

A General Preconditioning Scheme for Difference Measures in Deformable Registration

Darko Zikic¹, Maximilian Baust¹, Ali Kamen², and Nassir Navab¹

¹ Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

² Siemens Corporate Research (SCR), Princeton, USA

Abstract

We present a preconditioning scheme for improving the efficiency of optimization of arbitrary difference measures in deformable registration problems. This is of particular interest for high-dimensional registration problems with statistical difference measures such as MI, and the demons method, since in these cases the range of applicable optimization methods is limited. The proposed scheme is simple and computationally efficient: It performs an approximate normalization of the point-wise vectors of the difference gradient to unit length. The major contribution of this work is a theoretical analysis which demonstrates the improvement of the condition by our approach, which is furthermore shown to be an approximation to the optimal case for the analyzed model. Our scheme improves the convergence speed while adding only negligible computational cost, thus resulting in shorter effective runtimes. The theoretical findings are confirmed by experiments on 3D brain data.

1. Introduction

Given two input images, the target image I_T and the source image I_S , the task of registration is to compute the transformation ϕ , such that ϕ maps between the corresponding points in the input images. Intensity-based deformable registration is generally formulated as an optimization of an energy E combining a difference measure E_D and a regularization term E_R by

$$E(\phi) = \gamma E_D(\phi) + \lambda E_R(\phi), \quad (1)$$

with positive scalar factors γ and λ .¹ For high-dimensional deformation models, the difficulty of efficient optimization of (1) depends largely on the choice of E_D . While many standard methods can be used for point-wise difference measures such as the sum of squared differences (SSD), alternatives for efficient optimization of statistical measures

such as mutual information (MI) [18, 8], or correlation ratio (CR) [13] are more rare and complex. The goal of this work is to provide a simple scheme for efficient optimization of arbitrary difference measures.

For the treatment of multi-modal registration problems in medical image analysis, the use of statistical difference measures such as MI or CR is of particular interest. These measures operate on the joint probability distribution of the intensities of the two input images, hereby linking all points with the same intensities, which results in a non-local character of the measures. This is in contrast to difference measures which are defined by point-wise comparisons, such as SSD. The non-locality of statistical measures has consequences on the structure of the corresponding Hessian matrix, which is a major building block for many standard optimization methods. While the Hessian of the SSD measure is a sparse matrix with a small number of non-zero diagonals, the Hessian of MI is dense without any regular sparsity pattern. The size of Hessian matrix H for MI thus becomes prohibitive for practical treatment for high-dimensional settings - with the number of parameters being equal to the number of voxels times the dimension of the images, we get sizes of H in the order of $((3 \cdot 10^6) \times (3 \cdot 10^6))$ for medium size problems. The same problem holds for the Jacobian J of statistical difference measures. This property rules out a number of standard efficient optimization methods for the use with statistical measures, such as the Newton method (employing H), or as Gauss-Newton or Levenberg-Marquardt (employing $J^T J$), which are good choices for point-wise measures such as SSD. Please see also [9] for a more detailed description of this issue.

Therefore, efficient optimization methods for statistical difference measures in high-dimensional settings must circumvent the problem of dealing with dense matrices H or $J^T J$. Standard techniques to achieve this goal include quasi-Newton methods such as L-BFGS, nonlinear conjugate gradient (NL-CG), or preconditioning techniques [11].

While all of these methods can in principle be applied for approaches based on generic optimization of the energy in

¹ E in (1) is highly non-linear, and we compute only local solutions.

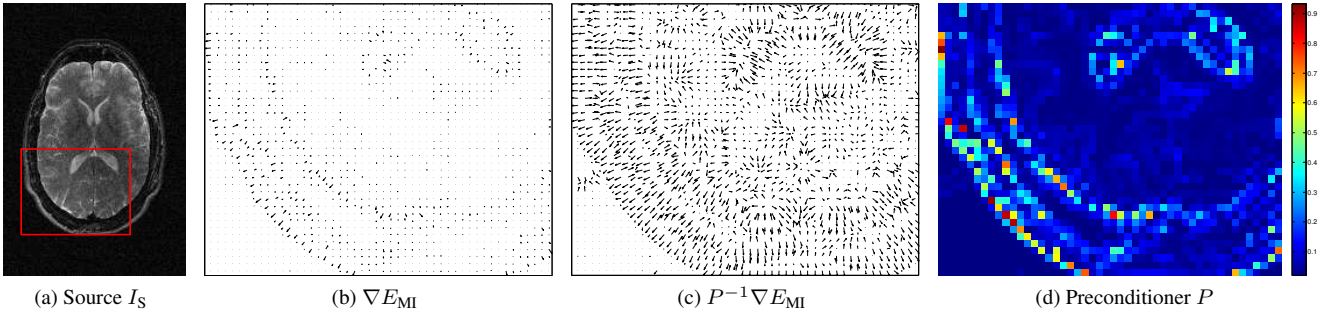


Figure 1: Example of the effect of the proposed preconditioning scheme for MI. The point-wise gradient vectors $\nabla E_D(x)$ are normalized to approximately same length by $P^{-1}\nabla E_D(x)$. Our analysis reveals that this simple strategy improves the condition for arbitrary difference measures E_D , which improves the convergence speed of gradient descent methods.

