Received: 6 October 2015 | Revised: 24 October 2016 | Accepted: 13 January 2017
DOI 10.1002/nla.2095

WILEY

RESEARCH ARTICLE

A stabilized multigrid solver for hyperelastic image registration

Lars Ruthotto[1] | **Chen Greif**[2] | **Jan Modersitzki**[3,4]

[1]Department of Mathematics and Computer Science, Emory University, 400 Dowman Dr, Atlanta, GA 30322, USA
[2]Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, BC V6T 1Z4, Canada
[3]Institute of Mathematics and Image Computing, Maria-Goeppert-Str. 3, 23562 Lübeck, Germany
[4]Fraunhofer MEVIS, Maria-Goeppert-Str. 3, 23562 Lübeck, Germany

Correspondence
Lars Ruthotto, Department of Mathematics and Computer Science, Emory University, 400 Dowman Dr, Atlanta, GA 30322, USA.
Email: lruthotto@emory.edu

Summary

Image registration is a central problem in a variety of areas involving imaging techniques and is known to be challenging and ill-posed. Regularization functionals based on hyperelasticity provide a powerful mechanism for limiting the ill-posedness. A key feature of hyperelastic image registration approaches is their ability to model large deformations while guaranteeing their invertibility, which is crucial in many applications. To ensure that numerical solutions satisfy this requirement, we discretize the variational problem using piecewise linear finite elements, and then solve the discrete optimization problem using the Gauss–Newton method. In this work, we focus on computational challenges arising in approximately solving the Hessian system. We show that the Hessian is a discretization of a strongly coupled system of partial differential equations whose coefficients can be severely inhomogeneous. Motivated by a local Fourier analysis, we stabilize the system by thresholding the coefficients. We propose a Galerkin-multigrid scheme with a collective pointwise smoother. We demonstrate the accuracy and effectiveness of the proposed scheme, first on a two-dimensional problem of a moderate size and then on a large-scale real-world application with almost 9 million degrees of freedom.

KEYWORDS
image registration, biomedical imaging, multigrid methods, numerical optimization

1 | INTRODUCTION

Image registration is a central task in a variety of areas involving imaging techniques, such as astronomy, geophysics, and medical imaging; see reviews[1–7] and references therein. Its goal is to automatically establish geometrical correspondences between two or more image data sets. This primarily entails finding a plausible transformation that aligns the given data sets. Registration is required whenever images taken at different times, from different devices, with different modalities, or of different subjects need to be compared or combined. A particular application of image registration that is relevant to this paper is medical imaging.

Image registration can be formulated as a variational problem. To this end, a distance measure is used to quantify the alignment of the data sets. Common choices for distance measures include the sum-of-squared-differences (SSD), normalized gradient fields, and mutual information; see Modersitzki[6] for an overview. In general, the distance functional depends nonconvexly on the transformation, and thus, the registration problem is known to be ill-posed.[2,8] Tikhonov regularization can be used to ensure existence of solutions, increase robustness against noise, and quantify the plausibility of the transformation. Probably, the most widely used regularization functional is the elastic regularization functional, which can be dated back to Fischler and Elschlager[9] and Broit.[10] It is based on the assumption that images deform like elastic material. Typically the strain tensor is linearized to yield computationally efficient schemes.[11] However, due to the linearization, elastic schemes are limited to small deformations by design. If this assumption is violated, the model may yield estimates of the deformation that are not invertible, which is implausible for elastic materials and not meaningful in many applications. Assuming regularity, local invertibility of a deformation is equivalent to its Jacobian determinant being positive; see Evans[12] [Section C.5]. Thus, invertible solutions can be ensured for linear elastic schemes by imposing constraints on the compression and expansion of tissue.[13]

To accurately compute elastic energies for large deformations, an extension to nonlinear elastic regularizers inspired by the theory of hyperelastic materials was first described by Droske and Rumpf[14] and is also used in other recent works.[15–17] Hyperelastic extensions consider Ogden materials for which the Jacobian determinant directly enters the regularization functional. Thus, expansion is penalized and compression is bounded, which ensures invertibility of the deformation without utilizing additional constraints.[14,18,19]

Ensuring that numerical solutions computed on a finite grid fulfill this requirement, however, is not trivial. Previously, we proposed a discretization that controls volume changes on a tetrahedral subdivision of a rectangular grid.[15] A series of discrete optimization problems are solved using a Gauss–Newton optimization scheme.[20] While this approach has already been used for solving a real-life registration problem in cardiac imaging in the work of Gigengack et al.,[21] a severe computational overhead so far limits broader applications to large-scale registration problems. A large fraction of computational time is spent on approximately solving the Hessian system to obtain a search direction. For practically relevant problem sizes, iterative solvers, such as preconditioned conjugate gradient (PCG) methods, are commonly used; see, Hestenes and Stiefel[22] for the original work on CG and Saad's book[23], for an introduction to preconditioning. In contrast to linear elastic schemes in hyperelastic approaches, the linear system couples all partial derivatives of the transformation. Furthermore, as we will see later, the operator and, in particular, the coupling pattern can change drastically in the course of optimization, and it imposes challenges on designing effective preconditioners.

In this paper, we employ a multigrid scheme for preconditioning the Gauss–Newton system. We motivate the derivation of the method by showing, using a local Fourier analysis (LFA), and by numerical experiments that a standard approach on the basis of a collective pointwise smoother and a Galerkin coarsening strategy may fail in practice; specifically, we show that h-ellipticity factors approach zero when the transformation between the image data requires large compression or expansion of volume. We then develop a stabilization procedure by locally thresholding the coefficients associated with the volume regularization. As we demonstrate, our scheme yields a drastic improvement of scalability for large-scale hyperelastic registration problems.

This paper is closely related to two of our previous works.[15,24] A hyperelastic registration problem was formulated in Burger et al.,[15] where a finite volume discretization based on a tetrahedral subdivision was implemented and the linear systems were solved with a simple Jacobi preconditioned CG scheme. We use the same theoretical formulation as in[15], but propose a new discretization which is based on a Galerkin finite element approach with piecewise linear basis functions. An advantage of our new approach is that the numerical solution inherits the invertibility properties from the continuous problem since the gradient of the transformation is piecewise constant, and thus, the Jacobian determinant is controlled exactly. In addition, we tie the Galerkin finite element framework in a seamless way into the multigrid solver by using a piecewise linear interpolation to obtain coarse grid operators. Haber et al.[24] derived and successfully applied a colored Vanka smoother for a volume constrained registration problem, which is a special case of hyperelastic image registration.[25] In this work, we apply a computationally economical collective pointwise smoother. While this smoother is less effective than the Vanka smoother, our numerical experiments indicate that the collective pointwise smoother yields a robust and efficient multigrid preconditioning scheme.

The remainder of the paper is organized as follows: In Section 2, we briefly introduce the hyperelastic image registration problem, derive a discretization using a Galerkin finite element method, and describe the numerical optimization framework used to solve the discrete minimization problems. In Section 3, we present a Galerkin multigrid method and the proposed stabilization. Section 4 provides numerical experiments. We demonstrate the accuracy and robustness of the proposed method on a 2D problem involving large, nonlinear transformations, as well as on a real-life 3D registration problem involving almost 9 million degrees of freedom.

2 | MATHEMATICAL FORMULATION AND DISCRETIZATION

In this section, we briefly introduce the hyperelastic image registration problem and introduce a Galerkin finite element discretization.

2.1 | Hyperelastic image registration

We are given two images $\mathcal{T}, \mathcal{R} : \Omega \subset \mathbb{R}^d \to \mathbb{R}$, continuously differentiable and compactly supported on the domain Ω. We restrict the following presentation to $d = 3$, which today is the most relevant case in medical imaging. Commonly, $\mathcal{T}$ is referred to as the template image and $\mathcal{R}$ is called the reference image. The goal is to find a transformation $y : \Omega \to \mathbb{R}^3$ such that ideally $\mathcal{T}(y(x)) \approx \mathcal{R}(x)$ for all $x \in \Omega$ and y is plausible in the given application, which is quantified by a regularization energy. For many medical applications, nonsmooth or noninvertible transformations are implausible.

