

Image Registration and Predictive Modeling: Learning the Metric on the Space of Diffeomorphisms

Ayagoz Mussabayeva^{1,3}, Alexey Kroshnin^{1,3}, Anvar Kurmukov^{1,3}, Yulia Dodonova¹, Li Shen⁴, Shan Cong⁵, Lei Wang⁶, and Boris A. Gutman^{1,2}

- ¹ The Institute for Information Transmission Problems, Moscow, Russia
² Department of Biomedical Engineering, Illinois Institute of Technology, Chicago, IL, USA
³ Higher School of Economics, Moscow, Russia
⁴ Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA
⁵ Indiana University, Indianapolis, IN, USA
⁶ Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Abstract. We present a method for metric optimization in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework, by treating the induced Riemannian metric on the space of diffeomorphisms as a kernel in a machine learning context. For simplicity, we choose the kernel Fischer Linear Discriminant Analysis (KLDA) as the framework. Optimizing the kernel parameters in an Expectation-Maximization framework, we define model fidelity via the hinge loss of the decision function. The resulting algorithm optimizes the parameters of the LDDMM norm-inducing differential operator as a solution to a group-wise registration and classification problem. In practice, this may lead to a biology-aware registration, focusing its attention on the predictive task at hand such as identifying the effects of disease. We first tested our algorithm on a synthetic dataset, showing that our parameter selection improves registration quality and classification accuracy. We then tested the algorithm on 3D subcortical shapes from the Schizophrenia cohort Schizconnect. Our Schizophrenia-Control predictive model showed significant improvement in ROC AUC compared to baseline parameters.

Keywords: image registration, machine learning, subcortical shape, expectation maximization, metric learning, LDDMM

1 Introduction

Image registration, and more generally geometric alignment underlies a large number of analyses in medical imaging, particularly neuroimaging. As one of the mainstays of medical image analysis, the problem has been addressed extensively over the last 2+ decades, with several flavors of robust algorithms [1]. A

number of registration approaches develop an explicit metric space comprised of the geometric objects of interest — anatomical shapes, diffusion tensors, images, etc. Prominent among these is the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [2]. Instead of treating images as objects of interest directly, LDDMM builds a space of image-matching diffeomorphisms using a Riemannian metric on velocity fields. This metric is induced by a differential operator which at once controls the nature of the metric space and regularizes the registration.

The structure of such a space — a manifold of smooth mappings with well-defined geodesics — enables generalizations of several standard statistical analysis methods. These methods adapted to the Riemannian setting has been repeatedly shown to improve their sensitivity and ability to discern population dynamics when compared to projecting the data onto Euclidean domains. Works in this area include computation of the geometric median and metric optimization for robust atlas estimation [3,4], time series geodesic regression [5], and principal geodesic analysis [6].

With the exception of [4], in the works above the metric is assumed to be fixed. Further, the metricity and Riemannian inner product with which the LDDMM space is endowed has not been used explicitly in predictive modeling up to now. In this work, we strive for two complementary aims: (1) to exploit the Riemannian metric on registration-defining velocities as a kernel in a classification task and (2) to optimize the metric to improve classification. We follow an Expectation-Maximization (EM) approach similar to [4], alternating between minimizing image misalignment for kernel estimation, and optimizing model quality over the kernel parameters. In this work, we choose the kernel Fischer linear discriminant classifier for simplicity, though other predictive models are admissible in our framework as well. It is our hope that by explicit tuning the diffeomorphism metric to questions of biological interest, the carefully crafted manifold properties of LDDMM will gain greater practical utility.

Our experiments consist of synthetic 2-dimensional shape classification, as well as classifying hippocampal shapes extracted from brain MRI of the Schizconnect schizophrenia study. In both cases, the classification accuracy and ROC area under the curve (AUC) improved significantly compared to default baseline kernel parameters.

