

# Sparsity and Scale: Compact Representations of Deformation for Diffeomorphic Registration

Stefan Sommer, Mads Nielsen, Xavier Pennec

► **To cite this version:**

Stefan Sommer, Mads Nielsen, Xavier Pennec. Sparsity and Scale: Compact Representations of Deformation for Diffeomorphic Registration. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2012), Jan 2012, Breckenridge, Colorado, United States. 2012, <10.1109/MMBIA.2012.6164753>. <hal-00641357>

**HAL Id: hal-00641357**

**<https://hal.inria.fr/hal-00641357>**

Submitted on 15 Nov 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sparsity and Scale: Compact Representations of Deformation for Diffeomorphic Registration

Stefan Sommer<sup>1</sup>

<sup>1</sup>Department of Computer Science  
Univ. of Copenhagen, Denmark

sommer@diku.dk

Mads Nielsen<sup>1,2</sup>

<sup>2</sup>BiomedIQ A/S  
Copenhagen, Denmark

Xavier Pennec<sup>3</sup>

<sup>3</sup>Asclepios Project-Team  
INRIA Sophia-Antipolis, France

## Abstract

*In order to detect small-scale deformations during disease propagation while allowing large-scale deformation needed for inter-subject registration, we wish to model deformation at multiple scales and represent the deformation at the relevant scales only. With the LDDMM registration framework, enforcing sparsity results in compact representations but with limited ability to represent deformation across scales. In contrast, the LDDKBM extension of LDDMM allows representations of deformation at multiple scales but it does not favour compactness and hence may represent deformation at more scales than necessary. In this paper, we combine a sparsity prior with the multi-scale framework resulting in an algorithm allowing compact representation of deformation across scales. We present a mathematical formulation of the algorithm and evaluate it on a dataset of annotated lung CT images.*

## 1. Introduction

Deformation captured in image registration occur at multiple scales: lungs deform at large scale during the respiratory phases while disease progression may only be detected at small scales. Similarly, large-scale deformation is needed when registering brains from different subjects while e.g. atrophy in the hippocampus occur at small scales during the progression of Alzheimer’s disease. Representing deformation at multiple scales is therefore useful when performing statistics on small-scale features over a population requiring large-scale inter-subject registration. In this paper, we develop a method that represents deformation at multiple scales while seeking to represent the deformation at the relevant scales only. The resulting sparse, multi-scale registration will extend the range of deformation expressed by single-scale models, and we show how the extra capacity of the method does not hamper generalization to test data.

## 1.1. Background

The LDDMM framework is widely used in the field of computational anatomy to model deformation and perform registration of geometric objects. It provides convenient parametrization of flows of diffeomorphisms and a complete mathematical setting ensuring existence of optimal warps and allowing meaningful statistics to be performed on the registration results. Essential to the framework is the notion of momentum, and recent work has shown that even infinite dimensional spaces of initial momenta can be successfully approximated using sparse sets of parameters [8]. Such compact representations can be obtained by imposing a sparse prior on a finite sampling of the momentum space. However, the notion of kernels and their shape and scale imposes restrictions on the finite dimensional parametrizations, and it limits the range of deformations the model is able to express.

On the other hand, including multiple scales and kernel shapes in the LDDMM framework has been the subject of several works [2, 11, 15, 14, 12]. The common goal has been to model deformation occurring at multiple scales resulting in improved registration results and scale-aware statistics. Deformation at different scales may however occur at *different spatial locations*, and we wish to represent deformation at different locations at the appropriate scales *only*. The current multi-scale frameworks have no provision for ensuring this: as an example consider registering two images of fairly uniform objects. The large-scale deformations can then be expected to be located at the center of the object while lower scale deformations occur close to the boundaries. The sum of kernels approach [12] will represent deformation at all scales at all spatial locations; the LDDKBM construction [15] is able to represent deformation at different spatial locations at different scales but it will only do so if equipped with a penalty on non-relevant scales.

Here, we combine sparse priors on the initial momenta with the multi-scale LDDKBM framework in order to ben-