To ensure invertibility of y even for large and nonlinear transformations, we resort to hyperelastic regularization and consider the variational problem

$$\mathcal{J} : \mathcal{A} \to \mathbb{R}^+, \quad \min_y \left\{ \mathcal{J}[y] = \mathcal{D}[\mathcal{T}, \mathcal{R}; y] + \mathcal{S}^{\text{hyper}}[y] \right\}, \tag{1}$$

whose ingredients are discussed subsequently. As the distance functional $\mathcal{D}$, quantifying the alignment of the images, is not the focus of this paper, we restrict the presentation to the SSD

$$\mathcal{D}[\mathcal{T}, \mathcal{R}; y] = \frac{1}{2} \int (\mathcal{T}(y(x)) - \mathcal{R}(x))^2 dx. \tag{2}$$

For this distance measure to be effective, intensity values should approximately agree at corresponding points in the reference and template image. This assumption is often satisfied when comparing images of the same modality, and hence, SSD is a prototype of uni-modal distance functionals. For a more general overview of distance measures see the work of Modersitzki.[6]

The hyperelastic registration functional as motivated and presented in our previous work[15] is given by

$$\begin{aligned} \mathcal{S}^{\text{hyper}}[y] &= \mathcal{S}^{\text{length}}[y] + \mathcal{S}^{\text{area}}[y] + \mathcal{S}^{\text{vol}}[y] \\ &= \frac{\alpha_1}{2} \|\nabla(y - y_{\text{ref}})\|^2 + \alpha_2 \int \phi(\text{cof}\nabla y) dx + \alpha_3 \int \psi(\text{det}\nabla y) dx, \end{aligned}$$

where the penalty functions $\phi : \mathbb{R}^{3\times 3} \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ are convex and $\alpha_1, \alpha_2, \alpha_3 > 0$ are regularization parameters. Choosing optimal regularization parameters for nonlinear inverse problems is a question of active research[26–28] and beyond the scope of this work. The impact of the regularization parameter in hyperelastic image registration is not trivial and extensively discussed in our previous work.[15] In the following we assume that adequate parameters are supplied by the user. A reference transformation y_{ref}, typically the identity of the result of a rigid preregistration, can be used to favor preferred solutions. For a geometrical interpretation of $\mathcal{S}^{\text{hyper}}$, we note that at a point x the columns of $\nabla y(x) \in \mathbb{R}^{3\times 3}$ are tangents of the coordinate lines in the deformed coordinate system passing through x; see, e.g., Ciarlet[19] (p. 30). Thus, the first term in $\mathcal{S}^{\text{hyper}}$ approximates changes in lengths. Similarly, the norm of each column of the cofactor matrix $\text{cof}\nabla y$ approximates the area of a surface element spanned by two lines of the deformed coordinate system, and the determinant of the Jacobian $\det \nabla y$ approximates volume changes induced by the transformation. Clearly, the actual choice of ϕ and ψ offers degrees of freedom. Here, we follow our previous work[15] and use

$$\phi(C) = \sum_{j=1}^{3} \max\left\{ \sum_{i=1}^{3} C_{ij}^2 - 1, 0 \right\} \quad \text{and} \quad \psi(x) = (x^2 - 1)^4 / x^2. \tag{3}$$

The area penalty function ϕ only penalizes area growth and is blind to shrinkage as it otherwise would not be convex.[15] The definition of $\mathcal{S}^{\text{hyper}}$ is valid for transformations in a subset of the Sobolev space $W^{1,2}(\Omega, \mathbb{R}^3)$

$$\begin{aligned} \mathcal{A} = \big\{ y \in W^{1,2}(\Omega, \mathbb{R}^3) : \; & \text{cof}\nabla y \in L^4(\Omega, \mathbb{R}^{3\times 3}), \det \nabla y \in L^2(\Omega, \mathbb{R}), \\ & \det \nabla y > 0 \text{ a.e.}, \; |\textstyle\int y(x) dx| < C_\Omega \big\}. \end{aligned}$$

The measurability requirements on $\text{cof}\nabla y$ and $\text{det}\nabla y$ are motivated by the choices of the penalty functions ϕ and ψ and requirements of a compensated compactness argument.[15,18,19] The volume inequality $\det \nabla y > 0$ implies that the transformation will be locally invertible and the boundedness of the transformation follows from the compact support of $\mathcal{T}$ and $\mathcal{R}$; see our previous work[15] for extended discussion and a proof of existence of minimizers $y \in \mathcal{A}$.

2.2 | Finite element discretization

We discretize the variational problem (1) using globally continuous and piecewise linear finite elements on a tetrahedral mesh. Let $V^1, V^2, \dots, V^{n_V} \in \mathbb{R}^3$ denote vertices and $T_1, T_2, \dots, T_{n_T}$ tetrahedra, where we assume $\text{vol}(T_i) > 0$ for all $i = 1, 2, \dots, n_T$. We search for a minimizer in the space

$$\mathcal{A}_h = \left\{ y \in \mathcal{C}(\Omega, \mathbb{R}^3) : y\big|_{T_i} \in \Pi^1(T_i, \mathbb{R}^3) \text{ for } i = 1, 2, \dots, n_T \right\} \subset \mathcal{A},$$

where Π_1 denotes the space of first order vector-valued polynomials. We use standard nodal Lagrange hat functions $b^1, b^2, \dots, b^{n_V} : \Omega \to \mathbb{R}$ to construct a basis of $\mathcal{A}_h$. The coefficients $\mathbf{y}$ of $y^h \in \mathcal{A}_h$ with respect to that basis are stored component-wise in a column vector of size $3n_V$. We thus have

$$y^h(x) = \sum_{j=1}^{n_V} \begin{pmatrix} \mathbf{y}_j^1 \\ \mathbf{y}_j^2 \\ \mathbf{y}_j^3 \end{pmatrix} b^j(x),$$

where $\mathbf{y}_j^i = y^{h,i}(V^j)$ and in the same way we obtain $\mathbf{y}_{\text{ref}}$ with $(\mathbf{y}_{\text{ref}})_j^i = y_{\text{ref}}^i(V^j)$. In the following, $I_k \in \mathbb{R}^{k \times k}$ is an identity matrix, $1_k \in \mathbb{R}^k$ is a vector of all ones, and the operators $\otimes$ and $\odot$ are the Kronecker- and Hadamard-products, respectively. The entries of the Jacobian matrix ∇y^h are piecewise constant inside each triangle. We store the nine components in a column vector $(B\mathbf{y}) \in \mathbb{R}^{9n_T}$, where B is the discrete vector gradient operator

$$B = I_3 \otimes \nabla^h, \ \nabla^h = \begin{pmatrix} \partial_1^h \\ \partial_2^h \\ \partial_3^h \end{pmatrix}, \ \partial_k^h \in \mathbb{R}^{n_T \times n_V}, \text{ and } \left(\partial_k^h\right)_{i,j} = \partial_k b^i(V^j).$$

Interpolation from nodes to barycenters of the tetrahedra is done by the averaging matrix

$$\mathbf{A} = I_3 \otimes A \in \mathbb{R}^{3n_T \times 3n_V} \text{ with } A_{i,j} = \begin{cases} 1/4, & \text{if } V^j \text{ is node of } T_i \\ 0, & \text{otherwise} \end{cases}.$$

We denote the volume of the tetrahedra by

$$\mathbf{v} \in \mathbb{R}^{n_T}, \text{with } \mathbf{v}_i = \operatorname{vol}(T_i) \text{ and } \mathbf{V} = \operatorname{diag}(\mathbf{v}).$$

We approximate the SSD distance functional by using a midpoint quadrature rule

$$D(\mathbf{y}) = \frac{1}{2} \operatorname{res}(\mathbf{y})^T \mathbf{V} \operatorname{res}(\mathbf{y}), \text{ where } \operatorname{res}(\mathbf{y}) = \mathcal{T}(\mathbf{A}\mathbf{y}) - \mathcal{R}(\mathbf{x}_c),$$

where $\mathcal{R}(\mathbf{x}_c) \in \mathbb{R}^{n_T}$ evaluates the reference image at the barycenters of the undeformed tetrahedral mesh. Similarly, $\mathcal{T}(\mathbf{A}\mathbf{y})$ computes the image intensities on the barycenters of the deformed mesh. Interpolation is used to build continuous images models for the reference and template image from discrete image date; see also the work of Modersitzki.[6] Because the Jacobian matrix of $y^h \in \mathcal{A}_h$ is piecewise constant, the hyperelastic regularization functional is evaluated exactly by

$$S^{\text{hyper}}(\mathbf{y}^h) = \frac{\alpha_1}{2} (\mathbf{y} - \mathbf{y}_{\text{ref}})^T B^T (I_9 \otimes \mathbf{V}) B (\mathbf{y} - \mathbf{y}_{\text{ref}}) + \alpha_2 \mathbf{v}^T \phi(\operatorname{cof} B\mathbf{y}) + \alpha_3 \mathbf{v}^T \psi(\det B\mathbf{y}).$$