(1), the use of L-BFGS and NL-CG is not theoretically justified for the standard interpretation of the popular *demons* registration method as an alternating optimization process, in which the optimization steps for E_D are interleaved with smoothing operations, which correspond to optimization of E_R [9]. Since L-BFGS and NL-CG operate by utilizing the information about the energy gradient from subsequent iterations, this process for E_D is disturbed by the smoothing step, which makes this information inconsistent [9]. Presumably in consequence of these properties, optimization of E_D by steepest gradient descent (SGD) is still the standard choice for multi-modal demons registration.

In this work, we propose a simple but theoretically justified preconditioning scheme for the optimization of arbitrary difference measures. Since the derivation of the method relies only on the general properties of a generic difference measure, without any approach-specific considerations, it has the advantage of being applicable to any deformable registration method, and for any difference term. The major idea behind the proposed preconditioning scheme is extremely simple: Modify the gradient of the difference measure, such that the force vectors $\nabla E_D(x)$ at single points in space have possibly similar magnitudes, please compare Fig. 1 for an illustration. In this work, we perform a theoretical analysis which demonstrates that the above scheme actually improves the condition of the problem, which results in faster convergence. Actually, the analysis yields that our scheme is an approximation to the optimal preconditioning of the problem, with respect to the made assumptions. Also, due to the simplicity of the scheme, only minimal computational cost is added compared to SGD.

While the proposed approach is generally applicable in any registration framework, it seems particularly interesting for the demons method, due to its previously discussed restrictions with respect to available optimization methods. Therefore, in this work, we focus on this setting and demonstrate the potential of our scheme in the demons framework.

1.1. Related Work

The preconditioner that we discuss in this work is related to our choice of the Riemannian metric in [19]: It corresponds to the part of the metric related to E_D . While in [19], the term is chosen heuristically, we provide an analysis which demonstrates that our scheme actually improves the condition for arbitrary difference measures. This result can be used to theoretically justify the findings in [19]. On the other hand, the results in [19] can be seen as an exemplary application of the proposed scheme also to other registration methods than the demons approach.

With respect to the demons method, most applications are employing it for mono-modal registration based on SSD, with the original force term, which is based on the Gauss-Newton method. The demons framework has also been applied in some multi-modal settings with statistical difference measures. However, most of these works assume the steepest gradient descent (SGD) update step for the forces [2, 5]. To our best knowledge, the only exception to the use of SGD is the work by Modat *et al.* in [9], where the force is computed as an update of the non-linear conjugate gradient (NL-CG) optimization method [11]. Boiled down, the NL-CG method computes the force term as a specific weighted average of the gradients of the difference measure from the last two iterations [11].

While NL-CG does not require an approximation of the Hessian, which would be problematic for statistical difference measures, it has the following disadvantages (also discussed in [9]). First, the usage of NL-CG is theoretically not justified for demons, since the gradient information about the difference measure from consecutive iterations is not consistent due to the interleaved regularization step [9]. While, in spite of this point, NL-CG reportedly can improve the convergence speed compared to the SGD, no quantitative comparison is performed in [9]. Furthermore, NL-CG requires a precise step size estimation in order to achieve an improvement in convergence [9, 11]. Since pre-

cise step-size search in deformable registration has the cost of several iterations of the algorithm, the increased computational cost can easily outweigh the improvement in convergence speed.

In the context of high-dimensional elastic registration, the L-BFGS method [11] has been employed e.g. in [10]. The essence of L-BFGS is the approximation of the effects of applying H^{-1} to the energy gradient, without actually setting up or inverting H . Since the approximation in L-BFGS also relies on accumulation of the gradient information from several iterations, it shares the same disadvantages as NL-CG, which makes it unsuitable for the demons framework. Compared to NL-CG, the additionally required storage for the n last gradients is a further disadvantage.

Preconditioning schemes have not been studied widely for deformable registration to our best knowledge. An exception is [7], which proposes a specific scheme for SSD.

In contrast to the above methods, the use of the proposed preconditioning approach is applicable to arbitrary difference measures, and it is justified also in the demons setting, since it does not require information from previous iterations. Furthermore, the proposed method does not require a precise step size search, and improves the convergence speed already with a simple fixed step size strategy, thereby directly translating the gain in convergence speed to an effective improvement of runtime.

2. Methodology

We proceed in the following manner. First, to present the setting for the application of the proposed preconditioning scheme, we provide a brief description of the demons framework in Sec. 2.1, and we also discuss how the scheme can be employed for other methods in Sec. 2.2. Then we present the actual preconditioning scheme in Sec. 2.3. The analysis in Sec. 2.4 shows that even the restricted optimization of difference measures for deformable registration is in general ill-conditioned, and that the proposed scheme improves the condition of the problem. The theoretical findings are confirmed by the experiments in Sec. 3.