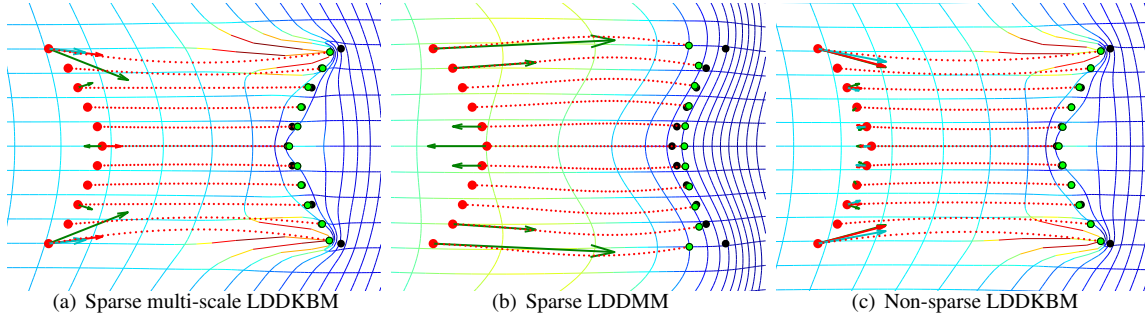


Figure 1. Matching eleven landmarks (red) to eleven landmarks (black) and results (green) with the three different registration methods considered in this paper: (a) the proposed method, sparse LDDKBM (multi-scale); (b) sparse LDDMM (single-scale); and (c) non-sparse LDDKBM (multi-scale). The arrows show the initial momentum with different colors for each of the three scales. Initially square grids are shown deformed by each diffeomorphism; the grids are colored with the log-trace of Cauchy-Green strain tensor. The inherent single-scale behaviour of sparse LDDMM causes large deviation between the landmarks and results (black and green) as figure (b) shows. Non-sparse LDDKBM matches the points well but the lack of sparse prior results in a non-compact representation as the non-zero momentum vectors on the initial points in (c) shows. The proposed method (a) obtains a compact representation (few non-zero momentum vectors) and the sparser representation continues to provide a good match between the landmarks and results (black and green). This happens without sacrificing warp regularity: the deformed grid with sparse prior (a) is similar to the deformed grid without the prior (c).

efit from both methods. LDDKBM was explicitly constructed to allow *sparse deformation description across space and scales*, and, therefore, the framework allows the sparse priors to be applied *independently* across the multiple scales. The method is thus able to faithfully represent deformations that are located at different places at different scales. The resulting multi-scale, sparse parametrization greatly extends the *range* of deformations expressed by the sparse models while allowing *compact* representations of deformation occurring at multiple scales.

## 1.2. Related Work

The deformable template model pioneered by Grenander [9] and the flow approach by Christensen et al. [6] together with the theoretical contributions of Dupuis et al. and Trounev [7, 16] started the development of the LDDMM deformation model. Beg et al. [1] developed algorithms for computing optimal diffeomorphisms in the framework, and the momentum representation has been used for statistics and for momentum based algorithms for the landmark matching problem [17]. The review paper [19] and the book [18] provide excellent overviews of the theory and applications of LDDMM in medical imaging.

Multi-scale extensions of LDDMM have been treated in several recent works. Sommer et al. represented deformation at different scales by adding a kernel bundle to the standard LDDMM construction [15, 14, 13]; the resulting framework is denoted LDDKBM. Independently, Bruveris et al. developed an extension of LDDMM allowing two scales through the use of semi-direct product groups [2]. This construction matches LDDKBM for the special case of two scales. Risser et al. [12] included scale in LDDMM

by adding kernels of different scales. Interestingly, it has recently been showed that the latter method finds the same optimal warps as LDDKBM [3] when the  $L^2$ -norm is combined linearly across scales. The two methods differ in representation and they will handle a sparse prior as proposed in this paper differently.

The literature on sparse representations and sparse penalty functions is wide, and we will in this paper limit the discussion to a small set of such priors [4]. A control point formulation of LDDMM template-based image registration has been developed by Durrleman et al. [8]. Sparsity is enforced by a  $\log -L^1$  penalty on the initial momenta, and the prior guides a search towards low-dimensional representations of deformation for populations of images. The method was developed for image registration but the sparse prior introduced apply to any finite dimensional LDDMM implementation. The fixed size of the kernels does however limit the expressiveness of the model. The fundamental idea behind the present paper is to remove this limitation by using kernels of multiple scales; we intend to fuse the sparse prior with the LDDKBM multi-scale extension.

## 1.3. Content and Outline

In the next section, we briefly describe the LDDMM registration framework and the LDDKBM multi-scale extension. We then continue with our main contribution: applying a sparse prior to LDDKBM to introduce compact, multi-scale representations. The resulting construction will apply to LDDKBM registration in general though the focus here will be on landmarks and finite dimensional image parametrizations. After studying implementation details, we apply the method on synthetic data as well as on a

dataset of annotated lung CT images, and we show how the increased capacity to model deformation does not hamper its ability to generalize to test data.

The paper thus contributes by

- (1) showing how the limitations in sparse LDDMM and multi-scale LDDKBM complements each other,
- (2) combining the sparse prior for LDDMM with the scale representation in LDDKBM to represent deformation compactly across scales,
- (3) and evaluating the resulting algorithm to show the obtained compactness and conserved generalization ability.

## 2. LDDMM Registration and Scale

The Large Deformation Diffeomorphic Kernel Bundle Mapping (LDDKBM) framework extends the single-scale Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework by allowing regularization at multiple-scales in the registration. We here provide a brief overview of the registration problem and how it is treated in both frameworks. For further details, we refer to the paper [15] introducing LDDKBM and the monograph [18] with extensive details on LDDMM.

