

Spatially-Varying Metric Learning for Diffeomorphic Image Registration: A Variational Framework^{*}

François-Xavier Vialard¹ and Laurent Risser²

¹ Université Paris Dauphine, CEREMADE (UMR 7534), France

² CNRS, Institut de Mathématiques de Toulouse (UMR 5219), France

Abstract. This paper introduces a variational strategy to learn spatially-varying metrics on large groups of images, in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework. Spatially-varying metrics we learn not only favor local deformations but also correlated deformations in different image regions and in different directions. In addition, metric parameters can be efficiently estimated using a gradient descent method. We first describe the general strategy and then show how to use it on 3D medical images with reasonable computational resources. Our method is assessed on the 3D brain images of the LPBA40 dataset. Results are compared with ANTS-SyN and LDDMM with spatially-homogeneous metrics.

1 Introduction

Diffeomorphic image registration often consists in minimizing an objective function which contains a similarity term and a regularization term. In this work, we focus on the regularization term which can be seen as a prior when approximating the biomechanical properties of the registered structures (*e.g.* deformation smoothness, magnitude, ...). In the context of *Large Deformation Diffeomorphic Metric Mapping* (LDDMM) different works have addressed this question. Sum of kernels strategies were proposed in [7,13] to account for multi-scale effects and therefore obtain plausible deformations while still preserving good matchings. Other approaches do not learn the metric but a distribution on initial momenta which defines the optimal deformations between a template and a learning set of images [12]. Such approaches however require to first choose a metric and the momenta distribution or their PCAs strongly depend on this choice.

To overcome these limitations, the Bayesian approach is often the method of choice, allowing to learn parameters distribution [14]. Full Bayesian approaches require the use of stochastic optimization methods which are slow. Sometimes approximations such as variational Bayes are preferred. For small deformations,

^{*} This work was supported by the ANR DEMOS grant, the *Chaire Havas-Dauphine Économie des nouvelles données*, and the AO1 grant MALAC3D from Université Paul Sabatier. The authors also thank the reviewers for their constructive and insightful comments.

[1] estimates the covariance matrix of the deformations parametrized by a set of points. The approach has been extended to large deformations in [4] by reducing the metric learning to a set of control points. In a non-diffeomorphic context, variational Bayes methods were also introduced in [11,10] to perform image registration with the automatic tuning of global or spatially-varying regularisation parameters. The motivation of this method is however to be flexible and not to learn optimal metrics on large groups of images, which differs from our goal.

Spatially-varying metrics have recently been introduced in LDDMM by changing the Eulerian point of view on the regularization strategy to a Lagrangian point of view [8]. Although this strategy is appealing, its practical use is limited, as the tuning of a large amount of metric parameters is made manually. To overcome this issue, we propose in this paper a fully variational approach to estimate a kernel matrix M , which parametrizes the regularization term.

We present the variational approach to learn M in section 2. Strategies to reduce the problem dimensionality are developed in section 3. Results are finally given in section 4.

2 Mathematical Model

2.1 LDDMM Registration

Our model is based on the LDDMM framework. Let $(I_n)_{n=1,\dots,N}$ be a population of N images and T be a given template. Registering the template T onto the image I_n consists in minimizing:

$$\mathcal{J}_{I_n}(v, K) = \frac{1}{2} \int_0^1 \|v(t)\|_V^2 dt + E(\phi(1) \cdot T, I_n), \quad (1)$$