In line with the organization of the gradient, the entries of the cofactor matrix are stored in a column-vector of length $9n_T$. The first derivative of the discretized objective function is

$$dJ(\mathbf{y}) = d\mathcal{D}(\mathbf{y}) + dS^{\text{length}}(\mathbf{y}) + dS^{\text{area}}(\mathbf{y}) + dS^{\text{vol}}(\mathbf{y}),$$

where we use chain rule to obtain

$$\begin{aligned} dD(\mathbf{y}) &= \operatorname{res}(\mathbf{y})^T \mathbf{V} \nabla \mathcal{T}(\mathbf{A}\mathbf{y}) \mathbf{A}, \\ dS^{\text{length}}(\mathbf{y}) &= \alpha_1 (\mathbf{y} - \mathbf{y}_{\text{ref}})^T B^T (I_9 \otimes \mathbf{V}) B, \\ dS^{\text{area}}(\mathbf{y}) &= \alpha_2 ((I_9 \otimes \mathbf{v}) \odot \phi'(\operatorname{cof} B\mathbf{y}))^T d\operatorname{cof} B\mathbf{y}, \\ dS^{\text{vol}}(\mathbf{y}) &= \alpha_3 \mathbf{v}^T \psi'(\det B\mathbf{y}) \, d\det B\mathbf{y}. \end{aligned}$$

Using the abbreviation $D_i^j = \operatorname{diag}(\partial_i \mathbf{y}^j) \in \mathbb{R}^{n_T \times n_T}$, we have

$$d\operatorname{cof} B\mathbf{y} = \left(\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & D_3^3 & -D_2^3 & 0 & -D_3^2 & D_2^2 \\ 0 & 0 & 0 & -D_3^3 & 0 & D_1^3 & D_3^2 & 0 & -D_1^2 \\ 0 & 0 & 0 & D_2^3 & -D_1^3 & 0 & -D_2^2 & D_1^2 & 0 \\ \hline 0 & -D_3^3 & D_2^3 & 0 & 0 & 0 & 0 & D_3^1 & -D_2^1 \\ D_3^3 & 0 & -D_1^3 & 0 & 0 & 0 & -D_3^1 & 0 & D_1^1 \\ -D_2^3 & D_1^3 & 0 & 0 & 0 & 0 & D_1^2 & -D_1^1 & 0 \\ \hline 0 & D_3^2 & -D_2^2 & 0 & -D_3^1 & D_2^1 & 0 & 0 & 0 \\ -D_3^2 & 0 & D_1^2 & D_3^1 & 0 & -D_1^1 & 0 & 0 & 0 \\ D_2^2 & -D_1^2 & 0 & -D_2^1 & D_1^1 & 0 & 0 & 0 & 0 \end{array}\right) B \in \mathbb{R}^{9n_T \times 3n_V}.$$

Using the abbreviation $C_i^j = \operatorname{diag}\left((\operatorname{cof} B\mathbf{y})_{i,j}\right) \in \mathbb{R}^{n_T \times n_T}$, we have

$$d\det B\mathbf{y} = \left(C_1^1,\ C_1^2,\ C_1^3,\ C_2^1,\ C_2^2,\ C_2^3,\ C_3^1,\ C_3^2,\ C_3^3 \right)\ B \in \mathbb{R}^{n_T \times 3n_V}.$$

Following the concepts of the image registration framework,[6] we approximate the Hessian of the objective functional to ensure positive semidefiniteness and avoid computations of second derivatives of the (generally noisy) template image as follows:

$$H(\mathbf{y}) \approx d_2 D(\mathbf{y}) + d_2 S^{\text{length}} + d_2 S^{\text{area}}(\mathbf{y}) + d_2 S^{\text{vol}}(\mathbf{y}), \tag{4}$$

with the summands

$$d_2 D(\mathbf{y}) = \mathbf{A}^T (\nabla \mathcal{T}(\mathbf{A}\mathbf{y}))^T \mathbf{V} \nabla \mathcal{T}(\mathbf{A}\mathbf{y}) \mathbf{A}, \tag{5}$$

$$d_2 S^{\text{length}} = \alpha_1 B^T (I_9 \otimes \mathbf{V}) B, \tag{6}$$

$$d_2 S^{\text{area}}(\mathbf{y}) = \alpha_2 (d\operatorname{cof} B\mathbf{y})^T (\operatorname{diag}(I_9 \otimes \mathbf{v})\ \phi''(\operatorname{cof} B\mathbf{y}))\ d\operatorname{cof} B\mathbf{y}, \tag{7}$$

$$d_2 S^{\text{vol}}(\mathbf{y}) = \alpha_3 (d\det B\mathbf{y})^T\ \operatorname{diag}(\mathbf{v} \odot \psi''(\det B\mathbf{y}))\ d\det B\mathbf{y}. \tag{8}$$

Note that in (4), the dependency of the first term on $\mathbf{y}$ is only of low order, and the second term is constant with respect to $\mathbf{y}$. In contrast, the Hessian of the area and volume regularization functionals strongly depend on $\mathbf{y}$, which is a significant source of complications for iterative solvers for the full system. Note in particular that $\psi''(\det B\mathbf{y}) \to \infty$ as $\det B\mathbf{y} \to 0^+$ or $\det B\mathbf{y} \to \infty$. Thus, ill-conditioning is to be expected for large volume changes.

2.3 | Mesh generation

A typical solution strategy in image registration is to approximately solve the variational problem (1) on a coarse-to-fine hierarchy of discretizations often referred to as multilevel strategy, see the work of Modersitzki,[6] or cascadic multigrid described by Bornemann and Deuflhard.[29] The key motivation is to reduce the risk of being trapped in a local minimum and to obtain good starting guesses for the correction steps on finer discretization levels. While the above description of our method requires no assumptions on the mesh, we now describe structured meshes for the purpose of providing a clear description of multilevel and multigrid method. In imaging, the domain is typically rectangular and the data is provided on a regular grid of pixels or voxels. As in our previous work,[15] we use a nodal grid as a base mesh and divide each cell symmetrically into four triangles (2D) or 24 tetrahedra (3D) by introducing additional points in the cell centers and on the faces; see Figure 1. Unlike subdivisions into only two triangles (2D) or five or six tetrahedra (3D), these partitions avoid a bias with respect to a discretization direction, which may lead to implausible solutions; see Ruthotto[25] (Figure 4.5) for a numerical example.

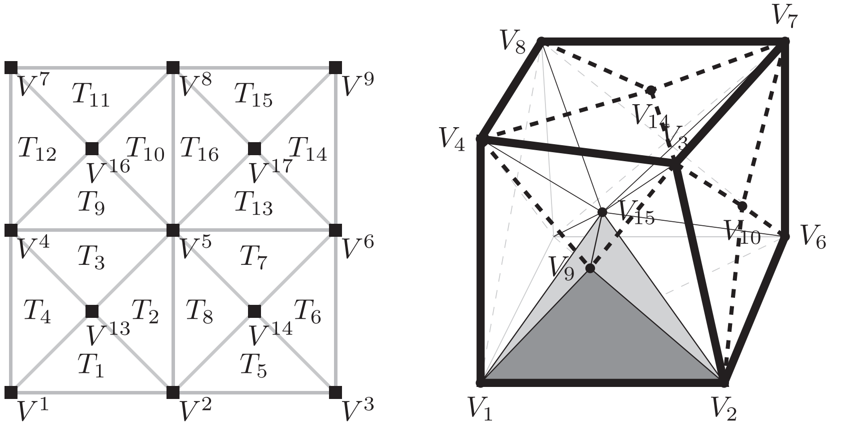

FIGURE 1 Subdivision used for generating nested families of structured meshes.[15] Left: triangulation of a $m = [2, 2]$ grid. Right: tetrahedral subdivision of a voxel

2.4 | Numerical optimization

The variational problem (1) is solved on a family of nested finite dimensional space $\mathcal{A}_{h_c} \subset \ldots \subset \mathcal{A}_{h_f} \subset \mathcal{A}$ using a standard Gauss–Newton method.[20] Thus, a large fraction of the computation time is spent on the solution of the linear system

$$H(\mathbf{y}^k)\delta\mathbf{y}^k = -dJ(\mathbf{y}^k), \tag{9}$$

where dJ is the gradient and H the approximated Hessian of the objective function J at the kth iterate $\mathbf{y}^k$ and $\delta\mathbf{y}^k$ denotes the search direction. For efficient solution of large-scale registration problems, fast iterative solvers are crucial. As outlined above, a particular challenge arises from the dependency of the Hessian on the current transformation.