Registration of geometric objects in the current setting is performed by defining an action of diffeomorphisms on the objects before searching for diffeomorphisms matching the objects through the action. For example, in order to register landmarks  $x_1, \dots, x_N$  and  $y_1, \dots, y_N$  in  $\mathbb{R}^d$ ,  $d = 2, 3$ , we search for a diffeomorphism  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that  $\varphi(x_i) = y_i$ . Equivalently, if we wish to register the moving image  $I_m$  to the fixed image  $I_f$ , we search for  $\varphi$  such that  $I_m \circ \varphi^{-1} = I_f$ . Frequently, a perfect match is not possible or even not desirable because noisy data may force the diffeomorphism to be highly irregular. Instead, the problem is stated in variational form as a search for  $\varphi$  minimizing

$$E(\varphi) = E_1(\varphi) + \lambda U(\varphi) \quad (1)$$

where  $E_1$  is a regularization measure,  $U$  a measure of the quality of the match which is defined using the data but dependent on  $\varphi$ , and  $\lambda > 0$  a weight. A simple and often used choice for  $U$  is the  $L^2$ -error which takes the form  $U(\varphi) = \sum_{i=1}^N \|\varphi(x_i) - y_i\|^2$  for landmarks. In the LDDKBM framework, the regularization measure  $E_1(\varphi)$  is defined as the minimum energy of paths of diffeomorphisms transporting the identity  $\text{Id}_\Omega$  to  $\varphi$ . We explain each part of the formal expression

$$E_1(\varphi) = \min_{w_t \in W, \varphi_{0t} = \varphi} \int_0^1 \|w_s\|_W^2 ds \quad (2)$$

in the following: The term  $\varphi_{0t}^{\Psi(w)}$  denotes the diffeomorphism path starting at  $\text{Id}_\Omega$  generated by the time-dependent vector field  $\Psi(w_t)$ , i.e.  $\varphi_{0t}^{\Psi(w)}$  satisfies the differential equation  $\partial_t \varphi_{0t}^{\Psi(w)} = \Psi(w_t) \circ \varphi_{0t}^v$ . The space  $W$  containing  $w_t$  is denoted the kernel bundle and consist of a family of vector spaces  $V_r$  parameterized by  $r$ , the scale. Each vector space  $V_r$  can be considered a subset of a tangent space  $V$  of a suitable Lie group of diffeomorphisms. The tangent space  $V$  is in turn a subset of the vector fields  $\Omega \rightarrow \mathbb{R}^d$ . The map  $\Psi$  collects parts  $w_r$  of a bundle vector  $w \in W$  at each scale  $r$  to one derivative vector in  $V$  by integration  $\Psi(w) = \int_{I_W} w_r dr$ . The norm  $\|\cdot\|_{V_r}$  on each  $V_r$  is allowed to vary with  $r$ , and the bundle norm  $\|\cdot\|_W$  is defined by

$$\|w\|_W^2 = \int_{I_W} \|w_r\|_{V_r}^2 dr,$$

i.e., the integral of the energy over all scales. Note that the weighting between the different scales can be modified by scaling the norms  $\|\cdot\|_{V_r}$ . The original LDDMM framework corresponds to the case of a single scale.

The bundle norm is chosen to penalize highly varying paths while allowing variation at different scales to be penalized differently. In short, a low value of  $E_1(\varphi)$  implies that the path to reach  $\varphi$ , and hence  $\varphi$  itself, is regular.

### 2.1. Velocity and Momentum

Tightly connected to the norm on each scale in the bundle are the notions of kernels and momenta. Since each  $V_r$  is chosen sufficiently regular in order to be a reproducing kernel Hilbert space [18], the existence of kernels  $K_r : \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$  with the reproducing property  $\langle K_r(\cdot, x)a, K_r(\cdot, y)b \rangle_{V_r} = a^T K_r(x, y)b$  for all points  $x, y \in \Omega$  and all vectors  $a, b \in \mathbb{R}^d$  is ensured. The momentum, named so because of its connection to momentum in classical mechanics, is an operator  $L_r$  on  $V_r$  defined by

$$\langle L_r v, w \rangle_{L^2(\Omega)} = \int_{\Omega} (L_r v(x))^T w(x) dx = \langle v, w \rangle_{V_r}, \quad (3)$$

and it therefore connects the inner product on  $V_r$  with the inner product in  $L^2(\Omega)$ . The value  $L_r v$  might be singular and in fact  $L_r(K_r(\cdot, y)a)(x)$  is the Dirac measure  $\delta_y(x)a$ . For landmark matching with  $N$  landmarks, the momentum will be located as Dirac measures at the point positions  $x_{t,l}$ ,  $l = 1, \dots, N$  at time  $t$ , and the velocity  $w_{t,r}$  at each scale  $r$  at time  $t$  is then given by

$$w_{t,r} = \sum_{l=1}^N K_r(\cdot, x_{t,l}) a_{t,l,r}. \quad (4)$$

### 2.2. Optimal Warps

Optimal paths for (2) are governed by the KB-EPDiff equations which extends the EPDiff equations for LDDMM

[14]. These evolution equations assert that the bundle velocity  $w_0$  of the path at time  $t = 0$  changes in a specific way throughout the evolution of the path from  $t = 0$  to its end at  $t = 1$ . This property is denoted momentum conservation, and it allows a search for a  $\varphi$  minimizing (1) to be phrased in terms of the initial bundle velocity: if we assume  $\varphi = \varphi_{01}^{\Psi(w)}$  then the values of both  $E_1(\varphi)$  and  $U(\varphi)$  are determined by  $w_0$  and we optimize

$$E(w_0) = E_1(w_0) + \lambda U(w_0) \quad (5)$$

instead of (1). In practice, this can be done by giving an initial guess for  $w_0$ , calculating the gradient  $\nabla E(w_0) = \nabla E_1(w_0) + \lambda \nabla U(w_0)$ , and updating  $w_0$  in a gradient descent or similar fashion. Because of the connection (4) between velocity and momentum, the search can equivalently be formulated in terms of the initial momentum.