where the path $\phi(t)$, $t \in [0, 1]$ is encoded by the velocity field $v(t)$, $t \in [0, 1]$: $\phi(0) = Id$ and $\partial_t \phi(t) = v(t) \circ \phi(t)$. Importantly, the optimal diffeomorphism depends on a smoothing kernel K which defines the metric V . In LDDMM, K is usually translation and rotation invariant so that the Fourier transform can be used to write the metric as $\|\mathbf{v}_t\|_V^2 = \langle \mathcal{F}(\mathbf{v}_t) \mathcal{F}(K)^{-1}, \mathcal{F}(\mathbf{v}_t) \rangle_{L^2}$, where $\mathcal{F}(\cdot)$ is the Fourier transform and $\langle \cdot, \cdot \rangle_{L^2}$ is the L^2 inner product [3]. Following [3], the energy can be minimized using a gradient descent with $\nabla_{\mathbf{v}} E_t = \mathbf{v}_t - K \star P_t$, where $P_t = \text{Det } J_{\phi_{t,1}} \nabla T_t(T_t - I_t)$ is the momentum at time t and \star denotes the convolution operator. Images T_t and I_t are also T and I_n transported at time t by $\phi(t)$: $T_t = \phi(t) \cdot T$ and $I_t = \phi(t) \cdot \phi^{-1}(1) \cdot I_n$.

2.2 Spatially Varying Metrics

A mathematical interpretation of spatially-varying metrics in LDDMM has recently been given by [8], opening the opportunity to design metrics adapted to the different structures contained in the template T . Based on this interpretation, we design a set of kernels expressing spatially-varying metrics. More

specifically, we use (symmetric) positive definite matrices M as a parametrization of this set of kernels. In order to ensure smoothness of the deformations, any kernel of this set has to satisfy the constraint that the Hilbert space of vector fields is embedded in the Banach space of C^1 vector fields. To enforce this constraint, we propose the following parametrization,

$$\mathcal{K} = \{\hat{K}M\hat{K} \mid M \text{ SDP operator on } L^2(\mathbb{R}^d, \mathbb{R}^d)\}, \quad (2)$$

where \hat{K} is a spatially-homogeneous smoothing kernel (typically Gaussian). Instead of using K as in section 2.1, \mathcal{K} smoothes the vector field P_t as follows:

1. P_t is first convoluted with \hat{K} : $\Gamma_t = \hat{K} \star P_t$.
2. Matrix M multiplies the values of the 3D vector field Γ_t as follows: We suppose that Γ_t has a size (N_x, N_y, N_z) and denote $\Gamma_t(x_i, y_i, z_i, d_i)$ the value of Γ_t at point (x_i, y_i, z_i) and in direction d_i . Γ_t is first reshaped as a vector, so that $\Gamma_t(x_i, y_i, z_i, d_i)$ is located at the index $\nu_i = x_i + y_i N_x + z_i N_x N_y + d_i N_x N_y N_z$ of the vector. The vector is first multiplied with M and then reshaped as a 3D vector field.
3. The vector field resulting from step 2 is finally convoluted with \hat{K} .

By construction $M(\nu_i, \nu_j)$ therefore correlates the velocities of Γ_t at points (x_i, y_i, z_i) and (x_j, y_j, z_j) and in directions d_i and d_j , respectively. Remark that if $M = Id$ and \hat{K} is a Gaussian kernel of standard deviation σ , then \mathcal{K} is a Gaussian kernel of standard deviation $\sqrt{2}\sigma$. If M has non-null and heterogeneous values on its diagonal only, the metric will *favor* the deformations at specific locations and in specific directions. More interestingly, if M contains non-null terms outside of its diagonal, it will *favor* deformations in correlated locations and/or directions. Of course, this correlation can be non-local.

2.3 Learning Optimal Metrics

To shorten the notations, we use $\mathcal{J}_{I_n}(v, M)$ instead of $\mathcal{J}_{I_n}(v, \hat{K}M\hat{K})$. The variational model consists in minimizing the functional, with β a positive real:

$$\mathcal{F}(M) = \frac{\beta}{2} d_{S^{++}}^2(M, Id) + \frac{1}{N} \sum_{n=1}^N \min_v \mathcal{J}_{I_n}(v, M), \quad (3)$$