3 | A STABILIZED MULTIGRID SOLVER

This section presents and analyzes a multigrid method and novel stabilization scheme for effectively computing the search direction $\delta\mathbf{y}$ in Equation 9.

Multigrid methods are widely used to approximate solutions to linear systems, especially systems that arising from discretizations of elliptic partial differential equations such as the system in (Equation 9); see, for example, the monographs.[30–32] The primary observation underlying multigrid for elliptic problems is that high-frequency components in the error are smoothed out quickly by basic iterative solvers such as Gauss–Seidel, whereas the error in the relatively smooth components remains large. Multigrid is thus comprised of the following two components: first, a small number of iterations of a basic iterative scheme (called henceforth a smoother) is applied on the original grid, which generates a smooth error. The second component is a coarse-grid correction: the error is projected onto a coarser mesh, where some of the previously smooth modes are represented by higher frequencies, relative to the (coarser) mesh size. There are several variants of multigrid. One of them, the V-cycle approach applied in this paper, is carried out iteratively, until a sufficiently coarse mesh is reached and the problem can be solved directly. At this point, a procedure of prolongation and postsmoothing is carried out, until the original mesh size is reached. What gives multigrid its power and popularity is the fact that for a large class of elliptic problems, the scheme converges independently of the mesh size, yielding linear complexity.

We first describe a collective pointwise smoother. We then describe our Galerkin based prolongation and restriction operators and comment on the coarse-grid solver. We provide an LFA for a simplified two-dimensional discretization by following the presentation in Trottenberg et al.[32] Motivated by the observation that effectiveness of the multigrid scheme depends primarily on the volume change introduced by the transformation, we derive a stabilized scheme in Section 3.4. We demonstrate the effectiveness of the stabilization scheme by experimentally determining smoothing factors and two-grid convergence factors for a 2D test problem that requires large deformations with severe volume changes in Section 4.

3.1 | Smoother

In hyperelastic image registration, a strong coupling between the components of $\delta\mathbf{y}^k$ in Equation 9 is introduced by the area and volume regularization; see also Equation 4. This difficulty becomes all the more pronounced as the Gauss–Newton iteration evolves and the iterates typically introduce larger compression and expansion of volume. This is indeed a known problem in volume-constrained image registration approaches.[24] Consequently, decoupled pointwise smoothers like Jacobi-schemes that are used for generic elliptic problems are expected to yield poor smoothing factors.[32] This is also confirmed by our experiments. Thus, it is necessary to consider more elaborate smoothing strategies.

We therefore consider a collective smoother.[32] The idea is to include the coupling between the d components of the vector field $y : \Omega \to \mathbb{R}^d$, but to ignore the coupling between different nodes. To this end, at each node, a $d \times d$ system is solved directly (in parallel) to update the coefficients of the vector field. Thus, the arithmetic costs of this smoother grow linearly with respect to the number of nodes in the mesh. Furthermore, because in our applications the spatial dimension, d, is either two or three, the computational costs of applying the collective smoother are comparable to a Jacobi iteration.

3.2 | Prolongation and restriction

We use a standard Galerkin coarsening strategy as described in Trottenberg et al.,[32] which is compatible with our finite element discretization. Specifically, because the quantities in Equation 9 relate to piecewise linear functions, prolongation is equivalent to their evaluation on a finer mesh. This linear operation is summarized in a prolongation matrix and the restriction operator is its transpose.

3.3 | Local fourier analysis and h-ellipticity

We now study the h-ellipticity of our finite element discretization of Equation 9 in two dimensions. The MATLAB script that is used to carry out the numerical computations is available from the first author's website.

For simplicity of presentation, we assume a rectangular domain $\Omega = (0,1)^2$ and a uniform grid, with grid size $h = h_1 = h_2 = 1/m$. Examining Equation 4, we observe that the Hessian of the regularizer, namely, $d_2S^{\text{length}} + d_2S^{\text{area}}(\mathbf{y}) + d_2S^{\text{vol}}(\mathbf{y})$, is symmetric positive semidefinite, with a null space comprised of constant translations in the coordinate directions. The operator $d_2D(\mathbf{y})$ is data-dependent and positive semidefinite, but in general, it does not have constant translations in its null space. Therefore, the fact that there is no intersection generates the effect that when combined together to form the full Hessian, $H(\mathbf{y})$, the eigenvalues are typically shifted further away from zero, compared to the Hessian of the regularizer. In the sequel, we will therefore consider the worst case scenario corresponding to $d_2D(\mathbf{y}) = 0$.

We begin by deriving discrete operators and stencils. To perform smoothing analysis, we assume that all operators are defined on an infinite grid and consider the case of constant coefficients; see Trottenberg et al.[32,Chapter 4] To that end, we consider linear transformations $y(x) = Ax$ for $A \in \mathbb{R}^{2\times 2}$, which leads to $\nabla y(x) = A^\top$.

Let us write the Hessian in a symmetric 2×2 block matrix fashion, that is,

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix},$$

and introduce the shortcuts $w = \psi''(\det A^\top)$ and $c = [a_{22}, -a_{12}, -a_{21}, a_{11}]$ for the entries of the cofactor matrix. We then have

$$H_{11} = \left(\alpha_1 + \alpha_3 w c_1^2\right)\ \partial_1^T\partial_1 + 2\alpha_3 w c_1 c_2\ \partial_1^T\partial_2 + \left(\alpha_1 + \alpha_3 w c_2^2\right)\ \partial_2^T\partial_2, \tag{10}$$

$$H_{12} = \alpha_3 w \left(c_1 c_3\ \partial_1^T\partial_1 + (c_2 c_3 + c_1 c_4)\ \partial_2^T\partial_1 + c_2 c_4 \partial_2^T\partial_2\right), \tag{11}$$

$$H_{22} = \left(\alpha_1 + \alpha_3 w c_3^2\right)\ \partial_1^T\partial_1 + 2\alpha_3 w c_3 c_4\ \partial_1^T\partial_2 + \left(\alpha_1 + \alpha_3 w c_4^2\right)\ \partial_2^T\partial_2. \tag{12}$$

The stencils of the blocks in H can thus be derived as linear combinations of the stencils of the discrete analogues of $\partial_1^T\partial_1$, $\partial_2^T\partial_1 = \partial_1^T\partial_2$ and $\partial_2^T\partial_2$. Because the stencil at nodal grid points differs from the stencil at cell-centered grid points for the symmetric triangulation, we divide the discretized transformation into four parts, that is, $\mathbf{y} = \left(\mathbf{y}_1^n, \mathbf{y}_1^c, \mathbf{y}_2^n, \mathbf{y}_2^c\right)^\top$. Here, the superscripts n and c are used to denote the nodal and cell-centered part, respectively. At nodal grid points, we then obtain the stencils

$$\partial_1^T\partial_1 \sim D_{1,1}^{n,n} + D_{1,1}^{n,c} = \frac{1}{2h^2}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 4 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} + \frac{1}{2h^2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\partial_2^T\partial_2 \sim (D_{1,1}^{n,n} + D_{1,1}^{n,c})^\top,\ \partial_1^T\partial_2 \sim D_{1,2}^{n,n} + D_{1,2}^{n,c} = 0 + \frac{1}{2h^2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