2.1. The Demons Framework

Due to its efficiency and simplicity of implementation, the demons method [14] has become a popular choice in numerous applications. We consider a general demons framework which computes the transformation $\phi = \text{Id} + u$ by

$$f = \text{compute_force}(E_D) \quad (2)$$

$$g = \tau\gamma G_{\sigma_{\text{fl}}} * f \quad (3)$$

$$u_{i+1} = G_{\sigma_{\text{el}}(\tau, \lambda)} * (u_i \circ (\text{Id} + g)) \quad (4)$$

The above framework contains several extensions of the original approach, which have been proposed in the literature. First, the original elastic-type approach (filtering by

$G_{\sigma_{\text{el}}}$ in (4)) is combined with fluid-type regularization (filtering by $G_{\sigma_{\text{fl}}}$ in (3)) which is applied not to the complete displacement, but only to the force f , as discussed for example in [12]. Second, assuming a group structure for deformations, composition is used as the update scheme in (4) [15, 2, 17]. Finally, the optimization now contains a step size parameter τ , which is included directly in (3), and enters (4) through the width of the Gaussian kernel which is defined by $\sigma_{\text{el}}(\tau, \lambda) = 2\sqrt{\tau\lambda}$. The discretization of images and the displacements is performed on a dense, Cartesian grid, corresponding to position of image voxels, and linear interpolation for computation of values at off-grid points.

A common interpretation which ties in the demons approach within the energy model in (1), is to see the demons approach as optimization of the energy E , by alternating optimization of the expressions containing the difference measure E_D and the regularization term E_R [12, 1]. Within this interpretation, the computation of the forces in (2) corresponds to a step of the optimization of E_D in the L^2 space, the fluid regularization in (3) can be seen as a projection to a corresponding Sobolev space [2], and the regularization step in (4) corresponds to one optimization step of E_R .

The focus of this work is on the computation of the force term in (2). In the original approach this was performed by

$$f = \frac{1}{(I_T - I_S(\phi))^2 + \|\nabla I_S(\phi)\|^2} \underbrace{(I_T - I_S(\phi)) \nabla I_S(\phi)}_{= -\nabla E_{\text{SSD}}(\phi)}, \quad (5)$$

which can be seen as a modification of the gradient of the SSD. It was demonstrated that (5) corresponds to a step of a Newton-type optimization scheme on the SSD [12]. Most current approaches which extend the demons framework to other difference measures, compute the forces by employing the steepest descent scheme, that is

$$f = -\nabla E_D(\phi). \quad (6)$$

The only exception to this approach known to us is [9], where the forces are based on the NL-CG method.

In this work, we consider the computation of the demons forces based on a preconditioned gradient descent by

$$f = -P^{-1} \nabla E_D(\phi), \quad (7)$$

with the operator P^{-1} as discussed in Sec. 2.3.

2.2. Application to Other Methods

For registration methods based on generic optimization, a preconditioned gradient descent can be employed for the energy E from (1) by incrementally updating ϕ by

$$h = -K^{-1} \nabla E(\phi), \quad (8)$$

where K is a symmetric positive definite operator, which should present a numerically favorable approximation to

H_E . The art of preconditioning consists in designing such a K for a specific problem. While the difference terms E_D are non-linear, most regularization terms are of the form $E_R = 1/2 \langle A_R u, u \rangle$, with the corresponding Hessian being $H_{E_R} = A_R$. Thus, for the scheme from (8), we have to estimate only the preconditioning term P for the difference measure, and we get

$$h = -(P + A_R)^{-1} \nabla E(\phi). \quad (9)$$

The above equation shows how the preconditioning term P , which we discuss in Sec. 2.3 can be employed also for registration methods other than demons, cf. also [19].

2.3. Preconditioning Scheme

In this work, we propose a simple preconditioning scheme, which modifies the magnitudes of $\nabla E_D(x)$, such that they become possibly similar, please see Fig. 1 for an illustration. Point-wise, this can be achieved by defining the action of the preconditioner P as a multiplication of $\nabla E_D(\phi)(x)$ with a positive scalar by

$$P^{-1} \nabla E_D(\phi)(x) = \frac{1}{\|\nabla E_D(\phi)(x)\| + \sigma} \nabla E_D(\phi)(x). \quad (10)$$

This corresponds to a diagonal, positive definite operator P , with blocks corresponding to $x \in \Omega$ defined by $P|_x = \text{diag}(\|\nabla E_D(\phi)(x)\| + \sigma)$. While P being positive establishes the formal requirements for a proper gradient descent, it has not yet been shown that the application of P will actually improve the condition of E_D . We will demonstrate this property by the analysis in Sec. 2.4.

It is interesting to note that in the case of SSD, the proposed scheme from (10) approximates the original demons force in (5), since for small displacements, we have $|I_T - I_S(\phi)| \approx \|\nabla I_S(\phi)\|$, and

$$\begin{aligned} \|\nabla E_{\text{SSD}}(\phi(x))\| &= \|(I_T - I_S(\phi)) \nabla I_S(\phi)\| \\ &\approx \frac{1}{2} (I_T - I_S(\phi))^2 + \frac{1}{2} \|\nabla I_S(\phi)\|^2 \end{aligned} \quad (11)$$

Thus our approach can be seen as a natural generalization of the original demons to arbitrary difference measures.