### 3. Sparse LDDKBM

In a variety of applications, it is useful to obtain compact representations in the form of *sparse* solutions [4]. The standard method of obtaining sparsity is to add a penalty function to a variational formulation of the problem. The penalty function is also denoted a *sparse prior*.

Combining sparsity and multi-scale representations promises enhancements for both pairwise and group-wise registration: For statistics following pairwise registration with the aim of retrieving scale information, it is paramount to represent the deformation at the right scale *only*. Low-scale deformation may be represented by high-scale momenta but will require a higher number of non-zero parameters than if represented at the correct low-scale. Enforcing sparsity makes the low-scale representation more likely. This property is possible with *sparsity across scales* as discussed below.

For group-wise registration, each pair of images may be registered with a sparsely parameterized deformation. However, the non-zero momenta may have different spatial localization for the different pairs of images. Sparsity should therefore in this case be applied on a group level. Inter-subject registration may however emphasize the need for multi-scale representation: if modeling inter-subject differences using only a single large-scale, small scale features may be lost. If using only small-scale deformation, the representation will not be sparse.

Durrleman et al. [8] showed that the number of points in a finite control point formulation of LDDMM can be controlled by a  $\log - L^1$  like penalty term: a weight  $\gamma_{sp}$  and truncated log function

$$f_{\log^c}(x) = \max(\log(x), \log(c)) - \log(c)$$

is applied to the norms of the set of  $N$  single-scale momenta

resulting in the extension of (1) to the energy

$$E(\varphi) = E_1(\varphi) + \lambda U(\varphi) + \gamma_{sp} \sum_{l=1}^N f_{\log^c}(\|a_{0,l}\|). \quad (6)$$

The prior is added to all elements of a population of images, and it is shown that a fairly large reduction in the number of non-zero momenta does not affect the registration results much.

In the multi-scale case, the connection (3) between the momentum space and the kernel bundle can also be exploited in order to define penalty functions. Sparsity is generally formulated via the  $L_0$ -norm which on the bundle momenta take the form

$$\|w\|_{L^0} = \int_{I_W} \text{Area}\{Lw_r \neq 0\} dr.$$

This reduces to the number of non-zero momentum vectors  $\|w\|_{L^0} = \int_{I_W} |\{Lw_r \neq 0\}| dr$  in the finite-dimensional case. For sparse problems in general, optimization based on  $L_0$  penalty functions is a combinatorial problem and thus computationally prohibitive. Instead, the  $L_0$ -norm is approximated by the  $L_1$ -norm or similar functions.

In the multi-scale, finite dimensional setting, we parametrize the bundle momentum in the same way as the momentum is represented in the single-scale case: for  $N$  landmarks and  $R$  scales or, equivalently, for  $N$  control points and  $R$  scales in image registration,  $N \cdot R$  vectors  $a_{0,l,r}$  will specify the initial momentum. We then formulate a multi-scale sparse registration functional extending (6) by

$$E(\varphi) = E_1(\varphi) + \lambda U(\varphi) + \sum_{r=1}^R \gamma_{sp,r} \sum_{l=1}^N f(a_{0,l,r}) \quad (7)$$

and we require the evolution of  $a_{t,l,r}$  to follow the KB-EPDiff equations. Here  $\gamma_{sp,r}$  denote scale-dependent weights on the sparse prior  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ . As in the single-scale case, the idea is to push small momentum vectors towards zero without affecting large momenta much. We denote registration governed by (7) *sparse LDDKBM*.

#### 3.1. Choice of Prior

Approximations of the  $L_0$ -norm aiming to ease the complexity of the combinatorial optimization has been considered in many applications [4]. Though a full discussion of this subject out of scope of this paper, we will provide a brief rationale for our choice of penalty function. We note that ensuring convexity is not a major concern in this setting because the non-linearity of the connection between initial momenta and the match functional  $U$  makes the energy (1) non-convex even before adding the prior.