2 Methods

2.1 Metric on diffeomorphisms

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) was first introduced in [2]. The goal of the registration is to compute a diffeomorphism $\phi: \Omega \rightarrow \Omega$, where the Ω is the image domain. The diffeomorphism ϕ is generated by the flow of a time-dependent velocity vector field v , defined as follows:

$$\begin{aligned} \frac{\partial \phi(t, x)}{\partial t} &= v(t, \phi(t, x)), \\ \phi(0, x) &= id, \end{aligned} \tag{1}$$

where id is the identity transformation: $id(x) = x, \forall x \in \Omega$. This equation gives a path $\phi_t: \Omega \rightarrow \Omega, t \in [0, 1]$, in the space of diffeomorphisms. Estimation of the optimal diffeomorphism via the basic variational problem in the space of smooth velocity fields V on Ω takes the following form, constrained by 1:

$$v^* = \underset{v}{\operatorname{argmin}} \left(\int_0^1 \|v_t\|_L^2 dt + \frac{1}{\sigma^2} \|I_0 \circ \phi - I_1\|^2 \right). \quad (2)$$

The required smoothness is enforced by defining the norm $\|\cdot\|_L$ on the space V of smooth velocity vector fields through a Riemannian metric L .

The Riemannian metric L should be naturally defined by the geometric structure of the domain. The inner product $\|v\|_L^2 = \langle Lv, v \rangle$ can also be thought of as a metric between images, i.e. the minimal diffeomorphism required to transform the appearance of I_0 to be as similar as possible to I_1 . Since the diffeomorphism space is a Lie Group with respect to eq.1, the Riemannian metric defined suggests a right-invariance property. The original LDDMM work [2] defines L as a smooth differential self-adjoint operator $L = (\alpha\Delta + \beta E)$, where E is identity operator. Here, we choose to use an L based on the biharmonic operator $L = (\alpha\Delta + \beta E)^2$, as e.g. in [7]. The parameters (α, β) correspond to convexity and normalization terms, respectively. These parameters significantly affect the quality of the registration. It is not obvious how to select α and β , though they effectively define the geometric structure of the primal domain. Indeed, these are the parameters we optimize in our EM scheme below.

2.2 Predictive Model

Consider a standard binary classification problem: given a sample $(x_i, y(x_i))_{i=1}^n$, where $y(x_i) \in \{1, -1\}$ is a class label, find a classification function \hat{y} approximating the true one y . One of the standard linear techniques in statistical data analysis is the Fisher’s linear discriminant analysis (LDA). The kernel Fisher Discriminant Analysis (KLDA) introduced in [8] is a generalization of the classical LDA. There are several approaches to derive more general class separability criteria. KLDA derives a linear classification in the embedding feature space (RKHS[9]) induced by a kernel k , what corresponds to a non-linear decision function in the original, or “input” space. The main idea of LDA is to find a one-dimensional projection w in the feature space that maximizes the between-class variance while minimizing the within-class variance. KLDA seeks an analogous projection in the embedding space, where means (M_z) and covariance matrices (Σ_z) for each class $z \in \{-1, 1\}$ are computed. The (K)LDA cost function takes the following quadratic rational form:

$$J(w) = \frac{w^T(M_1 - M_{-1})(M_1 - M_{-1})^T w}{w^T(\Sigma_{-1} + \Sigma_1)w} = \frac{w^T M w}{w^T N w}, \quad (3)$$

where

$$(M_z)_i = \frac{1}{n_z} \sum_{x_\ell: y(x_\ell)=z} k(x_i, x_\ell),$$

$$(\Sigma_z)_{i,j} = \frac{1}{n_z} \sum_{x_\ell: y(x_\ell)=z} k(x_i, x_\ell)k(x_j, x_\ell) - (M_z)_i(M_z)_j.$$

Here n_z is a number of objects from class z in the sample.

The solution of the problem $J(w) \rightarrow \min$ is known to be $\hat{w} = N^{-1}(M_1 - M_{-1})$. The decision function for a new observation x is based on the projected distance to the training sample means, $w^T(M_z - x)$. The M-step in an EM formulation requires a differentiable measure of model quality which in our case is the accuracy of classification. The more common approach is to formulate a probabilistic model which leads to a log-likelihood optimization. Such an approach is used e.g. in [4]. In our case, this can be done by modeling the classifier’s output with a parametric distribution.