2.4. Analysis of the Preconditioning Scheme

The analysis of the proposed preconditioning is performed as follows. First, we define the model for the analysis of the condition, which operates by measuring the variation of the energy about the optimum, and is based on the energy gradients (Sec. 2.4.1). Second, for optimization of difference measures, which is an under-constrained problem for high-dimensional deformable registration, the analysis has to be restricted to permissible, meaningful directions (Sec. 2.4.2). Finally, we perform the condition analysis for a restricted optimization of difference measures in deformable registration (Sec. 2.4.3). We find that the proposed scheme approximates the optimal improvement of the condition, w.r.t. to the made assumptions.

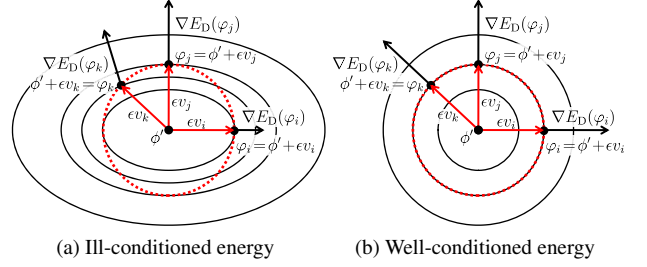


Figure 2: Illustration of the model for evaluating the condition of the optimization problem, cf. Eq. (14). For registration, the directions v correspond to displacement fields, and ϵv represent permissible small warpings of the source image about the optimal deformation, compare also Fig. 3.

2.4.1 Model for Condition Analysis

The condition of an optimization problem can be intuitively seen as the description of the geometry of the energy function about the optimum. A well-conditioned energy function has a possibly spherical shape. The process of preconditioning can be interpreted as modification of the underlying space to render the energy possibly spherical. For a critical point ϕ' , the perfectly conditioned problem can be formalized by

$$E_D(\phi' + v_i) = E_D(\phi' + v_j), \quad (13)$$

where the energy is varied in all possible directions v with a fixed length $\|v\| = r$. Intuitively, the model states that *for a perfectly conditioned problem, the change of energy about the optimum ϕ' should depend only on the distance to ϕ' and not the probing direction v .*

In order to tie the above condition to the gradient of the function, we employ the following alternative form

$$\langle \nabla E_D(\phi' + \epsilon v), v \rangle, \quad (14)$$

which describes a spherical function if (14) is constant for all directions v with a fixed length $\|v\| = r$. For a visualization, please see Fig. 2. If the energy is not perfectly conditioned, then the term in (14) varies, and we measure the quality of the shape of E_D by the variation of the values of (14) for all possible directions v , which is bounded by

$$\kappa(E_D) = \frac{\max_v \langle \nabla E_D(\phi' + \epsilon v), v \rangle}{\min_v \langle \nabla E_D(\phi' + \epsilon v), v \rangle}. \quad (15)$$

A high κ corresponds to an ill-conditioned problem, with the perfectly conditioned function expressed by $\kappa = 1$.

Our goal will be to determine a preconditioner P , which minimizes or at least reduces the value of

$$\kappa_P(E_D) = \frac{\max_v \langle P^{-1} \nabla E_D(\phi' + \epsilon v), v \rangle}{\min_v \langle P^{-1} \nabla E_D(\phi' + \epsilon v), v \rangle} \quad (16)$$

compared to no preconditioning, i.e. $P = \text{Id}$.

2.4.2 Restriction of Analysis to Permissible Directions

For optimization of difference measures in deformable registration, we have to restrict the set of allowed directions v , since in general, the optimization of the unregularized difference measure subject to a high-dimensional deformation is an under-constrained problem, and the condition is infinitely bad per definition. However, we are interested in improving the condition along the constrained dimensions of the problem. The problem is under-constrained since not all deformations result in changes of the energy. For example, this is the case for deformations in homogeneous image areas or along the level set lines of the source image. Therefore, we exclude such under-constrained directions from the analysis, and focus on deformations which do not contain any under-constrained components. Such “pure” displacements v are characterized by being point-wise parallel to the corresponding intensity gradient of the source image, i.e. $v(x) \parallel \nabla I_S(x)$. It is important to note that - with the assumption of group structure for deformations - the original energy gradient has the above permissible structure through $\nabla E_D(\phi)(x) = \omega(x)\nabla(I_S \circ \phi)(x)$, cf. e.g. [2]. An illustration of the permissible directions is given in Fig. 3.

Through the connection between ∇I_S and ∇E_D , we can formalize the deformations v as point-wise rescaling of the directions of the energy gradient

$$v(x) = \tilde{\alpha}(x) \nabla E_D(\phi)(x) = \alpha(x) \frac{\nabla E_D(\phi)(x)}{\|\nabla E_D(\phi)(x)\|}, \quad (17)$$

with $\alpha : \Omega \rightarrow \mathbb{R}$. We define the spatial subset $\Omega' = \{x \in \Omega : \|\nabla E_D(\phi)(x)\| \neq 0\}$, and set $\alpha(x)|_{\Omega'} \geq 0$, and otherwise $\alpha(x)|_{\Omega \setminus \Omega'} = 0$, and $v(x)|_{\Omega \setminus \Omega'} = 0$. Scaling by α instead of $\tilde{\alpha}$ in (17) will facilitate the further analysis. Please note that, given $\nabla E_D(\phi)$, the directions v are completely determined by the choice of α .