The most widely used approximation is probably the  $L^1$ -norm which provide sparse solutions but has the downside

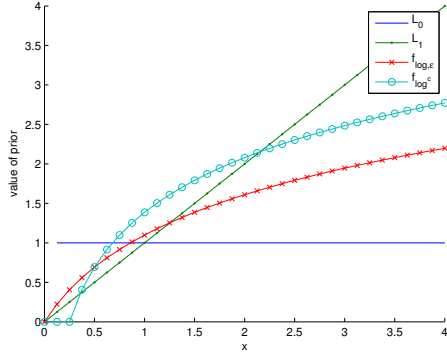


Figure 2. The  $L_0$ -norm for real valued data, and the approximations  $L_1$ ,  $f_{\log^c}$ , and  $f_{\log,\epsilon}$  with  $\epsilon = 0.5$  and  $c = 0.25$ . The  $L_1$ -norm provides poor approximation for large values. The truncated  $\log f_{\log^c}$  is not non-zero for small values.

of penalizing large momenta relatively hard, and it therefore provides poor approximation of the  $L_0$ -norm in such cases. The  $L^1$ -norm has been applied to LDDMM in addition to  $f_{\log^c}$  [10]. Candès et al. [4] proposes several penalty functions including the function

$$f_{\log,\epsilon}(x) = \log(1 + x/\epsilon).$$

Figure 2 illustrates the approximation of the  $L_0$ -norm provided by the  $L_1$ -norm,  $f_{\log^c}$ , and  $f_{\log,\epsilon}$ . Both  $f_{\log^c}$  and  $f_{\log,\epsilon}$  suffer less from the poor approximation for large momenta. Both necessitates a choice of parameter,  $c$  or  $\epsilon$ . Though  $f_{\log,\epsilon}$  may seem more natural than  $f_{\log^c}$  which is zero for small values, the gradient of  $f_{\log,\epsilon}$  may cause numerical issues close to zero. In the experiments section, we use  $f_{\log^c}$  to get results comparable with the single-scale algorithm in [8].

### 3.2. Sparsity Across Scales

An important quality of the sparse, multi-scale construction is that a momentum vector  $a_{0,l,r}$  at scale  $r$  may be zero while a momentum vector  $a_{0,l,r'}$  at scale  $r'$  for the *same point* may be non-zero. Hence, a purely low-scale deformation may be represented with momenta being non-zero at that particular scale only. The original motivation behind the LDDKBM construction was exactly to allow independent velocity at the different scales, and the behaviour of sparsity across scales is allowed by this fact. When not using sparse priors, Bruveris et al. [3] showed that, when the  $L^2$ -norm is combined linearly across scales, optimal deformations in the LDDKBM setting happen to coincide with results obtained with the sum of kernels approach [12] to multi-scale registration. With the latter method, momenta will only vanish at all scales simultaneously and hence the sought for sparsity across scales is not possible. The equivalence between the methods and the modeling benefits of

representing deformation only at the scales required indicate that modifying either the regularization or the match criterion in LDDKBM could improve results. Imposing a sparse prior constitutes exactly such a modification of the regularization criterion, and, as we will see in the experiments, optimal deformations with the prior do indeed exhibit sparsity across scales.

The weights  $\gamma_{sp,r}$  should ideally be chosen by cross-validation in same way the weight  $\lambda$  in (1) and the weighting between scales in the bundle are determined. At this point, we heuristically choose  $\gamma_{sp,r}$  either constant in  $r$  or  $\gamma_{sp,r} = \gamma_{sp}/r^\alpha$  for a fixed scalar  $\gamma_{sp}$  and exponent  $\alpha \geq 0$  in order to compensate for the often larger momenta at small scales. We will carry out a full examination of the influence of these weights in the future.

## 4. Implementation

We here describe how sparse LDDKBM can be computed in the case of landmark matching. Extending the method to images using a control point formulation similar to [8] poses no conceptual problem. The cost of adding the sparsity prior and computing its gradient is insignificant compared to the cost of integrating the flow equations described below. We do not experience an increase in the number of iterations of the optimization procedure, and, therefore, we see no significant increase in computation time.

Since the evolutions of the bundle momentum and velocity are required to follow the KB-EPDiff equations, we can optimize (7) in a gradient descent fashion. This can be done by giving an guess for the initial momentum  $a_0$ , calculating the gradient  $\nabla E(w_0) = \nabla E_1(w_0) + \lambda \nabla U(w_0)$  using  $w_0 = K a_0$ , adding the gradient from the sparsity term, and updating  $a_0$  by adding a vector proportional to the gradient.

The gradient  $\nabla E(w_0)$  can be computed using a two step algorithm: the initial bundle velocity  $w_0$  is transported forward in time to obtain the diffeomorphism  $\varphi$  before flowing the gradient at time  $t = 1$  backwards to obtain the gradient  $\nabla E(w_0)$  at  $t = 0$ . The KB-EPDiff equations governing the forward integration and the backwards gradient transport constitute non-linear ODEs which are finite if the set of scales  $I_W$  is finite. In practice,  $I_W$  is a discretization  $\{s_1, \dots, s_R\}$  of an interval  $[s_1, s_R]$  using  $R$  scalars. The ODEs can be integrated using standard Runge-Kutta integrators such as MATLAB's `ode45` solver.