In our discretization, each cell-centered point is connected to the four surrounding nodal points. Thus, the cell-centered stencils can be written as 3×3 matrices

$$\partial_1^T\partial_1 \sim D_{1,1}^{c,c} + D_{1,1}^{c,n} = \frac{1}{2h^2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{1}{2h^2}\begin{bmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix},$$

$$\partial_2^T\partial_2 \sim (D_{1,1}^{c,c} + D_{1,1}^{c,n})^\top, \quad \partial_1^T\partial_2 \sim D_{1,2}^{c,c} + D_{1,2}^{c,n} = 0 + \frac{1}{2h^2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}.$$

Having established notation and the various discrete operators and stencils, we are now ready to perform LFA. The fundamental quantities in LFA are complex grid functions

$$\varphi(\theta, x) = e^{\imath\theta x/h}, \tag{13}$$

where $\imath$ denotes the imaginary unit, x a spatial position, $\theta \in \Sigma$ a frequency with $\Sigma = [-\pi/2, 3\pi/2)^2$ denoting the continuous frequency domain that is partitioned into low and high frequencies

$$\Sigma^{\text{low}} = [-\pi/2, \pi/2)^2 \qquad \text{and} \qquad \Sigma^{\text{high}} = \Sigma \backslash \Sigma^{\text{low}}.$$

Applying the discretized differential operators to the grid function φ yields the Fourier symbols

$$\begin{aligned}
&L^{n,n}[\partial_1^T \partial_1](\theta) = 2 - \cos(\theta_1) + \cos(\theta_2),\\
&L^{n,c}[\partial_1^T \partial_1](\theta) = -2\cos(\theta_1/2)\cos(\theta_2/2) = L^{c,n}[\partial_1^T \partial_1](\theta)\\
&L^{c,c}[\partial_1^T \partial_1](\theta) = 2,\\
&L^{n,n}[\partial_2^T \partial_2](\theta) = 2 + \cos(\theta_1) - \cos(\theta_2),\\
&L^{n,c}[\partial_2^T \partial_2](\theta) = -2\cos(\theta_1/2)\cos(\theta_2/2) = L^{c,n}[\partial_2^T \partial_2](\theta)\\
&L^{c,c}[\partial_2^T \partial_2](\theta) = 2,\\
&L^{n,n}[\partial_2^T \partial_1](\theta) = 0 = L^{c,c}[\partial_2^T \partial_1](\theta),\\
&L^{n,c}[\partial_2^T \partial_1](\theta) = 2\sin(\theta_1/2)\sin(\theta_2/2) = L^{n,c}[\partial_2^T \partial_1](\theta).
\end{aligned}$$

These intermediate quantities are used to compute the h-ellipticity of the Hessian, H, which provides a quantitative criterion for the existence of pointwise smoothers for a given discrete operator; see Trottenberg et al.[32] for details. We obtain the Fourier symbol of the Hessian by combining the above symbols and Equations 10–12 to

$$L[H](\theta) = \left(\begin{array}{cc|cc}
L^{n,n}[H_{11}] & L^{n,c}[H_{11}] & L^{n,n}[H_{12}] & L^{n,n}[H_{12}]\\
L^{c,n}[H_{11}] & L^{c,c}[H_{11}] & L^{c,n}[H_{12}] & L^{c,c}[H_{12}]\\
\hline
L^{n,n}[H_{12}] & L^{n,c}[H_{12}] & L^{n,n}[H_{22}] & L^{n,n}[H_{22}]\\
L^{c,n}[H_{12}] & L^{c,c}[H_{12}] & L^{c,n}[H_{22}] & L^{c,c}[H_{22}]
\end{array}\right) \in \mathbb{C}^{4\times 4}.$$

The h-ellipticity of the operator H then is a qualitative measure for the sensitivity of H to high frequencies[32] and is given by,

$$E(H) = \frac{\min\{|\det L[H](\theta)| \,:\, \theta \in \Sigma^{\text{high}}\}}{\max\{|\det L[H](\theta)| \,:\, \theta \in \Sigma\}} \in [0, 1].$$

A small value of $E(H)$ indicates that there are high-frequency components in the null space of the discretized operator, in which case any multigrid scheme is expected to be ineffective. Thus, it is in general desirable to have $E(H)$ uniformly bounded away from 0 from below.

By design of our hyperelastic regularizer, the volume term grows to infinity as $\det \nabla y \to 0$, namely, when the transformation introduces large compressions. This is realized by the choice of the penalty function ψ in (3), which is also singular at the origin. Because ψ'' contributes a weighting to $d_2 S^{\text{vol}}$ in the Hessian as per Equation 8, and is also singular at the origin, we expect numerical difficulties for large volume compressions. To verify this, we compute the h-ellipticity for a family of compressions $y_c(x) = cx$ where we vary $c \in (0, 1]$ for 10 equidistantly spaced points. Keeping the regularization parameter α_3 fixed, we vary the weighting of the diffusion term $\alpha_1 \in [10^{-1}, 10^5]$ for 41 logarithmically spaced points. The frequency space is partitioned equidistantly into $[128, 128]$ cells. Figure 2 visualizes the h-ellipticity $E(H(\mathbf{y}_c), \alpha_1)$ in dependence of the compression factor c and regularization parameter α_1. As expected, two trends are clearly visible: first, for large compressions, that is, $c \ll 1$, h-ellipticity rapidly approaches 0. Second, h-ellipticity is larger for larger weights α_1. This indicates the existence of potentially effective smoothing schemes when the compression rate is limited and thus the effective weight on the volume term, $\alpha_3 \cdot \psi''$, is not considerably larger than the weight on the length term, α_1.

3.4 | Stabilizing the Hessian

The experiment described in Section 3.3 suggests that the finite element discretization is h-elliptic if the contribution of the volume regularization is not too dominant. In our experience the weighting of the volume term can vary dramatically from tetrahedron to tetrahedron depending on the volume changes. While in most applications volume changes can be limited to a reasonable range, there may be a small number of outliers with extreme volume change affecting the scaling of the Hessian. To improve the robustness of the multigrid scheme in these cases, we limit the effective weight by locally thresholding the second

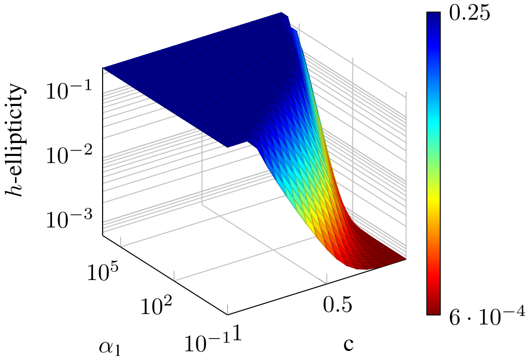

FIGURE 2 Results of our local Fourier analysis. Predictions of h-ellipticity are computed for mappings $y_c(x) = cx$ where $c \leqslant 1$ and regularization parameters for the diffusion term between 0.1 and 10^6. As expected, for given α_1 and increasing compression (decreasing the value of c) the measure of h-ellipticity converges to zero. Thus, the effectiveness of multigrid schemes for our discretization relies on a proper parameter choice

derivative of the volume penalty function in Equation 8

$$\widetilde{\psi''}(x) = \min\{\psi''(x), s \cdot \alpha_1/\alpha_3\}, \tag{14}$$

where s is a positive scalar strictly smaller than one. We denote the stabilized Hessian by H^{stab}.

In principle, the thresholding degrades the quality of the approximation of the Hessian in our Gauss–Newton scheme, which may result in a larger number of outer iterations. However, each such outer iteration is considerably cheaper in most cases, as the multigrid solver is more effective. Furthermore, it is important to stress that only the Hessian of the volume penalty function is altered, but not its value or gradient. Thus, transformations computed using our numerical scheme will have the features guaranteed by the theory and, most importantly, will be invertible.

4 | NUMERICAL RESULTS

In this section, we illustrate the effectiveness of the proposed multigrid scheme based on 2D and 3D problems. The 2D problem is the well-known one of transforming a disc to a C-shaped object, proposed by Christensen[33] and requires a large deformation with large range of compression and expansion. As a large-scale 3D problem, we consider the registration of computer tomography (CT) images of a human knee in straight and bent pose; image courtesy of T. Netsch, Philips Research Hamburg, Germany.

The numerical experiments are performed using the publicly available image registration toolbox flexible algorithms for image registration.[6] In particular, we use the implementation of the Gauss–Newton optimization, which stops if the change in the objective function, the norm of the update, and the norm of the gradient are all sufficiently small at the current iteration. The tolerances for these three conditions are constant in all experiments.