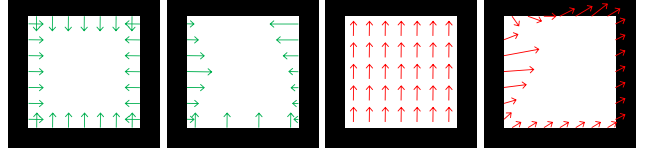
In order to restrict the directional vectors v to a sphere, such that (17) complies with (14), α has to be chosen such that $\|v\| = r$, which gives the following condition

$$r^2 = \|v\|^2 = \sum_{x \in \Omega'} \alpha(x)^2. \quad (18)$$

It is important to note that the choice of α is not unique, and we define the set of all permissible α values as A . Also, due to the construction in (17), A is only dependent on $|\Omega'|$, i.e. the number of non-zero point-wise vectors $\nabla E_D(x)$, and not their actual magnitudes or directions.

For the analysis we will employ the directions v as defined in (17), and substitute these to the criterion (14). Thus, we have to employ the test deformations φ with

$$\varphi = \phi' + \epsilon v \quad \text{with} \quad v(x) = \alpha(x) \frac{\nabla E_D(\varphi)(x)}{\|\nabla E_D(\varphi)(x)\|}. \quad (19)$$



(a) valid test dir. (b) valid test dir. (c) invalid test dir. (d) invalid test dir.

Figure 3: Illustration of the restriction to permissible test directions for condition analysis. Directions v are represented by displacement fields, superimposed onto the source image, warped by a scaling transformation. (a) gradient ∇E_D , and (b) general valid test field, obtained point-wise by $\alpha(x)\nabla E_D(x)$, with $\alpha(x) \geq 0$. (c) is invalid since its application results in no change of energy. (d) is invalid as the point-wise vectors $\nabla E_D(x)$ are re-oriented, thus containing a component along the image level set lines, which does not change the energy. Also, some $v(x)$ point in the opposite direction of the original gradient, i.e. $\langle \nabla E_D(x), v(x) \rangle < 0$.

It is easy to demonstrate the existence of such points for a quadratic function E . Consequently for $\epsilon \rightarrow 0$, the existence of such points can also be assumed.

Also, we restrict the class of preconditioners P , such that $P^{-1}\nabla E_D$ is permissible in the above sense. This is achieved by restricting P to a point-wise multiplication, such that $P^{-1}\nabla E_D(x) = p^{-1}(x)\nabla E_D(x)$ with $p > 0$. Our scheme from (10) falls within this class with

$$p^{-1}(x) = \frac{1}{\|\nabla E_D(\phi(x))\| + \sigma}. \quad (20)$$

2.4.3 Condition Analysis of Difference Measures

Finally, we argue that the proposed scheme actually improves the condition of the problem, by showing that the proposed P reduces the value of (16), compared to the original, un-preconditioned problem (i.e. $P = \text{Id}$).

For the shape of a general preconditioned difference measure, we find by applying a preconditioned version of (14) to directions from (19) that

$$\langle P^{-1}\nabla E_D(\varphi), v \rangle = \sum_{x \in \Omega'} \left\langle P^{-1}\nabla E_D, \alpha \frac{\nabla E_D}{\|\nabla E_D\|} \right\rangle \quad (21)$$

$$= \sum_{x \in \Omega'} \alpha \|p^{-1}\nabla E_D\|. \quad (22)$$

In the above we omit the arguments from $\alpha(x)$, $p^{-1}(x)$ and $\nabla E_D(\varphi)(x)$ on the right-hand side for space reasons. We get from (21) to (22) based on (20).

By inserting (22) into (16), we obtain

$$\kappa_P(E_D) = \frac{\max_{\alpha} \sum_{x \in \Omega'} \alpha(x) \|p^{-1}\nabla E_D(\varphi)(x)\|}{\min_{\alpha} \sum_{x \in \Omega'} \alpha(x) \|p^{-1}\nabla E_D(\varphi)(x)\|}. \quad (23)$$

A solution of the constrained maximization and minimization sub-problems from (23), with $1 = \sum \alpha(x)^2$ (cf. (18)), by Lagrange multipliers, reveals that (23) equals

$$\kappa_P(E_D) = \frac{\sqrt{\sum_{x \in \Omega'} \|p^{-1} \nabla E_D(\varphi)(x)\|^2}}{\min_{x \in \Omega'} \|p^{-1} \nabla E_D(\varphi)(x)\|}, \quad (24)$$

Note that (24) is independent on α or overall scaling of ∇E_D . The minimum of (24), and thus the optimal condition, is obtained if the magnitudes of all non-zero point-wise entries are equal: Assuming w.l.o.g. $\|p^{-1} \nabla E_D(\varphi)(x)\| = 1$, $\forall x \in \Omega'$, an increase of $\|p^{-1} \nabla E_D(\varphi)(x)\|$ for any $x \in \Omega'$ will increase the numerator of (24), while the denominator remains unchanged.