Below, we let  $x_1, \dots, x_N$  denote the  $N$  points which we will match against  $N$  landmarks. Similarly,  $x_{t,i}$  will denote the point positions at time  $t$  and the set of time-dependent vectors  $a_{t,r,i}$  is the momentum of the flow. The vectors have components at each scale and are connected to the bundle velocity  $w_{t,i,r}$  through the kernels  $K_r(\cdot, \cdot)$  as expressed by the first evolution equation below. We use Gaussian kernels  $K_r(x, y) = \exp(-\frac{\|x-y\|^2}{r^2}) \text{Id}_d$ . The KB-EPDiff equations

then take the form

$$\begin{aligned} \Psi(w_t) &= \int_{I_W} \sum_{l=1}^N K_r(\cdot, x_{t,l}) a_{t,r,l} dr \\ \frac{d}{dt} a_{t,r,i} &= - \left( \int_{I_W} \sum_{l=1}^N \left( \frac{\partial}{\partial x} K_s(x_{t,i}, x_{t,l}) a_{t,s,l} \right)^T ds \right) a_{t,r,i} \\ x_{t,i} &= \varphi_{0t}^{\Psi(w)}(x_{0,i}). \end{aligned} \quad (8)$$

Since  $\varphi$  is determined by  $w_0$ , which through the evolution of  $w_t$  is uniquely linked to  $w_1$ ,  $U(\varphi)$  is determined by  $w_1$ . The gradient  $\nabla U(w_1)$  is usually known; if  $U$  measure the  $L^2$ -error, the gradient is just the vector with the  $i$ th component being  $2(x_{1,i} - y_i)$  where  $y_i$  are the target points. To perform gradient descent using  $w_0$ , we need the gradient  $\nabla U(w_0)$  which can be obtained by differentiating (8) and solving the transpose system backwards.<sup>1</sup> The gradient  $\nabla E_1(w_0)$  can be solved simultaneously by adding it to the backwards ODE. Combined, the gradient  $\nabla E(w_0)$  can be found as the solution at  $t = 0$  of an affine, non-autonomous ODE  $\dot{y}_t = v_t + M_t y_t$  integrated from  $t = 1$  to  $t = 0$ . The linear component transports  $\nabla U(w_t)$  while the affine component transport  $\nabla E_1(w_t)$ .

The penalty functions considered here have gradients

$$\begin{aligned} \nabla f_{\log^c}(a_{0,l,r}) &= \gamma_{sp,r} a_{0,l,r} / \|a_{0,l,r}\|^2, \\ \nabla f_{\log,\epsilon}(a_{0,l,r}) &= \gamma_{sp,r} a_{0,l,r} / ((e + \|a_{0,l,r}\|) \|a_{0,l,r}\|). \end{aligned}$$

When  $\|a_{0,l,r}\|$  is less than  $c$ , it is considered zero, in which case we do not add the gradient to  $\nabla E(w_0)$  when using  $\nabla f_{\log^c}$ . Pruning of small values  $a_{0,l,r}$  may be done during the optimization process but does not seem to effect stability of the algorithm much.

## 5. Experiments

We perform two sets of experiments: we start by illustrating the difference between the sparse and non-sparse, single- and multi-scale algorithms on synthetic examples. We then evaluate the effect of the sparse prior on a dataset of manually annotated landmarks on CT lungs images in different respiratory phases. The latter experiment is performed for pairwise registration and will show that large reductions in parameters may be achieved, and that the extra degrees of freedom in the multi-scale algorithm do not hamper the ability of the method to generalize to the test data.

### 5.1. Synthetic Examples

Figure 1 presents a simple example which, in our minds, captures the intuition for why fusing sparsity and multiple scales has great potential. In the figure, we show the results of matching two sets of 11 points using the sparse

<sup>1</sup>Confer [18] for a description of this method in the LDDMM case.

<sup>2</sup>distance after match/distance before match.

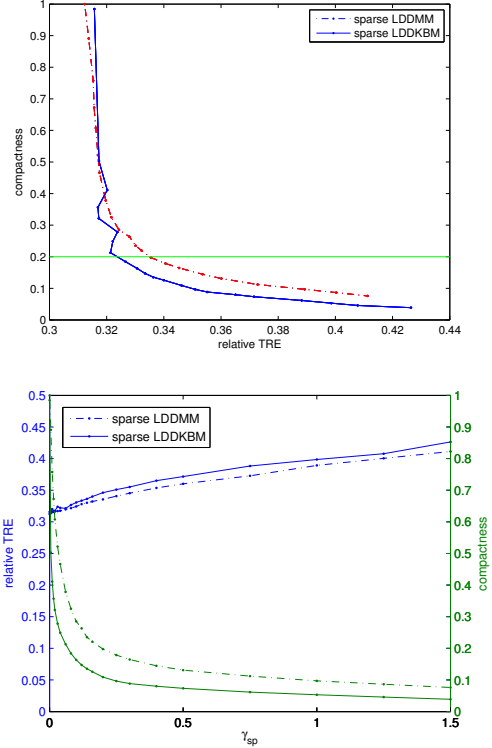


Figure 3. Top: relative TRE<sup>2</sup> increase (horizontal axis) versus relative compactness (vertical axis) when the varying sparsity weight  $\gamma_{sp}$ . Bottom: relative TRE increase (left axis, blue) and relative decrease in non-zero momenta (right axis, green) as function of  $\gamma_{sp}$ . Results averaged over 5 patients. With a factor 5 reduction in non-zero parameters (horizontal line, top), relative TRE for sparse LDDKBM is 0.324 in contrast to 0.336 for sparse LDDMM.