The first term is a regularizer of the kernel parameters so that that the minimization problem is well posed. Here, it favors parametrizations of M close to the identity matrix but other a priori correlation matrix could be used. The term $d_{S^{++}}^2(Id, M)$ can be chosen as the squared distance on the space of positive definite matrices given by $\|\log(M)\|^2$. Here again, other choices of regularizations could have been used such as the log-determinant divergence. We simply remark that this distance comes from a Riemannian metric denoted by g on S^{++} which makes it complete [6,2]: namely we consider S^{++} endowed with the inner product at S given by $tr(S^{-1}dSS^{-1}dS)$ where tr is the standard trace operator. The variational problem would have been ill-posed if the standard L^2

metric (Frobenius norm) had been used. Note that the energy term in \mathcal{J}_{I_n} is linear in M^{-1} . A direct calculation shows that the gradient of \mathcal{F} with respect to the metric g denoted by $\nabla_g \mathcal{F}$ is

$$\nabla_g \mathcal{F}(M) = \beta M \log(M) - \frac{1}{N} \sum_{n=1}^N \int_0^1 (M \hat{K} \star P_n(t)) \otimes (M \hat{K} \star P_n(t)) dt, \quad (4)$$

where $A \otimes B$ is the tensor product and is defined by $A \otimes B(f) = \langle B, f \rangle_{L^2} A$ for $A, B \in L^2(\mathbb{R}^d, \mathbb{R}^d)$. Momenta $P_n(t)$ are obtained at convergence of the diffeomorphic matching algorithm on $\mathcal{J}_{I_n}(v, M)$. Note that this tensor product is performed in the space of vectorized vector fields defined in step 2 of the algorithm section 2.2. We now develop a brief proof of how Eq. (4) is obtained: The second term is a minimization over v of each term independently and therefore can be rewritten as a function of M : $\sum_{n=1}^N \min_v \mathcal{J}_{I_n}(v_n(M), M)$, so that at convergence we have $\partial_1 \mathcal{J}_{I_n}(v_n(M), M) = 0$ for each $n = 1, \dots, N$, where ∂_1 is the partial derivative w.r.t. v . Thus, $\nabla_{L^2} \mathcal{F}(M)$ equals $\frac{\beta}{2} \nabla_{L^2} d_{S^{++}}^2(M, Id) + \frac{1}{N} \sum_{n=1}^N \min_v \partial_M \mathcal{J}_{I_n}(v, M)$. Using the linearity w.r.t. M^{-1} and the chain rule formula, we obtain $\partial_2 \mathcal{J}_{I_n}(v, M) = - \int_0^1 (\hat{K} \star P_n(t)) \otimes (\hat{K} \star P_n(t)) dt$. We then derive the Riemannian gradient using the formula $\nabla_g f(S) = S \nabla_{L^2} f(S) S$. The differentiation of the first term is standard on a Riemannian manifold and its gradient is given by the tangent vector of the geodesic between Id and M evaluated at M . Then the geodesic starting at identity and ending at M is given by $t \rightarrow e^{t \log(M)}$ (see [6,2]). As a consequence, $\nabla_g \frac{1}{2} d_{S^{++}}^2(M, Id) = M \log(M)$.

After initializing M to identity, the algorithm is a simple gradient descent which iterates: (1) Register T on the images I_n , $n = 1, \dots, N$ to obtain the momenta $P_n(t)$; (2) Compute the gradient $\nabla_g \mathcal{F}(M)$ using formula (4); (3) Update $M := M - \varepsilon \nabla_g \mathcal{F}(M)$, where ε is the chosen step length.

3 Reducing the Problem Dimension

In this section, we propose two straightforward solutions to make the learning problem of section 2 usable on most computers when treating 3D medical images. Other dimensionality reduction methods could be applied as discussed in section 5. Considering that the registered 3D images have N voxels, matrix M has a size $3N \times 3N$ which may be a huge amount of information to store.

3.1 Diagonal Matrix M

A straightforward solution to reduce the problem's dimension is to constrain M to have diagonal terms only. As explained in section 2, matrix M will only favor deformations in specific locations and directions. The amount of parameters to store and to estimate is however $3N$ instead of $3N \times 3N$, which makes it usable in 3D medical imaging with reasonable computational resources. Moreover, Eq. (4) becomes numerically obvious to compute as the logarithm of $3N$ scalars is computed instead of the logarithm of a $3N \times 3N$ matrix.