However, we found that such a formulation using the sigmoid distribution function leads to an unstable solution. Instead, we propose to use the hinge loss defined for KLDA as

$$h(x', z) = \max\{0, 1 - z y(x')\}$$

$$y(x') = \sum_{i=1}^n w_i \left(k(x_i, x') - \frac{(M_1)_i + (M_{-1})_i}{2} \right), \quad (4)$$

where $z \in \{-1, 1\}$ is a true label for the new observation x' and $k(x_i, x')$ is the inner product (i.e. the kernel) between x' and the training observation x_i . While both hinge loss and log-likelihood formulations eventually lead to some locally optimal solutions on simple problems, such as our synthetic dataset, the former exhibits greater stability. For the hippocampal data, only hinge loss minimization leads to a stable solution.

2.3 Learning the diffeomorphic metric

The main goal of this work is to use the registration-derived metric to classify images. Let us denote the Riemannian metric by $K_L(\alpha, \beta) = \langle Lv, v \rangle$. In practice, β plays an insignificant role and can be fixed, as multiplication of the velocity by a constant does not change the optimization problem in LDDMM. We focus on optimizing α , fixing $\beta = 1$ as a normalization term.

We optimize α in the EM framework as follows.

- E-step:

Register each pair of images in our training sample optimizing equation 2 to derive $K_L(x_i, x_j)$. Define $K(x_i, x_j) = \exp\{-\gamma K_L(x_i, x_j)\}$ and apply KLDA using $K(x_i, x_j)$. The parameter gamma is estimated by grid search to make a computation easier, but it can be also estimated by gradient descent. Estimate the hinge loss 4 given a fixed α .

- M-step:

Minimize the hinge loss 4 with respect to α .

The primary computational challenge above is in the M-step. Though the decision function is non-convex with respect to α , we seek a local minimum via gradient descent. We give the gradient direction with respect to $\theta = (\alpha, \beta)$ below, keeping in mind that β is fixed.

$$\frac{dh(x', z)}{d\theta} = \begin{cases} -z \frac{dy(x')}{d\theta}, & \text{if } zy(x') < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Using the matrix notation $y(x') = w^T (k(x') - (M_1 + M_{-1})/2)$ with $k_i(x') = K(x_i, x')$ one can obtain

$$\begin{aligned} \frac{dy(x')}{d\theta} &= \left(k(x') - \frac{M_1 + M_{-1}}{2} \right)^T \frac{dw}{d\theta} + w^T \left(\frac{dk(x')}{d\theta} - \frac{1}{2} \frac{d(M_1 + M_{-1})}{d\theta} \right), \\ dw &= -N^{-1}(dN)N^{-1}(M_1 - M_{-1}) + N^{-1}(dM_1 - dM_{-1}), \\ \frac{dK(x_i, x')}{d\theta} &= -\gamma K(x_i, x') \frac{dK_L(x_i, x')}{d\theta} \\ \frac{dK_L(x_i, x')}{d\theta} &= \begin{pmatrix} 2\langle (\alpha\Delta^2 + \beta\Delta)v_{x,x'}, v_{x,x'} \rangle, \\ 2\langle (\alpha\Delta + \beta E)v_{x,x'}, v_{x,x'} \rangle. \end{pmatrix} \end{aligned} \quad (6)$$

The resulting algorithm requires $n \times (n - 1)$ registrations at each EM step to train, and n registrations to a new image from each of n images in training sample to apply.

3 Experiments

To derive a baseline set of metrics between pairs of images, we selected α to maximize mutual information between registered images. This metric was then used to define the kernel in the KLDA classifier, the results of which we used as a baseline accuracy for our proposed method.

Our initial experiments were based on 100 images of rectangles, and 100 images of ellipses, each generated with a random locally affine deformation sufficiently noisy to obscure the original class of the image to the naked eye (**Figure 1**). Using 50 deformed ellipses and 50 deformed rectangles as a training dataset, we optimized α until hinge loss convergence. **Figure 2 (left)** shows that EM converges stably after several iterations based on ROC AUC. The final model, chosen based on the best training ROC AUC, performed nearly as well on the synthetic test dataset: ROC AUC = 0.84.