LDDKBM, the sparse LDDMM, and the non-sparse LDDKBM algorithms. In all cases, we search for a diffeomorphism transporting the moving points (red) to the fixed points (black). The green points show the results of the matchings, and the red dotted lines indicate the trajectory of the moving points along the diffeomorphism path. The initial momenta  $a_{0,l,r}$ ,  $l = 1, \dots, 11$  are shown with arrows. The Gaussian kernels have scale  $\sigma = 6$  for the single-scale LDDMM case and  $\sigma = 12, 6, 0.8$  for the multi-scale methods in grid units as indicated by the deformed grids,

The sparse prior on LDDMM forces vanishing momentum for 4 of the 11 points. Together with the fixed size kernel, this has a serious effect on the registration quality: the points are not quite well matched as seen by the large deviation between the landmarks and result points. The match is closer for the LDDKBM algorithm without sparse prior. However, all momenta at all scales are non-zero as shown by the non-zero momentum vectors, and the representation is far from compact. The multi-scale, sparse LDDKBM obtains the best of both worlds: even with vanishing mo-

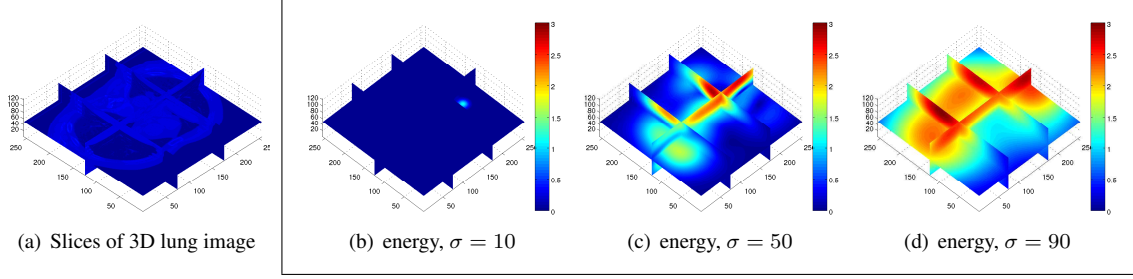


Figure 4. Slices of 3D lung image and the multi-scale initial vector field at three scales that when combined generate the sparse LDDKBM registration. Left to right: (a) slices of CT image, (b)-(d) squared  $L^2$ -norm of the components at each of the three scales  $\sigma = 10, 50, 90$  which in combination make up the multi-scale bundle vector  $w_0$  generating  $\varphi$  at  $t = 0$ . The uniform spread of the velocity provided by the large scale kernels results in a smooth flow even with only 20% percent non-zero momenta for that scale. The localized deformation field provided by the sparsity of the smaller momenta is in addition clearly visible.

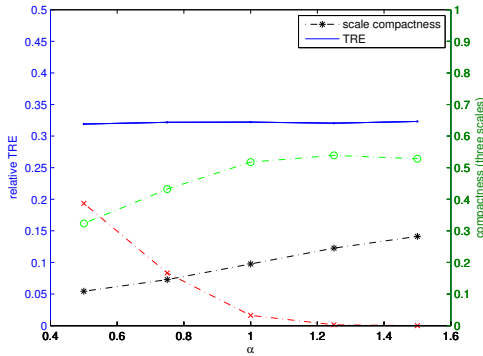


Figure 5. Relative  $TRE^2$  increase (left axis, blue) and relative decrease in non-zero momenta for each scale (right axis, red (x)/green (o)/black (\*)) as function of the scale weighting  $\alpha$  averaged over the 5 patients. The red (x) marked curve shows compactness for the smallest scale ( $\sigma = 10$ ), the green (o) marked curves for the midmost scale ( $\sigma = 50$ ) and the black (\*) marked curves for the largest scale ( $\sigma = 90$ ). The TRE and total number of non-zero parameters stay relatively constant though the distribution of non-zero parameters over scale varies.

menta for 6 of the 11 points, the match quality is comparable with non-sparse LDDMM. Of the  $3 \cdot 11$  momenta, 23 vanishes. The result shows that sparsity does indeed occur across scales: point 3 and 9 from above has non-vanishing momenta at only the smallest scale, and the central point (point 6 from above) has vanishing momentum only at the midmost scale.

## 5.2. Lung Registration

We now make use of the publicly available DIR [5] dataset of lung CT images and manually annotated landmarks. The dataset consists of five cases of CT images for different stages of the respiratory phases and annotated landmarks for the maximum inhale and exhale phases, re-

spectively. The images and landmarks are available on grids with voxel size varying slightly between the cases but close to  $1 \times 1 \times 2.5$  mm. Further details can be found in the reference.