3.2 Basis Projection

Another solution, which allows to model long distance and inter-axes correlations, is to project Γ_t on a 3D basis. In this work, we use a 3D B-spline basis, with elements denoted by ψ_l , $l \in \{1, \dots, L\}$. Each element has its origin at point $p_l = (p_l^x, p_l^y, p_l^z)$ and we suppose the p_l sampled on a spatially homogeneous grid with a larger step size than the image resolution. We also associate the vector $\alpha_l = (\alpha_l^x, \alpha_l^y, \alpha_l^z)$ to element l of the basis. We denote $\hat{\Gamma} = (\alpha_1^x, \dots, \alpha_L^x, \alpha_1^y, \dots, \alpha_L^y, \alpha_1^z, \dots, \alpha_L^z)$ the vector of size $d = 3L$ which will be used to learn M . A dense vector field v can be constructed from $\hat{\Gamma}$ using:

$$v(p) = \sum_{l=1}^L \alpha_l \psi(p - p_l), p \in \Omega \quad (5)$$

Projecting Γ_t on the basis only would induce a loss of information related to deformations at a finer scale than the grid step size. To address this issue, we perform the orthogonal projection of Γ_t on a closed subspace, $\Pi : L^2(\mathbb{R}^d, \mathbb{R}^d) \mapsto W$, to learn parameters only on this subspace:

$$\mathcal{X} = \{\hat{K} M \Pi \hat{K} + \hat{K} (Id - \Pi) \hat{K} \mid M \text{ SDP operator on } L^2(\mathbb{R}^d, \mathbb{R}^d)\}. \quad (6)$$

Let us interpret how Eq. (6) reduces the problem dimensionality. Instead of step 2 in the pseudo-algorithm of section 2.2: (1) We compute $\hat{\Gamma}_t$ ($\hat{\Gamma}$ at time t) by projecting Γ_t on the basis. (2) A vector field Q_1 is constructed using Eq. (5) with $\hat{\Gamma}_t$ and we define the residual vector field $R_t = \Gamma_t - Q_1$. (3) Another vector field Q_2 is constructed using Eq. (5) with $M \hat{\Gamma}_t$, the product of M and $\hat{\Gamma}_t$. (4) The vector field $Q_2 + R_t$ is finally the result of this modified step 2. The interest of this strategy is twofold: Spatially-varying metric can be learnt using the information projected on $\hat{\Gamma}_t$. Residue R_t which is related to deformations at a finer scale than the grid step size also ensures that all information, and not only the one projected on $\hat{\Gamma}_t$, is used to register the images. In addition to these modified steps, the gradient of Eq. (4) is now computed using the projected values of Γ_t and not the values of Γ_t directly.

4 Results

We assessed our method on the 40 subjects of the LONI Probabilistic Brain Atlas (LPBA40) [9]. All 3D images were affinely aligned to subject 5 using ANTS¹ and then resampled to a resolution of 2 mm. We then learnt different matrices M using the strategies of sections 3.1 and 3.2: T was the probability tissue map (TM) of subject 5, the I_n were the TM of subjects 1 to 30 (except 5) and \hat{K} was a Gaussian kernel of width $\sigma = 10$. For the strategy of section 3.2, we used a regular 3D grid sampled with a step size of 20mm. The parameter ε was semi-empirically tuned so that ε equals 0.01 divided by the

¹ <http://stnava.github.io/ANTS/>

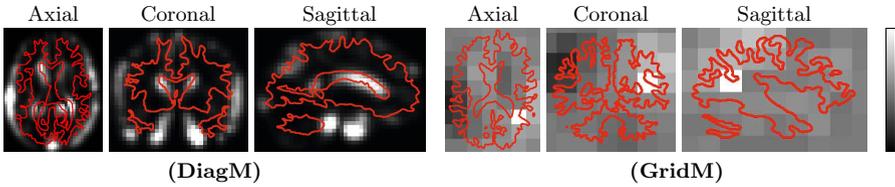


Fig. 1. Values out of M after the two learning steps of section 4. The values are represented at their corresponding location in the template image T . **(DiagM):** Values $M(j, j)$ for $j \in [1, \dots, N]$. Color bar ranges from 1 (black) to 1.04 (white). **(GridM):** $M(i, j)$ for a fixed i and $j \in [1, \dots, L]$. White point corresponds to $i = j$ and has an intensity of 1.03. Color bar ranges from -0.05 (black) to 0.05 (white) for other points. Red curves represent the boundary between white and grey matter in T .

