medical image registration and interpolation by optical flow with maximal rigidity. *Journal of Mathematical Imaging and Vision*, 23(1):47–65, 2005.

[22] Christian Knauer, Klaus Kriegel, and Fabian Stehn. Minimizing the weighted directed hausdorff distance between colored point sets under translations and rigid motions. In *FAW '09: Proceedings of the 3d International Workshop on Frontiers in Algorithmics*, pages 108–119, Berlin, Heidelberg, 2009. Springer-Verlag.

[23] Jan Modersitzki. *Numerical Methods for Image Registration*. Oxford Science Publications, 2004.

[24] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[25] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, August 2000.

[26] J. M. Ortega. The newton–kantorovich theorem. *The American Mathematical Monthly*, 75(6):658–660, Jun–Jul 1968.

[27] Wladimir Peckar, Christoph Schnörr, Karl Rohr, and H. Siegfried Stiehl. Parameter-free elastic deformation approach for 2d and 3d registration using prescribed displacements. *J. Math. Imaging Vis.*, 10(2):143–162, 1999.

[28] G. Plank, R. A. Burton, P. Hales, M. Bishop, T. Mansoori, M. O. Bernabeu, A. Garny, A. J. Prassl, C. Bollensdorff, F. Mason, F. Mahmood, B. Rodriguez, V. Grau, J. E. Schneider, D. Gavaghan, and P. Kohl. Generation of histo-anatomically representative models of the individual heart: tools and application. *Philos Transact A Math Phys Eng Sci*, 367(1896):2257–92, 2009.

[29] A. J. Prassl, F. Kickinger, H. Ahammer, V. Grau, J. E. Schneider, E. Hofer, E. J. Vigmond, N. A. Trayanova, and G. Plank. Automatically generated, anatomically accurate meshes for cardiac electrophysiology problems. *IEEE Trans Biomed Eng*, 56(5):1318–30, 2009.

[30] Dong-Gyu Sim, Oh-Kyu Kwon, and Rae-Hong Park. Object matching algorithms using robust hausdorff distance measures. *IEEE Transactions on Image Processing*, 8(3):425–429, March 1999.

[31] M.E. Taylor. *Partial Differential Equations: Basic Theory.* Springer-Verlag, New York, 1996.

[32] J. Toriwaki and H. Yoshida. *Fundamentals of Three-dimensional Digital Image Processing.* Springer, 2009.

[33] J. Tranum-Jensen, A. A. Wilde, J. T. Vermeulen, and M. J. Janse. Morphology of electrophysiologically identified junctions between purkinje fibers and ventricular muscle in rabbit and pig hearts. *Circ Res*, 69(2):429–437, Aug 1991.

[34] F.J. Vetter and A.D. McCulloch. Three-dimensional analysis of regional cardiac function: a model of rabbit ventricular anatomy. *Prog Biophys Mol Biol*, 69(2-3):157–83, 1998.

[35] Edward J Vigmond and Clyde Clements. Construction of a computer model to investigate sawtooth effects in the purkinje system. *IEEE Trans Biomed Eng*, 54(3):389–399, Mar 2007.

[36] Chunjiang Zhao, Wenkang Shi, and Yong Deng. A new hausdorff distance for image matching. *Pattern Recognition Letters*, 26(5):581–586, 2005.