Example 1. To illustrate the challenging effect of large compressions and expansions, we first consider the 2D-test problem of transforming a disc to a C-shaped object as proposed by Christensen[33]; see Figure 3. We use the SSD distance measure (Equation 2) and perform a multilevel registration using four levels with base meshes of size $16^2, 32^2, 62^2$, and 128^2. We use both a direct solver and the proposed multigrid PCG scheme with one presmoothing and one postsmoothing step and a stabilized Hessian ($s = 100$ in Equation 14). Using the direct method, we compute a transformation $\mathbf{y}_{\text{dir}}$ by performing 42, 5, 4, and 9 iterations on the respective levels, which reduces the distance measure from 100% to 0.31%. A comparable number of iterations is required for the proposed multigrid method with stabilized Hessian (40, 8, 5, and 5) yielding a transformation $\mathbf{y}_{\text{stab}}$ that reduces the image distance to 0.37%. Both transformations are almost identical as indicated by an relative error of

$$\text{RE}(\mathbf{y}_{\text{dir}}, \mathbf{y}_{\text{stab}}) = \|\mathbf{y}_{\text{dir}} - \mathbf{y}_{\text{stab}}\| / \|\mathbf{y}_{\text{dir}}\| = 0.0057.$$

Figure 3 shows visualizations of the initial template and reference image, and the computed transformation, which is—as guaranteed by our model—invertible ($\min_x \det \nabla y_{\text{dir}}(x) = 0.41$).

We test the effects of the stabilization proposed in Section 3.4 on the performance of the collective smoother and a two-grid scheme on the coarsest level as follows. At each Gauss–Newton iteration and for each mode in the discretized Fourier domain, we construct a grid function as described in Section 3 (see, for example, Equation 13), assign it as the right-hand side, perform one relaxation sweep as well as a two-grid cycle with one presmoothing and one postsmoothing step and a damping factor of

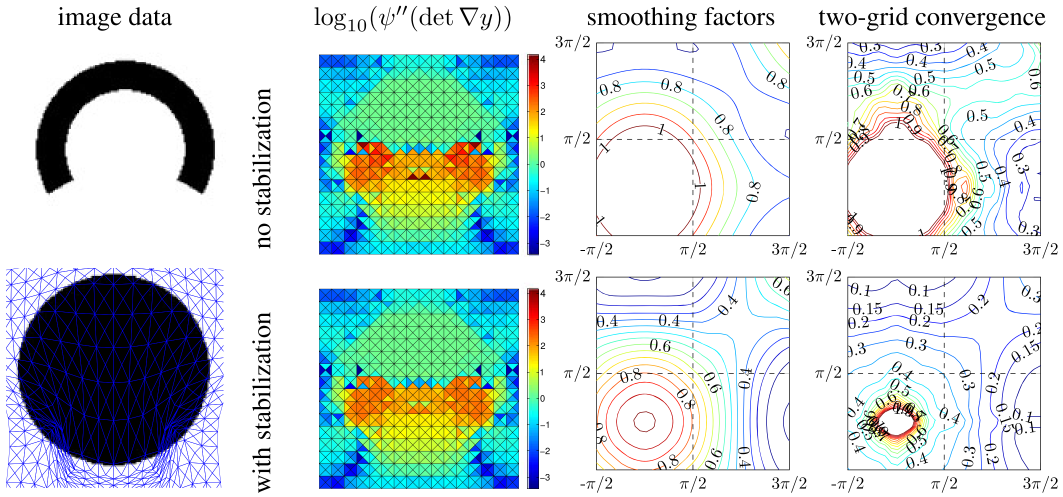

FIGURE 3 Impact of stabilization scheme exemplified for a 2D-test problem of transforming a disc to a C-shaped object as proposed by Christensen.[33] The Hessian of the registration function is considered for a transformation that introduces large volume changes, $\det \nabla y \in [0.05, 3.20]$; see left column. The second column shows the values of the volume penalty function on a logarithmic intensity scale at each triangle. The first row shows the original values that vary initially between more than 9 orders of magnitude whereas the second column shows the thresholded and therefore stabilized coefficients. The same color axis is used in both plots to enable comparison. The third and fourth column show estimated smoothing factors and two-grid convergence factors, respectively

2/3. Finally, we compute the relative reductions of the norm of the residuals. In the first row of Figure 3, we show the estimated smoothing factors for the 21st Gauss–Newton iteration, where the compression is extremely large ($\min_x \det \nabla y(x) = 0.14$). When using the original Hessian in Equation 4, the estimated smoothing factor is 1.03, which means that the collective smoother does not reduce the residual for all high frequencies. Using the proposed stabilization of the Hessian the smoothing factor is reduced to 0.79. As expected the stabilization scheme also improves the two-grid convergence as illustrated for the same outer iteration in the bottom row of Figure 3. Performing a two grid cycle using the original Hessian the largest residual norm for a high frequency right-hand side is 1.26 whereas using the stabilized Hessian the residual norm is below 0.42 for all high frequencies.

Example 2. As an example of a large-scale 3D registration problem, we consider the registration of CT images of a human knee in straight and bent position. The image data is given on a $128 \times 64 \times 128$ grid and a smooth representation is obtained using the cubic B-spline interpolation described by Modersitzki.[6] For motion estimation, we use a three-step multilevel strategy. The transformations are discretized on tetrahedral meshes obtained from base meshes with $24\times16\times24$, $48\times32\times48$, and $96\times64\times96$ cells as described in Section 2.3. On the coarsest level, a rigid registration is performed to correct for translations and rotations. To estimate the bend, we then perform hyperelastic image registration and compare the proposed multigrid preconditioner

TABLE 1 Convergence history for 3D bending knee example. For three discretization levels, we show the reduction of the objective function J in Equation 1, the number of PCG iterations and the relative residuals for Jacobi-PCG, and multigrid-PCG with the original and stabilized Hessian

		Jacobi-PCG			**Multigrid-PCG**			**Multigrid MG-PCG**		
	GNiter	**J**	**#iter**	**relres**	**J**	**#iter**	**relres**	**J**	**#iter**	**relres**
level 4	-1	8.7e7			8.7e7			8.7e7		
	0	1.3e7			1.3e7			1.3e7		
	1	6.0e6	43	9.63e-3	6.0e6	3	9.99e-3	6.0e6	3	9.99e-3
	2	5.4e6	77	9.67e-3	5.4e6	4	5.54e-3	5.4e6	4	4.67e-3
	3	5.2e6	69	1.00e-2	5.2e6	4	3.98e-3	5.2e6	4	3.00e-3
	4	5.2e6	66	9.72e-3	5.2e6	4	2.59e-3	5.2e6	4	2.62e-3
level 5	-1	1.2e8			1.2e8			1.2e8		
	0	1.0e7			1.0e7			1.0e7		
	1	8.0e6	35	9.46e-3	8.0e6	3	7.32e-3	8.0e6	3	7.32e-3
	2	7.2e6	48	9.99e-3	7.2e6	3	8.13e-3	7.2e6	3	8.14e-3
	3	6.6e6	62	9.64e-3	6.6e6	5	8.11e-3	6.6e6	4	3.92e-3
	4	6.5e6	92	9.84e-3	6.5e6	9	5.61e-3	6.5e6	4	4.88e-3
	5	6.4e6	85	9.50e-3	6.4e6	9	8.23e-3	6.4e6	4	3.94e-3
level 6	-1	1.5e8			1.5e8			1.5e8		
	0	9.4e6			9.4e6			9.4e6		
	1	8.2e6	43	9.64e-3	8.2e6	3	8.07e-3	8.2e6	3	8.01e-3
	2	8.1e6	98	9.35e-3	8.0e6	8	7.03e-3	8.0e6	4	6.36e-3

using three presmoothing and two postsmoothing steps and a damping factor of $2/3$ with original and stabilized Hessian to a Jacobi-PCG. The number of coarsening steps in the multigrid schemes are three, four, and five on the respective levels and the coarsest mesh consists of $[3, 2, 3]$ cells. The thresholding factor in Equation 14 was set to $s = 50$. The same regularization parameters $\alpha_1 = 100, \alpha_2 = 1, \alpha_3 = 10$ are used in all experiments. The parameters are determined manually by a rough search of the parameter space. To this end, we fix their ratios, vary only the order of magnitude of the overall regularization, and solve the problem approximately using only the first two levels to speed up computations. We then select a parameter that is small enough to yield an adequate reduction of the distance term while keeping the transformation smooth, that is, keeping the value of the regularization term and the range of the Jacobian determinant small.

