Generalization of Deformable Registration in Riemannian Sobolev Spaces

Darko Zikic[1], Maximilian Baust[1], Ali Kamen[2], and Nassir Navab[1]

[1] Computer Aided Medical Procedures (CAMP), TU München, Germany
[2] Siemens Corporate Research (SCR), Princeton, NJ, USA

Abstract. In this work we discuss the generalized treatment of the deformable registration problem in Sobolev spaces. We extend previous approaches in two points: 1) by employing a general energy model which includes a regularization term, and 2) by changing the notion of distance in the Sobolev space by problem-dependent Riemannian metrics. The actual choice of the metric is such that it has a preconditioning effect on the problem, it is applicable to arbitrary similarity measures, and features a simple implementation. The experiments demonstrate an improvement in convergence and runtime by several orders of magnitude in comparison to semi-implicit gradient flows in L^2. This translates to increased accuracy in practical scenarios. Furthermore, the proposed generalization establishes a theoretical link between gradient flow in Sobolev spaces and elastic registration methods.

Keywords: Deformable Registration, Sobolev Spaces, Riemannian Manifolds, Preconditioning.

1 Introduction

The goal of intensity-based deformable registration is the minimization of the similarity measure between a source image I_S, warped by a deformation ϕ, and a target image I_T. The d-dimensional images are defined as $I : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$, and the deformation $\phi \in \mathcal{H}$, $\phi : \Omega \rightarrow \mathbb{R}^d$ is defined in a Hilbert space $\mathcal{H}$, and is expressed in terms of the identity transformation and the displacement $u \in \mathcal{H}$ as $\phi = \mathrm{Id} + u$. The problem is generally modeled as a minimization of an energy consisting of a similarity measure E_D and a regularization term E_R, that is

$$E(\phi) = E_D(\phi) + \lambda E_R(\phi) \ . \tag{1}$$

There are numerous choices for E_D and E_R, cf. [1,2]. The optimization of (1) by a gradient flow in $\mathcal{H}$ consists of iterative application of the evolution rule

$$\partial_t \phi = -\tau \cdot \nabla_{\mathcal{H}} E(\phi) \qquad \text{// update as negative multiple of gradient} \tag{2}$$

$$\phi = \phi \oplus \partial_t \phi \qquad \text{// application of the update} \tag{3}$$

Here, $\nabla_{\mathcal{H}}$ is the gradient in $\mathcal{H}$, and $\oplus$ defines the appropriate update operation [3].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 586–593, 2010.

A common choice for $\mathcal{H}$ is the space of square integrable functions L^2, cf. e.g. [2]. In [4,3], Sobolev spaces are discussed as a choice for $\mathcal{H}$. Compared to L^2, Sobolev spaces have the advantage to contain only functions with certain regularity properties, which is favorable for deformable registration. Due to this inherent regularity, problems which are ill-posed in L^2 can be well-posed in Sobolev spaces. This is the case for the minimization of (1) with $\lambda = 0$. In L^2, this is an ill-posed problem and a regularization term E_R is necessary to allow a numerical treatment, while the problem is well-posed in an appropriate Sobolev space. This has been recognized and put to use in [4,3]. In these approaches, no regularization term E_R is used, and the required smoothness is instead achieved by choosing an appropriate "geometric setting" by employing Sobolev spaces [3].

The motivation behind omitting E_R in these works was to enable the recovery of large deformations, which can be prohibited by strong regularization. This motivation is the same as in the so-called fluid approaches [5], which were shown to be equivalent to minimization of E_D in a suitable Sobolev space [4]. The increased flexibility of fluid approaches comes at the cost of inhibited propagation of ϕ into homogeneous and low contrast image regions, which is otherwise achieved by E_R in the general model (1). Thus, depending on input data, omitting E_R altogether can present a drawback.