Please note that the constant value of $\|p^{-1} \nabla E_D(\varphi)(x)\|$ for all $x \in \Omega'$ corresponds to a gradient field obtained by the proposed preconditioning (10) for $\sigma \rightarrow 0$. On the other hand, for $\sigma \rightarrow \infty$, we arrive at the original energy shape (i.e. $P = \text{Id}$), compare also Fig. 4. This shows that for finite values of σ , the proposed scheme improves the condition of the original problem, and for $\sigma \rightarrow 0$ it actually approximates the optimal case for the examined model. Since we consider the model assumptions to be reasonable, we see the above result as a strong indication that a noticeable improvement of convergence speed can be expected in real applications.

3. Evaluation

3.1. Influence of the σ Parameter

The performance of the proposed method depends on the setting of the σ parameter from Eq. (10). For large values of σ , the preconditioning effect disappears, and the proposed method behaves as SGD. On real data, too low values of σ will enhance noise, which will lead to inrobust performance. To visualize the effect of the σ parameter in a multi-modal setting, we perform a test with MI, on a synthetic 3D data set, with the source image I_S shown in Fig. 4(a). The corresponding target image was set to $I_T = 1 - I_S \circ \phi_{GT}$, and the ground truth ϕ_{GT} was generated by a B-Spline FFD. Fig. 4(b) shows the increasing preconditioning effect for decreasing values of σ . In all our experiments, the σ value is given relative to $\max_x \|\nabla E_{D1}(x)\|$ in first iteration by stating σ' , s.t. $\sigma = \sigma' \max_x \|\nabla E_{D1}(x)\|$. The actual choice of σ depends on the used difference measure, the level of noise in the input images, and the chosen amount of regularization. From our experience, smaller σ values are suited for: (1) SSD rather than MI (probable reason: approximations in implementation of MI introduce “noise”); (2) images from same rather than different modalities; (3) strong regularization, which effectively counteracts noise.

3.2. Tests on Brain Images

We test on simulated 3D MRI brain images (T1,T2,PD) from the BrainWeb project [3], with noise-level of 3%,

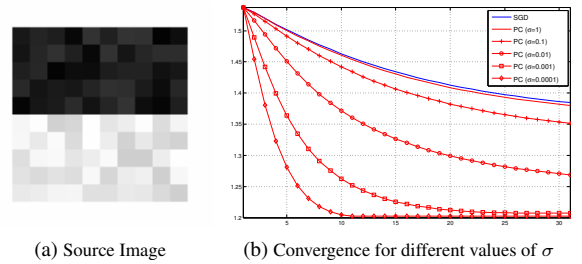


Figure 4: Influence of the σ parameter. Small values of σ result in a stronger preconditioning, and faster convergence. For $\sigma \rightarrow \infty$ the proposed method approximates SGD.

intensity non-uniformity of 20%, and element spacing of 1mm/vx. We perform the 6 possible registrations between the different modalities with different algorithm settings.

The target image is created by applying a ground truth displacement u_{GT} to one of the input images. The quality of the results is assessed by the mean end-point error in the region of interest Ω_M (the head), measured in millimeters.

The ground truth fields are generated in two steps. This two step approach is performed to generate random deformations, which do not have high-frequency displacements in homogeneous areas, which can not be recovered by any method and can overshadow the results of the evaluation. While seemingly complex, such an approach is commonly used, cf. e.g. [9]. First, a combination of cubic B-spline FFDs with different resolutions is created. This field is used to warp the one of images, and then, a registration with the DROP software [4] is performed. The resulting deformation is employed as the ground truth field in the experiments. The second step produces deformation fields which are mostly smooth in homogeneous image regions, and thus reduces the amount of this regularization-related error in the experiments. We employ DROP since it is based on gradient-free optimization and thus can be expected to be less affected by the condition of difference measures.

The implementation of the MI follows [6] with chosen histogram size of 40 and the standard deviation of 0.1 for the Parzen windowing. The tests are performed on two levels to simulate a realistic application setting.

As for σ , the difference weight γ is defined as $\gamma = \gamma' / \max \|\nabla E_{D1}(x)\|$ by setting γ' . This facilitates the use of different energy measures. We perform no explicit step size search. The only modification of the step size τ is performed if the maximal update exceeds a certain given value μ . Then, τ is modified, s.t. $\max \|f(x)\| = \mu$.

Fluid Demons. The first test is performed with fluid demons. For the coarse and fine level, the settings are $\sigma_{fl} = 4, 4[\text{mm}]$, $\gamma' = 1, 2$, $\tau = 1, 1$, $\mu = 1, 1[\text{mm}]$. For the proposed method, we set $\sigma' = 0.1, 0.1$. The results of the test are summarized in Fig. 5.