maximum of $\left| \sum_{n=1}^N \int_0^1 (M \hat{K} \star P_n(t)) \otimes (M \hat{K} \star P_n(t)) dt \right|$ at the first iteration of the algorithm. Parameter β is also equal to $0.025 \hat{\beta}$ divided by the maximum of $|\varepsilon M \log(M)|$ at the second algorithm iteration. We used four values of $\hat{\beta}$ to test different regularization levels of M .

We can see in Fig. 1(DiagM) the location of the deformations favored by the diagonal matrix M learnt with $\hat{\beta} = 1$ on the x-axis. For M learnt on a grid with $\hat{\beta} = 1$, Fig. 1(GridM) also indicates how the motion on the x-axis at point i is empirically correlated with the motion at other locations j on the same axis. Note that this information is computed for all grid nodes and not only i .

We then compared different registration strategies by registering the TM of subject 5 on those of subjects 31 to 40: **(DiagM)** and **(GridM)** LDDMM registration with the kernels learnt using the strategies of sections 3.1 and 3.2, respectively. **(K_{ref})** Same as (DiagM) or (GridM) with M equals identity (nothing learnt). This is equivalent to LDDMM registration using a Gaussian kernel with $\sigma = 10\sqrt{2}$. **(K_σ)** To compare (GridM) with results obtained using stronger spatially-homogeneous regularizations than in **(K_{ref})**, we also performed LDDMM registration with different values of σ . Finally, to compare our results with state of the art strategies we finally performed ANTS-SyN registration with the regularisation parameters of [5] **(SyN)** and LDDMM registration with the multiscale strategy of [7]² with σ sampled between 2 and 20mm **(K_{fine})**. After registration, we computed the target overlaps (TO) between the segmented cortical regions (given in LPBA40) to measure the matching quality and the determinant of the Jacobians (DetJ) to quantify the deformation smoothness. We also performed Kruskal-Wallis rank tests, where $p \leq 0.05$ was chosen as significance threshold, to compare different strategies.

As shown Table 1, we obtained more accurate TO using (SyN) and (K_{fine}) than other methods. This is because other methods derive from (K_{ref}) and are then constrained to register the images at a relatively large scale, as shown by the DetJ of Table 1. Note that the relatively large kernel of (K_{ref}) was chosen to

² <http://sourceforge.net/projects/utilzreg/>

emphasize the effect of M with the grid of (GridM). Learning M on finer grids would allow to learn multiscale kernels derived from (K_{fine}) for instance. On 3D brain images this would however either require more advanced dimensionality reduction techniques than those of section 3 or a very large amount of memory. Comparing (K_{ref}) with the methods using M shows the effect of our strategy, as (K_{ref}) is strictly equivalent to the other methods if M is identity: Remark first that, in our tests, no significant difference was found between (K_{ref}) and the different (DiagM). We also did not find significant differences between (K_{ref}) and (GridM) close to identity (eg: (GridM1) in Table 1) or (K_{σ}) with σ close to $10\sqrt{2}$ (eg: (K_{20}) in Table 1). No significant difference is also found between (GridM1) and (K_{20}). Significantly different TO and maximum DetJ to those of (K_{ref}) were found using stronger spatial regularization, e.g. (GridM2) and (K_{30}) in Table 1. Comparing these two strategies leads to our key result: For a similar TO, (GridM2) has significantly lower DetJ $_{Max}$ and DetJ $_{Std}$ than (K_{30}). Remark finally that the average DetJ $_{Std}$ of (K_{30}) is higher than the one of (K_{20}) although the one of (GridM2) is lower than the one of (GridM1). The spatially-varying kernel therefore seems to take advantage of the learnt information with non-local correlations to estimate smoother meaningful deformations.