The transformations computed with the multigrid schemes were visually and quantitatively almost identical to those computed with the Jacobi preconditioner:

$$\mathrm{RE}(\mathbf{y}_{\mathrm{PCG}}, \mathbf{y}_{\mathrm{MG}(H)}) = 6.3 \cdot 10^{-4} \text{ and } \mathrm{RE}(\mathbf{y}_{\mathrm{PCG}}, \mathbf{y}_{\mathrm{MG}(H^{\mathrm{stab}})}) = 6.3 \cdot 10^{-4}.$$

Also the Jacobian determinant was in about the same range between 0.1 and 6.9. We visualize the registration results obtained using the stabilized multigrid scheme in Figure 4.

The convergence history for the three schemes including the number of CG iterations performed in each Gauss–Newton step is given in Table 1. It can be seen that all preconditioners led in our case to similar reductions of the objective functions and (apart from the stabilized multigrid preconditioner on level 5) needed an identical number of outer iterations to converge.

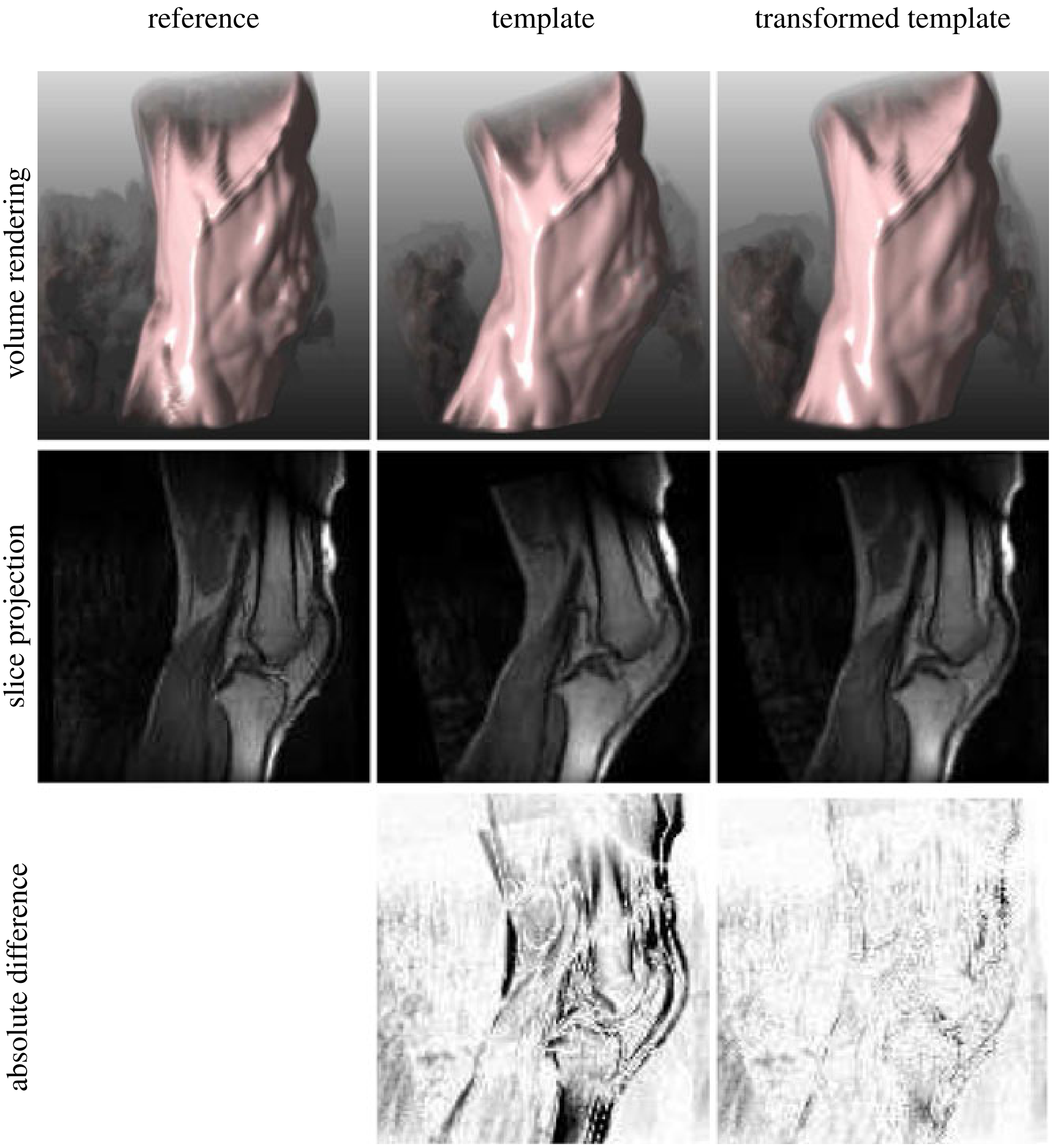

FIGURE 4 Result of 3D hyperelastic registration of CT data of a straight and bent knee, image data from T. Netsch, Philips Research Hamburg, Germany. The top row shows slice projections and the bottom row visualizes the initial and final absolute differences

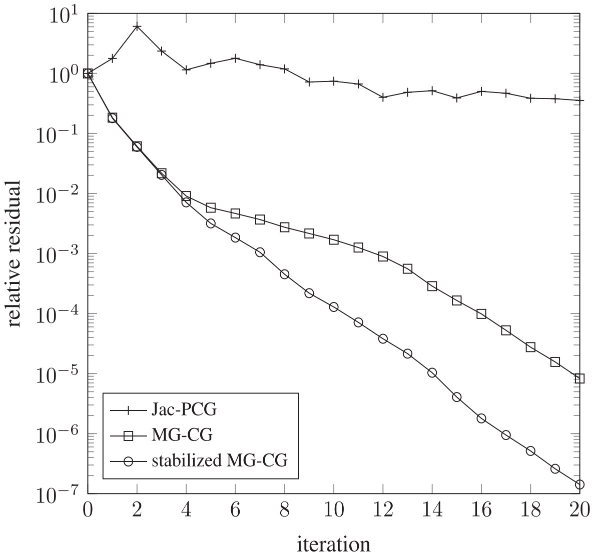

FIGURE 5 Convergence of preconditioned conjugate gradient solver with different preconditioners at the final outer iteration of the large-scale 3D registration problem. MG-CG = multigrid preconditioned conjugate gradient; PCG = preconditioned conjugate gradient

While the performance of the Jacobi-preconditioner varies considerably, and the performance of the multigrid preconditioner moderately, the stabilization scheme is highly effective. Most notably, the number of outer iterations is not increased severely due to the thresholding and that the approximate solution leads to a comparable reduction in the distance measure (4.07% versus 4.12%).

A comparison of the convergence of the three preconditioning schemes is provided in Figure 5. We perform 20 iterations of a PCG scheme for the system in Equation 9 at the final outer iteration on the finest discretization level (about 9 million unknowns). As mentioned above, this iterate introduced a large volume change. It can be seen in Figure 5 that the Jacobi-PCG scheme fails to reduce the relative residual. Both multigrid preconditioned schemes are highly effective, but the stabilized scheme (which entails approximately the same computational cost) is superior to the unstabilized one.

5 | SUMMARY AND CONCLUDING REMARKS

Regularization functionals based on hyperelasticity are a powerful option for applications of image registration entailing large nonlinear deformations. As demonstrated in this paper, they do, however, introduce several computational challenges. To reduce the computational costs associated with determining a search direction, we apply a stabilized multigrid scheme as a preconditioner for the large and sparse Hessian systems that arise throughout the Gauss–Newton iteration. Specifically, we have developed a finite element discretization of the variational problem and applied a Galerkin-based multigrid approach using a collective smoother that relaxes the components of the error individually at each node. We have provided an h-ellipticity analysis, which reveals challenges related to the presence of large compression or expansion factors. On the basis of this observation, we derived a stabilized scheme that limits the contribution of the volume regularizer to the Hessian and does not increase computational costs. The stabilized Hessian can be applied effectively using the proposed multigrid scheme, and our experiments show that the thresholding does not affect the solution to the hyperelastic registration problem.