In this work, we generalize previous approaches for deformable registration in Sobolev spaces in two points. The first point is that we consider the complete energy from (1), including the regularization term E_R. Compared to traditional fluid-type Sobolev-based approaches (obtained for $\lambda = 0$), the inclusion of E_R is of advantage for treatment of images with large low contrast regions. Besides, this generalization provides a theoretically interesting interpretation of elastic registration methods, by identifying the semi-implicit time discretization of a gradient flow in L^2 (cf. e.g. [2]) as steepest descent in a suitable Sobolev space. In [6], regularization is used together with a Sobolev space setting. This is conceptually quite a different approach and the regularization is performed on velocities, adding a temporal dimension to the problem.

The second generalization is the use of problem-dependent Riemannian metrics for definition of Sobolev spaces. The use of Sobolev spaces in [4,3] is directed at restricting the space of deformations to a certain class, such as diffeomorphisms. Our approach builds on these results by preserving the geometric setting, and extends it by changing the notion of distance in these spaces by Riemannian metrics. This provides us with a flexible theoretical framework, allowing to change the properties of the underlying space, such that its numerical treatment becomes more efficient. We design the Riemannian metrics based on the specific problem and the input data, such that the metric has a preconditioning effect on the optimization problem. We present a strategy to generate the Riemannian metric for arbitrary similarity measures used for registration of medical images, and show that the resulting algorithm exhibits a significantly improved convergence, resulting in much shorter overall runtimes.

By the rationale for the metric choice, our method relates to work on preconditioned gradient descent [7,8]. In contrast to these methods which assume a mono-modal scenario and employ SSD, the proposed technique makes no such assumptions, and is shown to work in multi-modal settings with MI.

It is important to note that while the choice of $\mathcal{H}$ influences the path of the optimization process and the resulting local minimum, it does not affect the definition of (1), such that the optimization operates on the same energy. Thus, by selection of an appropriate metric, we can hope to construct "shorter" paths on the given energy, resulting in more efficient methods.

2 Method

After a brief introduction of Sobolev spaces, and a discussion of previous uses for deformable registration in Sec. 2.1, we generalize the standard approach by using Riemannian metrics for definition of Sobolev spaces in Sec. 2.2. Section 2.3 motivates the selection of metrics based on preconditioning, and in Sec. 2.4, we propose the construction of such metrics for registration purposes. Section 2.5 highlights relations of some well-known methods to the proposed approach.

2.1 Sobolev Spaces and Sobolev Gradients

The following discussion is based on [9,10]. The Sobolev space H^k on $\mathbb{R}^n$ is a Hilbert space of functions whose derivatives up to the order k are square integrable, that is $H^k = \{f : \|f\|_{H^k} < \infty\}$, with

$$\|f\|_{H^k} = \left(\sum_{i=0}^{k} \langle f^{(i)}, f^{(i)} \rangle_{L^2} \right)^{\frac{1}{2}} \equiv \langle f, f \rangle_{H^k}^{\frac{1}{2}} \quad . \tag{4}$$

The H^k scalar product can be written in terms of the L^2 scalar product with the use of the vector-valued differential operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}^k$, $\mathcal{L} = (D^0 \dots D^k)$, consisting of differential operators D^i of order i, as

$$\langle f, f \rangle_{H^k} = \sum_{i=0}^{k} \langle f^{(i)}, f^{(i)} \rangle_{L^2} = \langle \mathcal{L}f, \mathcal{L}f \rangle_{L^2} = \langle \mathcal{L}^*\mathcal{L}f, f \rangle_{L^2} \quad , \tag{5}$$

where $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$, and the differential operator $\mathcal{L}^*\mathcal{L}$ has the form

$$\mathcal{L}^*\mathcal{L} = \sum_{i=0}^{k} (-1)^i \Delta^i \quad . \tag{6}$$

To express the Sobolev gradient by the L^2 gradient, we use the definition of gradient in the space $\mathcal{H}$ as the entity $\nabla_{\mathcal{H}} f$ which can be used to represent the directional derivative $\partial_h f$ by $\partial_h f = \langle \nabla_{\mathcal{H}} f, h \rangle_{\mathcal{H}}$. By applying this for $\mathcal{H} = L^2$ and $\mathcal{H} = H^k$, we can equate the resulting right hand sides, and use (5), yielding

$$\langle \nabla_{L^2} f, h \rangle_{L^2} = \langle \nabla_{H^k} f, h \rangle_{H^k} = \langle \mathcal{L}^*\mathcal{L}\nabla_{L^2} f, h \rangle_{L^2} \quad . \tag{7}$$