Our 3D hippocampal shape sample was derived from the SchizConnect brain MRI dataset [10]. We used right hippocampal segmentations extracted with FreeSurfer [11] from 227 Schizophrenia (SCZ) patients and 496 controls (CTL). All shapes were affinely registered to the ENIGMA hippocampal shape atlas [12], and their binary mask was computed from the transformed mesh model.

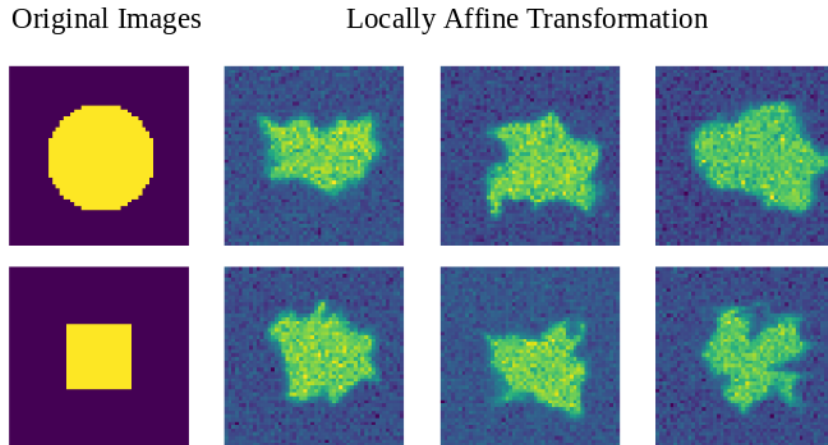


Fig. 1: Synthetic data generation

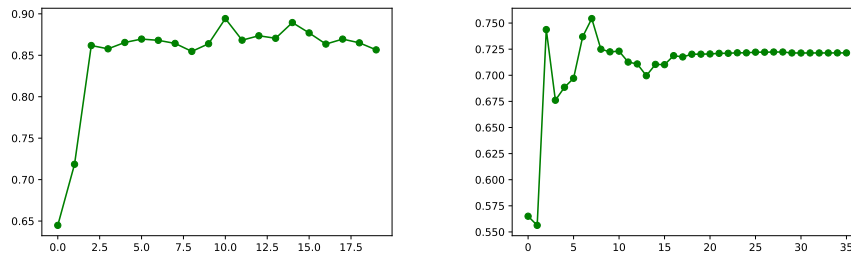


Fig. 2: ROC area under the curve vs. EM iterations on (left) synthetic data and (right) hippocampal shape

We again used 100 training examples (50 CTL, 50 SCZ) in all our experiments below, using the remaining sample as a test dataset. To derive baseline results to compare with our algorithm’s performance on hippocampal shapes, we constructed two additional discriminative models.

- (1) A logistic regression model simply using the vectorized binary mask. No spatial information is used in this model.
- (2) A KLDA model constructed using LDDMM metrics optimized for registration quality.

ROC AUC scores for the three models are shown in 1.

As expected, ignoring spatial information leads to significant drop in performance. It is also encouraging to see improvement in the classification accuracy

	Logistic Regression	Maximum MI	Optimized LDDMM-kernel
ROC AUC	0.36±0.02	0.72 ± 0.06	0.75 ± 0.06

Table 1: ROC AUC scores for three models

when the LDDMM metric is optimized for this explicitly. The stability of the EM algorithm trained on hippocampal shapes is comparable to stability when synthetic data, as seen in **Figure 2) (right)**. To visualize the difference in the kernel-based models, we project the mean difference between SCZ subjects and controls in the scalar momenta defining the registration velocity fields [7], as seen in **Figure 3**.

4 Conclusion

We have presented a method to optimize registration parameters for improved classification performance. Method exploits the geodesic distance on the space of diffeomorphisms as an image similarity measure to be learned in the fashion of traditional metric learning [13]. Our aim in this work was twofold: 1. to show that the metricity of a high dimensional space of geometric objects can be successfully used to improve predictive modeling, and (2) to suggest a means of making the sophisticated mathematical machinery of constructions such a LDDMM more useful in medical imaging practice. As a first attempt, we believe this work shows progress towards both goals. A stable LDDMM metric optimization is devised, and classification accuracy in our real-world application is indeed improved. The main drawback is the significant computational burden, as $N \times N$ training registrations are required. One approach to alleviate this problem is to lift the classification problem onto the tangent space at identity, thus requiring only N training registrations to an atlas, similar to [4]. Other generalizations of the idea presented here are possible both in LDDMM and other metric frameworks. We hope our work will inspire these generalizations to be developed.