To illustrate the effectiveness of our algorithm, we provide numerical results for two-dimensional and three-dimensional problems. The 2D problem is highly ill-conditioned and introduces large volume changes; the 3D example is a registration example of CT images of a human knee in straight and bent position with about 9 million degrees of freedom. As shown in our numerical experiments, the novel scheme is highly accurate and effective.

From a linear algebra point of view, we have identified and addressed challenges arising for highly nonlinear transformations. Our h-ellipticity analysis shows that the Hessian systems are expected to be increasingly ill-conditioned as the Gauss–Newton iteration progresses and the current transformation introduces larger volume changes. This is due to the dominance of the Hessian associated to the volume regularization, which has a high-dimensional null space. This challenge is addressed by thresholding the coefficients. Our 2D example shows that this stabilization strategy effectively deals with the ill-conditioning; see Figure 3.

The coefficients in the Hessian can vary considerably in space and thus operator-induced multigrid schemes[32,34] are expected to be more effective. However, operator-induced multigrid schemes are in general considerably more expensive than the Galerkin-coarsening approach presented here.

As our findings suggest, the effectiveness of the smoother most importantly depends on the Jacobian determinant. In future work, we will explore spatial adaptive smoothing strategies that account for this criterion. One idea is to perform Jacobi iterations in regions with smaller and more involving smoothing steps in regions with large volume changes. Another issue that may be addressed is the choice of presmoothing and postsmoothing steps. In our numerical experiments, choices are motivated by practical experiments. Deriving guidelines further informed by analytical considerations would be very beneficial in practice. Another item for future work is the full-scale parallel implementation of the multigrid scheme and comparisons to other subspace correction methods for nonlinear PDEs.[35]

ACKNOWLEDGEMENT

We wish to thank Eldad Haber for his ideas and helpful suggestions in many fruitful discussions.

REFERENCES

1. Brown LG. A survey of image registration techniques. ACM Comput Surv (CSUR). 1992;24(4):325–376. https://doi.org/10.1145/146370.146374
2. Fischer B, Modersitzki J. Ill-posed medicine—An introduction to image registration. Inverse Problems. 2008;24(3):034008. https://doi.org/10.1088/0266-5611/24/3/034008.
3. Glasbey CA, Mardia KV. A review of image-warping methods. J of Appl Stat. 1998;25(2):155–171. https://doi.org/10.1080/02664769823151.
4. Hill DLG, Batchelor PG, Holden M, Hawkes DJ. Medical image registration. Phys Med Biol. 2001;46(3):R1–R45. https://doi.org/10.1088/0031-9155/46/3/201.
5. Maintz JBA, Viergever MA. A survey of medical image registration. Med Image Anal. 1998;2(1). https://doi.org/10.1016/S1361-8415(01)80026-8.
6. Modersitzki J. FAIR: Flexible algorithms for image registration. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 2009. https://doi.org/10.1137/1.9780898718843.
7. Zitová B, Flusser J. Image registration methods: A survey. Image Vision Comput. 2003;21(11):977–1000. https://doi.org/10.1016/S0262-8856(03)00137-9.
8. Weickert J, Schnörr C. A theoretical framework for convex regularizers in PDE-based computation of image motion. Int J Comput Vision. 2001;45(3):245–264. https://doi.org/10.1023/A:1013614317973.
9. Fischler MA, Elschlager RA. The representation and matching of pictorial structures. IEEE Trans Comput. 1973;22(1):67–92. https://doi.org/10.1109/T-C.1973.223602.
10. Broit C. Optimal registration of deformed images. *PhD thesis*, University of Pennsylvania, 1981.
11. Haber E, Modersitzki J. A multilevel method for image registration. SIAM J on Sci Comput. 2006;27(5):1594–1607. http://dx.doi.org/10.1137/040608106.
12. Evans LC. Partial differential equations. Providence: American Mathematical Society; 1998. https://doi.org/10.1090/gsm/019.
13. Haber E, Modersitzki J. Numerical methods for volume preserving image registration. Inverse Prob. 2004;20(5):1621. https://doi.org/10.1088/0266-5611/20/5/018.
14. Droske M, Rumpf M. A variational approach to nonrigid morphological image registration. SIAM J Appl Math. 2004;64(2):668–687. https://doi.org/10.1137/S0036139902419528.
15. Burger M, Modersitzki J, Ruthotto L. A hyperelastic regularization energy for image registration. SIAM J on Sci Comput. 2013;35(1):B132–B148. https://doi.org/10.1137/110835955.
16. Darkner S, Hansen MS, Larsen R, Hansen MF. Efficient hyperelastic regularization for registration. Image Analysis. Berlin: Springer, 2011. p. 295–305. https://doi.org/10.1007/978-3-642-21227-7 28.
17. Yanovsky I, Le Guyader C, Leow A, Toga AW, Thompson PM, Vese L. Unbiased volumetric registration via nonlinear elastic regularization. 2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy. New York: 2008. https://hal.inria.fr/inria-00629762/document.
18. Ball JM. Convexity conditions and existence theorems in nonlinear elasticity. Archive for Rational Mechanics and Analysis. 1976;63(4):337–403. https://doi.org/10.1007/BF00279992.
19. Ciarlet PG. New York: Mathematical elasticity: Three dimensional elasticity. New York: North Holland; 1988.
20. Nocedal J, Wright S. Springer series in operations research and financial engineering. New York: Springer; 1999.
21. Gigengack F, Ruthotto L, Burger M, Wolters CH, Jiang X, Schäfers KP. Motion correction in dual gated cardiac PET using mass-preserving image registration. IEEE Transactions on Medical Imaging. 2012;31(3):698–712. https://doi.org/10.1109/TMI.2011.2175402.
22. Hestenes MR, Stiefel E. Methods of conjugate gradients for solving linear systems. J Res Nat Bur Stand. 1952;49(6):409–436. https://doi.org/10.6028/jres.049.044.
23. Saad Y. Iterative methods for sparse linear systems. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 2003. https://doi.org/10.1137/1.9780898718003.
24. Haber E, Horesh R, Modersitzki J. Numerical optimization for constrained image registration. Numer Linear Algebra with Appl. 2010;17(2-3):343–359. https://doi.org/10.1002/nla.715.
25. Ruthotto L. Hyperelastic image registration. *PhD thesis*, University of Münster, 2012.
26. Vogel CR. Computational methods for inverse problems. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 2002. https://doi.org/10.1137/1.9780898717570.
27. Haber E, Oldenburg D. A GCV based method for nonlinear ill-posed problems. Comput Geosci. 2000;4:41–63. https://doi.org/10.1023/A:1011599530422.
28. Hansen PC. Rank-deficient and discrete ill-posed problems. SIAM Monographs on Mathematical Modeling and Computation. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 1998. https://doi.org/10.1137/1.9780898719697.
29. Bornemann FA, Deuflhard P. The cascadic multigrid method for elliptic problems. Numerische Mathematik. 1996;75(2):135–152. doi:10.1007/s002110050234.
30. Brandt A, Livne OE. Multigrid techniques: 1984 guide with applications to fluid dynamics, vol. 67. Philadelphia, PA: SIAM; 2011. http://bookstore.siam.org/cl67/.
31. Hackbusch W. Multi-grid methods and applications, vol. 4. Berlin: Springer-Verlag; 1985.

32. Trottenberg U, Oosterlee CW, Schuller A. Multigrid. USA: Academic Press; 2001.
33. Christensen GE. Deformable shape models for anatomy. *PhD thesis*, Washington University, 1994.
34. Knapek S. Matrix-dependent multigrid homogenization for diffusion problems. SIAM J on Sci Comput. 1998;20(2):515–533. https://doi.org/10.1137/S1064827596304848.
35. Tai XC, Xu JC. Global and uniform convergence of subspace correction methods for some convex optimization problems. Math Comput. 2002;71(237):105–124. https://doi.org/10.1090/S0025-5718-01-01311-4.

How to cite this article: Ruthotto L, Greif C, Modersitzki J. A stabilized multigrid solver for hyperelastic image registration. *Numer Linear Algebra Appl*. 2017;e2095. https://doi.org/10.1002/nla.2095