From this, we can express the Sobolev gradient in terms of the L^2 gradient as

$$\nabla_{H^k} f = (\mathcal{L}^*\mathcal{L})^{-1} \nabla_{L^2} f \ . \tag{8}$$

In previous works, some modifications of the operator $\mathcal{L}^*\mathcal{L}$ have been proposed. In [3], a general form

$$\mathcal{L}_\alpha^*\mathcal{L}_\alpha = (\alpha\mathcal{L})^*(\alpha\mathcal{L}) = \sum_{i=0}^{k}(-1)^i\alpha_i\Delta^i \ , \tag{9}$$

is considered, with $\alpha_i \in \mathbb{R}$. Two specific instances of (9) are considered in more detail in [3]. We focus on the first one with $\alpha_0 = 1$, and $\alpha_1 = \alpha \in \mathbb{R}$, resulting in

$$\mathcal{L}_\alpha^*\mathcal{L}_\alpha = \mathrm{Id} - \alpha\Delta \ . \tag{10}$$

2.2 Generalization of Sobolev Gradients to Riemannian Manifolds

We generalize (4), by introducing Riemannian metric tensors M_i to the single scalar products by

$$\|f\|_{H_M^k} = \left(\sum_{i=0}^{k}\langle M_i f^{(i)}, f^{(i)}\rangle_{L^2}\right)^{\frac{1}{2}} \ , \tag{11}$$

thus treating the single derivatives in Riemannian manifolds. In contrast to the use of scalars α_i in (9), we employ Riemannian metric tensors $M_i = M_i'^* M_i'$, which are by definition symmetric positive definite, and vary smoothly in the space of deformations. With the operator $\mathcal{L}_M = (M_0' D^0 \dots M_k' D^k)$, we get

$$\mathcal{L}_M^*\mathcal{L}_M = \sum_{i=0}^{k}(-1)^i \ {\nabla^i}^\top M_i \nabla^i \ . \tag{12}$$

According to (8), the gradient with respect to H_k^1 reads

$$\nabla_{H_M^k} f = (\mathcal{L}_M^*\mathcal{L}_M)^{-1} \nabla_{L^2} f \ . \tag{13}$$

Please note that we do not change the class of functions contained in the Sobolev space, but only the notion of distance, because the positive definiteness of all M_i ensures the existence of $0 < c, C < \infty$ such that

$$c \cdot \|f\|_{H^k} \leq \|f\|_{H_M^k} = (\langle\mathcal{L}_M f, \mathcal{L}_M f\rangle_{L^2})^{\frac{1}{2}} \leq C \cdot \|f\|_{H^k} \ . \tag{14}$$

Thus $H^k = \{f : \|f\|_{H^k} < \infty\} = \{f : \|f\|_{H_M^k} < \infty\} = H_M^k$.

In the remainder of the paper, we will restrict our treatment to $k = 1$. The obtained results are however readily transferable to general settings. Corresponding to (10), for the generalized formulation of H_M^1 we get

$$\mathcal{L}_M^*\mathcal{L}_M = M_0 - \mathrm{div}(M_1\nabla) \ . \tag{15}$$

In summary, the computation of the update in (2) for H_M^1 is now based on

$$\nabla_{H_M^1} E(\phi) = (M_0 - \mathrm{div}(M_1\nabla))^{-1} (\nabla_{L^2} E_D(\phi) + \lambda\nabla_{L^2} E_R(\phi)) \ . \tag{16}$$