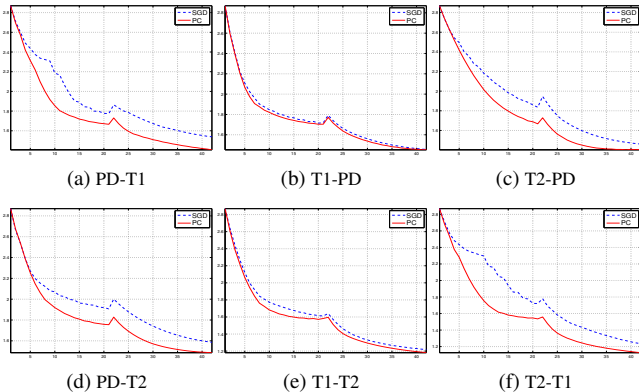


Figure 5: Comparison of proposed method (PC) and steepest descent (SGD) for fluid demons. As in all following figures, we show the convergence plot measuring the development of the mean end-point error in [mm] over iterations.

Elastic Demons #1. The first test with elastic demons is performed with limited update steps, with the limits $\mu = 1, 1$ [mm]. This approach is of particular interest, as several demons approaches require small updates in order to generate diffeomorphisms, cf. e.g. [2, 16], and the proposed method could thus yield a faster convergence for these approaches. The remaining parameters were: $\sigma_{fl} = 2, 2$ [mm], $\sigma_{el} = 1.4, 0.7$ [mm], $\gamma' = 1, 2$, $\tau = 1, 1$. For the proposed method, we set $\sigma' = 0.4, 0.4$. See Fig. 6.

Elastic Demons #2. The steepest descent method can yield significantly better results if larger updates are allowed. In this case, the result is improved since the majority of updates are assigned larger and more meaningful values, however, the maximal updates become too large and lead to local oscillations, which results in divergence.

In this experiment, the proposed method (PC) did not have to be limited, and the maximal occurring updates were always below 3 mm. For SGD, a limit of 10 and 2 mm had to be imposed for the coarse and the fine level, to avoid divergent behavior on several data sets. The remaining joint parameters were: $\sigma_{fl} = 2, 2$ [mm], $\sigma_{el} = 1.4, 0.7$ [mm], $\tau = 1, 1$. For SGD, we had $\gamma' = 2, 1.3$, and for PC we used $\sigma' = 0.4, 0.4$, and $\gamma' = 1.5, 1.3$. See Fig. 7 for the results.

Discussion. The results show that in general, the proposed method requires a smaller number of iterations to reach the same level of error as SGD. We observe this improvement in convergence speed in almost all tests, compared to a carefully tuned SGD method. Please note that we use the same preconditioning parameter $\sigma' = 0.1$ in all fluid tests, and $\sigma' = 0.4$ in all elastic tests.

With respect to the actual computation time, the proposed scheme introduces only a minimal overhead. As an

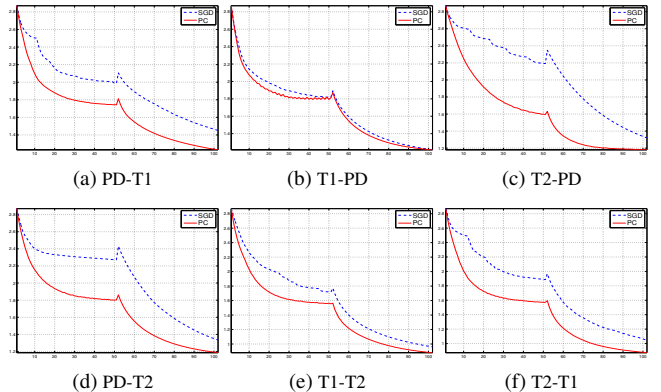


Figure 6: Elastic Demons #1: Point-wise updates are limited to 1mm. This scenario is of particular interest for those demons approaches which require small updates to generate diffeomorphisms, cf. e.g. [2, 16].

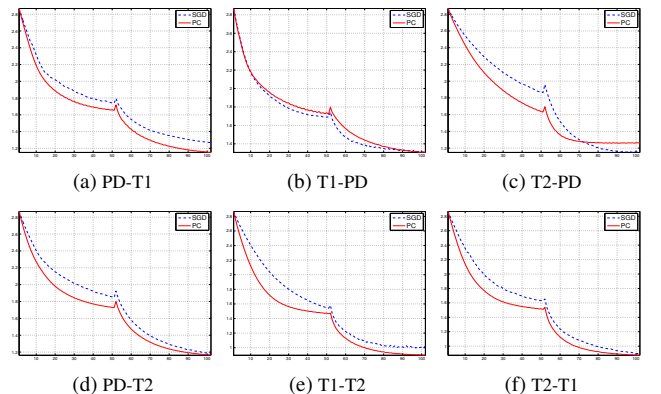


Figure 7: Elastic Demons #2: Point-wise updates are uncontrolled for PC (however, always smaller than 3mm), and for SGD had to be limited 10 and 2mm on the coarser and finer level respectively, to avoid divergence.

example, for the fluid demons test (cf. Fig 5), SGD takes 114 sec, compared to 116 sec for the PC version (C++ implementation, CPU: Intel®Core™2 Duo P8700 2.53GHz, RAM: 4GB). For elastic demons, which feature an additional costly smoothing step, the difference in runtime is even less prominent.