Table 1. Average results obtained on the 3D brain images of the LPBA40 dataset

	No Reg	SyN	K_{fine}	K_{ref}	DiagM	GridM1	GridM2	K_{20}	K_{30}
TO	0.665	0.750	0.732	0.712	0.711	0.710	0.704	0.710	0.704
DetJ $_{Max}$	1	3.17	4.65	1.66	1.66	1.61	1.41	1.62	1.50
DetJ $_{Min}$	1	0.047	0.46	0.67	0.68	0.70	0.67	0.73	0.66
DetJ $_{Std}$	0	0.17	0.11	0.063	0.062	0.059	0.049	0.056	0.063

5 Discussion

In this paper, we addressed the problem of regularization in diffeomorphic registration using a new variational approach to learn spatially-varying metrics in the LDDMM setting. We parametrized the space of spatially-varying metrics with positive definite matrices M and used the logarithm norm on this space as a Tychonov regularizer to make the variational problem well posed. We also gave a semi-analytical expression of the minimized energy gradients relative to M , as well as two strategies to keep the problem’s dimensionality reasonable, making our learning strategy doable on standard computers. Note that the proposed method has a natural *maximum a posteriori* interpretation and thus, Bayesian methods could be developed in coherence with this variational model. We obtained encouraging results showing that the spatially-varying metrics we learnt allowed to register 3D brain images with smoother deformations than by using a spatially-homogeneous metric, for similar structure overlaps.

Future work will focus on dimensionality reduction methods to learn M . Our goal is to learn multiscale smoothing kernels, making our strategy more pertinent

on large 3D medical images. An exciting perspective of this work would also be the statistical analysis of the spatially-varying metric parameters defined here.

References

1. Allasonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. *J. R. Statist. Soc. B* 69(1), 3–29 (2007)
2. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* 29(1), 328–347 (2007)
3. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61(2), 139–157 (2005)
4. Durrleman, S., Allasonnière, S., Joshi, S.: Sparse adaptive parameterization of variability in image ensembles. *IJCV* 101(1), 161–183 (2013)
5. Klein, A., Ghosh, S.S., Avants, B.B., Yeo, B.T.T., Fischl, B., Ardekani, B.A., Gee, J.C., Mann, J.J., Parsey, R.V.: Evaluation of volume-based and surface-based brain image registration methods. *NeuroImage* 51(1), 214–220 (2010)
6. Moakher, M., Zerai, M.: The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data. *JMIV* 40(2), 171–187 (2011)
7. Risser, L., Vialard, F.X., Wolz, R., Murgasova, M., Holm, D.D., Rueckert, D.: Simultaneous Multi-scale Registration Using Large Deformation Diffeomorphic Metric Mapping. *IEEE Transactions on Medical Imaging* 30(10), 1746–1759 (2011)
8. Schmah, T., Risser, L., Vialard, F.X.: Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part I. LNCS*, vol. 8149, pp. 203–210. Springer, Heidelberg (2013)
9. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* 39, 1064–1080 (2008)
10. Simpson, I., Schnabel, J., Groves, A., Andersson, J., Woolrich, M.: Probabilistic inference of regularisation in non-rigid registration. *NeuroImage* 59(3) (2012)
11. Simpson, I.J.A., Woolrich, M.W., Cardoso, M.J., Cash, D.M., Modat, M., Schnabel, J.A., Ourselin, S.: A bayesian approach for spatially adaptive regularisation in non-rigid registration. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part II. LNCS*, vol. 8150, pp. 10–18. Springer, Heidelberg (2013)
12. Singh, N., Fletcher, P.T., Preston, J.S., King, R.D., Marron, J., Weiner, M.W., Joshi, S.: Quantifying anatomical shape variations in neurological disorders. *Medical Image Analysis* (2014)
13. Sommer, S., Lauze, F., Nielsen, M., Pennec, X.: Kernel bundle EPDiff: Evolution equations for multi-scale diffeomorphic image registration. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) *SSVM 2011. LNCS*, vol. 6667, pp. 677–688. Springer, Heidelberg (2012)
14. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) *IPMI 2013. LNCS*, vol. 7917, pp. 37–48. Springer, Heidelberg (2013)