2.3 Selection of Metric Based on Preconditioning

The general formulation from (16) provides us with a framework in which we can choose the metrics M_i such that the resulting algorithms have advantageous properties. We propose to select these metrics such that the convergence of the algorithm is improved. This can be done by interpreting (13) as the result of preconditioning of the given problem in the L^2 setting. Preconditioning is a standard technique for improvement of convergence rate [11]. To this end, consider the second-order Taylor approximation of f in a Hilbert space $\mathcal{H}$

$$f(x+\alpha h) = f(x) + \alpha\langle h, \nabla_{\mathcal{H}} f(x)\rangle_{\mathcal{H}} + \frac{\alpha^2}{2}\langle h, H_{\mathcal{H}}(f)h\rangle_{\mathcal{H}} + \mathcal{O}(\alpha^3) \ . \tag{17}$$

For a critical point x' with $\nabla f(x') = 0$, the first order term in (17) disappears and H dominantly describes the shape of f about x', so that the condition of H has a direct impact on the convergence of gradient-based methods, see also [10].

As for the gradient in Eq. (8), we can express $H_{H^1_M}$ in terms of H_{L^2} by

$$H_{H^1_M} = \left(\mathcal{L}_M^* \mathcal{L}_M\right)^{-1} H_{L^2} \ . \tag{18}$$

Now, we can influence the condition of $H_{H^1_M}$ by an appropriate choice of $\mathcal{L}_M^*\mathcal{L}_M$ as an approximation to H_{L^2}, thus improving the convergence properties. The choice of $\mathcal{L}_M^*\mathcal{L}_M$ should be simple, efficient, and numerically stable.

2.4 Metric Selection for Deformable Registration

While the structure of $E_D = \int \mathcal{D}_D$ is general as the integration can be performed over the spatial domain (e.g. for SSD), or the intensity domain for statistical measures (CC, CR, MI), the regularization is mostly formulated as a least-squares term in the spatial domain, so that with a differential operator $\mathcal{D}_R$ we can rewrite (1) as $E(\phi) = E_D(\phi) + 1/2 \cdot \lambda\langle \mathcal{D}_R u, \mathcal{D}_R u\rangle_{\mathcal{H}}$. The corresponding L^2 Hessian reads

$$H_{L^2}(E) = H_{L^2}(E_D) + \lambda H_{L^2}(E_R) = H_{L^2}(E_D) + \lambda \mathcal{D}_R^* \mathcal{D}_R \ . \tag{19}$$

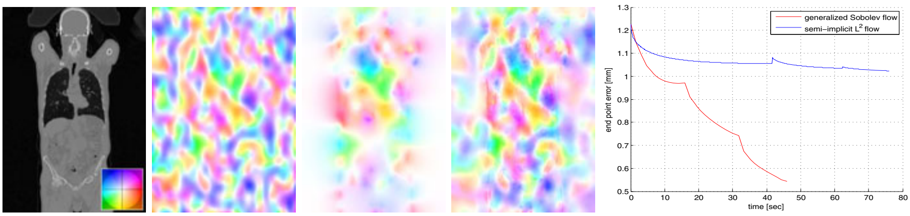
2.4.1 Preconditioning of Regularized Energies by Sobolev Gradients

The first observation is that the use of Sobolev spaces can result in preconditioning of the general energy from (1). To this end, the Sobolev space must be based on the same differential operator $\mathcal{D}_R$ which is employed for the regularization, that is $\mathcal{L} = (\mathrm{Id}, \mathcal{D}_R)$. Thus, with $M_0 = \mathrm{Id}$ and $M_1 = \lambda\mathrm{Id}$ in (15) we get

$$\mathcal{L}_M^* \mathcal{L}_M = \mathrm{Id} + \lambda \mathcal{D}_R^* \mathcal{D}_R \ , \tag{20}$$

which provides a preconditioner for general energies with regularization terms, since $\mathcal{L}^*\mathcal{L}$ from (20) is an approximation to $H_{L^2}(E)$ in (19). This tells us that the steepest descent can be expected to converge much faster in Sobolev spaces than in L^2, since it can be seen as a preconditioned version.

Please note that Sobolev spaces based on (20) are of the form (9) and were employed in [4,3]. In these works however, the above preconditioning argument does not hold since no regularization term E_R is employed.