The effect of the preconditioning depends mostly on the distribution of gradient magnitudes and is more prominent for images with multiple distinct clusters of gradient magnitudes (CGM). For example, in Fig. 4, I_S has two CGMs: one with weaker gradients between most blocks, and one with stronger gradients resulting from the high-contrast horizontal line. Without the “stronger” CGM, the preconditioning influence would be less prominent. Consequently, the effect for real data also depends on its CGMs. We in-

cluded all results to show that in some cases the effect can be limited - e.g. for the combination of T1 and PD. Our statement is that our simple technique performs at least as well as SGD in general, and better in most cases, without requiring tedious tuning.

The results for SGD are better in Fig. 7 than in Fig. 6 due to an extreme tuning of the step sizes. These are on the limit of the robust behavior for SGD and reaching such performance requires manual tuning for each data set. Our approach reaches its result with more conservative settings.

A further interesting observation is that the performance of our method seems to be less dependent on attribution of image modality to source or target: e.g. the results of PD-T1 and T1-PD should lead to same errors (Fig. 6: a,b). This is not too surprising, since for SGD, the process depends heavily on I_S through $\nabla E_D(x) = \omega_D(x)\nabla I_S(x)$ (cf. [2]), and thus on the choice of I_S . In our approach this dependence is strongly reduced by the preconditioning.

4. Summary

In this paper, we present a simple and theoretically justified preconditioning scheme for arbitrary difference measures in deformable registration. This approach is of particular interest for cases where other standard optimization methods become too complex, or are not applicable. Important examples of such scenarios are multi-modal registration problems with statistical difference measures (e.g. MI) and high-dimensional deformation models, and especially the multi-modal demons registration.

Our theoretical analysis demonstrates that the proposed scheme improves the condition of the problem, and that it actually even approximates the optimal case for the examined model. More specifically, we show that 1) point-wise multiplication of $\nabla E_D(x)$ can be seen as an important class of preconditioners for deformable registration, and 2) that the strategy to normalize the point-wise lengths $\|p^{-1}(x)\nabla E_D(x)\|$ to the same value is optimal for this class. Our actual strategy is only an approximation to this optimal case, since it performs a damped normalization, which makes the process robust to noise.

Due to its simplicity, our scheme is not only easy to implement for any difference measure, but it also has only a negligible computational overhead, such that the improvement in convergence speed is directly transferred to an improvement of runtime. The fact that the proposed method does not require any information from previous iteration steps makes it particularly suited for demons registration - this is in contrast to NL-CG or L-BFGS.

The performed experiments show an improvement of the convergence speed compared to SGD, which is currently the standard approach for multi-modal demons.

References

- [1] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache. Iconic feature based nonrigid registration: the pasha algorithm. *Comp. Vis. and Im. Underst. (CVIU)*, 2003. 3
- [2] C. Chef d'hotel, G. Hermosillo, and O. Faugeras. Flows of diffeomorphisms for multimodal image registration. In *Intl. Symp. on Biomedical Imaging (ISBI)*, 2002. 2, 3, 5, 7, 8
- [3] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, G. B. Pike, and A. C. Evans. Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage*, 5:425, 1997. 6
- [4] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense Image Registration through MRFs and Efficient Linear Programming. *Med. Image Anal.*, 2008. 6
- [5] A. Guimond, A. Roche, N. Ayache, and J. Meunier. Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. *IEEE Transactions on Medical Imaging (TMI)*, 20(1):58–69, 2002. 2
- [6] G. Hermosillo, C. Chef d'Hotel, and O. Faugeras. Variational methods for multimodal image matching. *IJCV*, 2002. 6
- [7] S. Klein, M. Staring, P. Andersson, and J. Pluim. Pre-conditioned stochastic gradient descent optimisation for monomodal image registration. In *MICCAI*, 2011. 3
- [8] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Im.*, 1997. 1
- [9] M. Modat, T. Vercauteren, G. R. Ridgway, D. J. Hawkes, N. C. Fox, and S. Ourselin. Diffeomorphic demons using normalized mutual information, evaluation on multimodal brain mr images. In *SPIE Med. Imag.*, 2010. 1, 2, 3, 6
- [10] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. SIAM, 2009. 3
- [11] J. Nocedal and S. Wright. *Numerical optimization*. Springer, 2000. 1, 2, 3
- [12] X. Pennec, P. Cachier, and N. Ayache. Understanding the demon's algorithm: 3d non-rigid registration by gradient descent. In *MICCAI*, 1999. 3
- [13] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. *Proc. MICCAI*, 98:1115–1124, 1998. 1
- [14] J. Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998. 3
- [15] A. Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *IJCV*, 28(3):213–221, 1998. 3
- [16] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. In *MICCAI*, 2008. 7
- [17] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 2009. 3
- [18] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996. 1
- [19] D. Zikic, M. Baust, A. Kamen, and N. Navab. Generalization of deformable registration in riemannian sobolev spaces. In *MICCAI*, 2010. 2, 4