5 Acknowledgements

This work was funded in part by the Russian Science Foundation grant 17-11-01390.

References

1. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage* **46**(3) (2009) 786–802

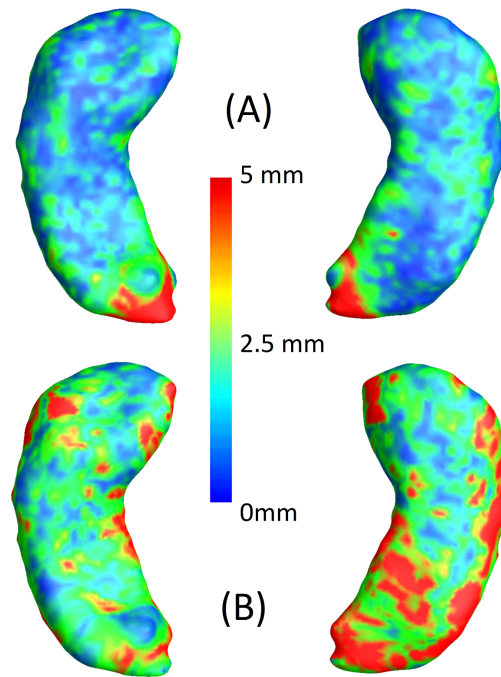


Fig. 3: **Mean momentum difference between Schizophrenics and healthy subjects**, using (A) a classification-optimized metric and (B) a metric optimized for pairwise mutual information. The effect in the latter is diffuse, while the classification-aware metric focuses on the hippocampal tail.

2. Beg, M.F., Miller, M.I., Trouné, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision* **61**(2) (February 2005) 139–157
3. Fletcher, P.T., Venkatasubramanian, S., Joshi, S.C.: The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage* **45** **1 Suppl** (2009) S143–52
4. 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.: *Information Processing in Medical Imaging*, Berlin, Heidelberg, Springer Berlin Heidelberg (2013) 37–48
5. Hong, Y., Golland, P., Zhang, M.: Fast geodesic regression for population-based image analysis. In Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S., eds.: *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Cham, Springer International Publishing (2017) 317–325

6. Zhang, M., Fletcher, P.T.: Probabilistic principal geodesic analysis. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. NIPS'13, USA, Curran Associates Inc. (2013) 1178–1186
7. Mang, A., Gholami, A., Biros, G.: Distributed-memory large deformation diffeomorphic 3d image registration. SC16: International Conference for High Performance Computing, Networking, Storage and Analysis (2016) 842–853
8. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. In: Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop., Ieee (1999) 41–48
9. Aronszajn, N.: Theory of reproducing kernels. Transactions of the American mathematical society **68**(3) (1950) 337–404
10. Wang, L., Alpert, K.I., Calhoun, V.D., Cobia, D.J., Keator, D.B., King, M.D., Kogan, A., Landis, D., Tallis, M., Turner, M.D., Potkin, S.G., Turner, J.A., Ambite, J.L.: Schizconnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. NeuroImage **124** (2016) 1155 – 1167
Sharing the wealth: Brain Imaging Repositories in 2015.
11. Fischl, B.: Freesurfer. Neuroimage **62**(2) (2012) 774–781
12. Roshchupkin*, G.V., Gutman*, B.A., Vernooij, M.W., Jahanshad, N., Martin, N.G., Hofman, A., McMahon, K.L., van der Lee, S.J., van Duijn, C.M., de Zubicaray, G.I., Uitterlinden, A.G., Wright, M.J., Niessen, W.J., Thompson, P.M., Ikram**, M.A., Adams**, H.H.H.: Heritability of the shape of subcortical brain structures in the general population. Nature Communications **7** (2016) 13738
13. Bellet, A., Habrard, A., Sebban, M.: A Survey on Metric Learning for Feature Vectors and Structured Data. ArXiv e-prints (June 2013)