(a) source (b) GT exampl. (c) L^2 displ. (d) our result (e) convergence plot

Fig. 1. Mono-modal random study. Results in (e) are the mean of 100 trials, w.r.t. computation time. Displacements in (b)-(d) are color-coded, c.f. (a). The proposed method clearly outperforms the semi-implicit L^2 flow in terms in speed and accuracy.

2.4.2 Further Preconditioning by Generalized Sobolev Gradients

The next step to improve the condition of $H_{H^1_M}$ in (18) is to choose M_0 such that $M_0 \approx H_{E_D}$. It is crucial that: 1) the approximation can be computed efficiently, and 2) the approximation is applicable to arbitrary similarity measures. Simple standard preconditioning techniques such as Jacobi preconditioning ($M_0 = \text{diag}\,(H)$) proved ineffective in our experiments.

We propose to compose the Riemannian metric M_0 as a block diagonal matrix, where each $d \times d$ block $M_0(x) \in \mathbb{R}^d$ corresponds to a spatial position $x \in \Omega$. The single blocks $M_0(x)$ are scaled structure tensors of the energy term for single spatial positions $x \in \Omega$, evaluated at the current estimate u, that is

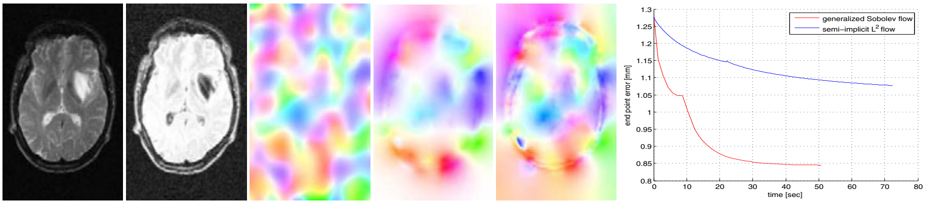
$$M_0(x) = \sigma \text{Id} + \frac{1}{\|\nabla_{L^2} E_D(u)(x)\|} \; \nabla_{L^2} E_D(u)(x) \; \nabla_{L^2} E_D(u)(x)^\top \;\; , \qquad (21)$$

where $\sigma \in \mathbb{R}$ is a small stabilization parameter, which assures the positiveness of M_0, and $\nabla_{L^2} E_D(u)(x) \mathbb{R}^d$ are the sub-vector of the energy gradient at single spatial positions $x \in \Omega$. This choice is applicable for arbitrary similarity measures, it results in a highly sparse, symmetric positive definite d-diagonal metric M_0, which is extremely efficient to compute as $\nabla_{L^2} E_D(u)$ is already calculated in every iteration. Furthermore, the solution of the resulting linear system with the operator $(M_0 - \lambda \Delta)$ can be performed by standard fast linear system solvers [12].

The motivation for the choice of (21) is based on the analysis of the Gauß-Newton optimization for an energy based on SSD and diffusion regularization. It can be shown that in this case, the effect of the preconditioning by the Gauß-Newton method can be seen as an approximate normalization of the magnitudes of the point-wise sub-vectors $\nabla_{L^2} E_D(u)(x)$. This choice is closely related to the analyses of the demons method provided in [13] and [14]. For more details please see the supplementary material at `http://campar.in.tum.de/personal/zikic/miccai2010/index.html`.

2.5 Relation to Other Methods

It is interesting to observe that some well-known methods can be seen as special cases of the proposed approach. Starting from the evolution rule derived in Eq.



(a) source (b) target (c) GT (d) L^2 displ. (e) our result (f) convergence plot

Fig. 2. Random study in a multi-modal setting, using MI, demonstrates applicability to statistical similarity measures. (b)-(e) depict entities from one trial.

(16), we note that for $M_0 = \text{Id}$, and $M_1 = \lambda\tau\text{Id}$ the Sobolev flow is equivalent to semi-implicit discretization of time in the Euler-Lagrange term arising in the L^2-based gradient flow approach [2]. Furthermore, the classical optical flow method by Horn and Schunck [15] can be seen as a generalized Sobolev flow, and is obtained by employing diffusion regularization, and SSD as similarity measure, with $M_0 = J_f^\top J_f$, with J_f being the Jacobian of $f = I_T - I_S \circ \phi$, and $M_1 = \lambda\text{Id}$.

3 Evaluation

We perform 2D random studies in a controlled environment with known ground truth to demonstrate the improvement in convergence and precision, which result from the proposed approach. We compare the proposed method to the standard semi-implicit approach as described in Sec. 2.5. Per study, we perform 100 trials in each of which the source image is warped by a random ground truth deformation ϕ_{GT}, generated by B-Spline FFDs, with maximal displacements of 5mm. Method parameters (α, τ) for the semi-implicit approach are carefully tuned for best possible performance. We monitor the mean euclidean distance between ϕ_{GT} and the estimated deformation (end-point error) in every iteration. A standard multi-level scheme is employed.

The first study is performed on a CT image with SSD as similarity measure, cf. Fig. 1. We demonstrate the applicability of the proposed approach to statistical similarity measures by a study with MI in a multi-modal scenario (Fig. 2). To this end, we employ an MR-T2 image (from http://www.insight-journal.org/RIRE/), with intensities rescaled to $[0, 1]$ as I_S, and register it to $I_T = \widetilde{I_S} \circ \phi_{\text{GT}}$ which includes a non-linear modification of intensities by $\widetilde{I_S}(x) = I_S(x) \cdot (1 - I_S(x))$, in order to simulate a multi-modal scenario. We observe a clear improvement in terms of convergence speed and the actual runtime for the proposed method. The effectively resulting accuracy is also drastically improved, especially in low gradient regions. This is consistent with our choice of the metric in 2.4.2. We observed the same behavior in experiments for SAD and CC as similarity measures. While the single iterations of the proposed method take longer than for the semi-implicit approach, due to the extreme improvement of convergence rate, far less iterations are needed, which results in a significant reduction of the overall runtime. For example, the results in Fig. 1 feature 30 iterations for the proposed, and 550 for the semi-implicit method.

It is important to note that the decrease of the energy is very similar for both approaches (cf. supplementary material) . Based on the inspection of energy logs alone, the semi-implicit method might be considered converged even if the actual error is still significant, c.f. Figs. 1, 2. This premature convergence is a serious pitfall for real applications in which the actual error cannot be measured.

4 Summary

We propose a generalization of previous work on deformable registration in Sobolev spaces by using an explicit regularization term in the energy model, and by modification of the notion of distance by introduction of Riemannian metrics. The general framework in combination with the choice of the Riemannian metric based on the idea of preconditioning leads to a simple and yet powerful method, which outperforms flow strategies in L^2 in terms of speed, and improves the resulting accuracy, especially in low-gradient areas, thus preventing possible premature convergence in real applications.

References

1. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision, IJCV (2002)
2. Modersitzki, J.: Numerical methods for image registration. Oxford Univ. Pr., Oxford (2004)
3. Chefd'hotel, C.: Geometric Methods in Computer Vision and Image Processing: Contributions and Applications. PhD thesis, ENS Cachan (2005)
4. Trouvé, A.: Diffeomorphisms groups and pattern matching in image analysis. International Journal of Computer Vision, IJCV (1998)
5. Christensen, G., Rabbitt, R., Miller, M.: Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing, TIP (1996)
6. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision, IJCV (2005)
7. Haber, E., Modersitzki, J.: A multilevel method for image registration. SIAM Journal on Scientific Computing (2006)
8. Klein, S.: Optimisation methods for medical image registration. PhD thesis, Image Sciences Institute, UMC Utrecht (2008)
9. Neuberger, J.: Sobolev gradients and differential equations. Springer, Berlin (1997)
10. Renka, R.: A simple explanation of the sobolev gradient method (2006)
11. Nocedal, J., Wright, S.: Numerical optimization. Springer, Heidelberg (2000)
12. Saad, Y.: Iterative methods for sparse linear systems. SIAM, Philadelphia (2003)
13. Pennec, X., Cachier, P., Ayache, N.: Understanding the demon's algorithm: 3d nonrigid registration by gradient descent. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 597–605. Springer, Heidelberg (1999)
14. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage (2009)
15. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence (1981)