For each case, the 300 publicly available landmarks,  $x_1^I, \dots, x_{300}^I$  for the maximum inhale phases and  $x_1^E, \dots, x_{300}^E$  for the maximum exhale phase, correspond pairwise. We drive the registration using random subsets of these landmarks and compute the target registration error (TRE) on the landmarks not used to drive the registration.

It has previously been shown [15] that LDDKBM provide as good results as LDDMM even after tuning the single-scale used in LDDMM to the best possible value using cross-validation. LDDKBM uses a range of scales and hence need no tuning. For the experiments here, we provide LDDMM with the advantage by selecting the best possible scale of  $\sigma = 50$  voxels. For LDDKBM, we use three scales,  $\sigma = 10, 50, 90$  voxels without any cross-validation tuning. We fix the match criterion weight in (7) to  $\lambda = 8$  in accordance with [15] and average over all 5 patients and several randomly selected subsets of points to drive the matching. The scale parameters vary by  $\gamma_{sp,r} = \gamma_{sp}/r$ . The value of  $\gamma_{sp,50} = \gamma_{sp}/50$  is used for LDDMM in accordance with the choice scale.

Figure 3 shows the connection between relative  $TRE^2$ , the sparsity weight  $\gamma_{sp}$ , and the fraction of momenta being non-zero after the match. As seen from the top figure, a reduction in the number of non-zero momenta of a factor 5 to 10 can be obtained for LDDKBM with only slightly increasing TRE. Sparse LDDKBM obtains the largest reduction of non-zero parameters for a given increase in relative TRE. LDDMM still provides the smallest number of total parameters but the gap narrows as TRE increases. This fact should be viewed in the light that LDDMM is already tuned to the best scale, and, importantly, that LDDKBM has more degrees of freedom than LDDMM. The extra capacity does thus not lead to reduced generalization ability. The bottom figure shows the reduction in non-zero momenta leveling



out while the relative TRE increases, though at a relatively slow pace. The absence of a sharp increase in relative TRE makes the method fairly robust the actual choice of  $\lambda_{sp}$ .

The weighting of the sparsity parameter across scales can be controlled by letting  $\gamma_{sp,r} = \gamma_{sp}/r^\alpha$  and varying  $\alpha$ . To explore this, we select  $\gamma_{sp} = 0.04$  which is a reasonable value from Figure 3, and plot the relative TRE against  $\alpha$  in Figure 5. In addition, the figure shows how the distribution of non-zero parameters at the different scales varies with  $\alpha$ . The increased penalty at small scales for  $\alpha > 1$  and corresponding increased penalty for large scales for  $\alpha < 1$  is clearly visible.

To illustrate the result of one lung registration with sparse LDDKBM, Figure 4 shows the energy of the initial velocity field for the three LDDKBM scales separately with  $\gamma_{sp} = 0.04$ . The uniform spread of the velocity provided by the large scale kernels results in a smooth deformation even with only 20% percent non-zero momenta at that scale. The localized deformation field provided by the sparsity of the smaller momenta is in addition clearly visible.

## 6. Conclusion and Outlook

We have successfully added sparse priors to the multi-scale LDDKBM extension of LDDMM. The resulting algorithm allows sparsity across scales, and the multiple scales extend the range of deformation the algorithm is able to model significantly. Since the priors are applied independently across scales, the model is able to faithfully represent deformation located at different places at different scales.

We have shown the intuition behind the method on synthetic data. In addition, when applying the method to a dataset of lung CT images, we show that the expected sparsity may indeed be achieved with only minor increase in registration error and that the extra capacity of the algorithm does not affect generalization ability. The method may as well be applied to images in the finite dimensional setting promising similar results and to group-wise registration extending the pairwise experiments presented here.

The full potential of the method will be apparent when applied to population analysis of e.g. atrophy during Alzheimer’s disease. The present paper shows that from both a theoretical and a practical point of view, the multi-scale ideas from LDDKBM and sparsity prior from sparse LDDMM benefit greatly from each other, and the presented method, sparse LDDKBM, provides compact representation of deformation across scales without hampering generalization ability.

## References

[1] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV*, 61(2):139–157, 2005. 2

[2] M. Bruveris, F. Gay-Balmaz, D. D. Holm, and T. S. Ratiu. The momentum map representation of images. *0912.2990*, Dec. 2009. 1, 2

[3] M. Bruveris, L. Risser, and F. Vialard. Mixture of kernels and iterated Semi-Direct product of diffeomorphism groups. *arXiv:1108.2472*, Aug. 2011. 2, 5

[4] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, Oct. 2008. 2, 4, 5

[5] R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology*, 54(7):1849–1870, Apr. 2009. 7

[6] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10), 2002. 2

[7] P. Dupuis, U. Grenander, and M. I. Miller. Variational problems on flows of diffeomorphisms for image matching. 1998. 2

[8] S. Durrleman, M. Prastawa, G. Gerig, and S. Joshi. Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. *IPMI*, 22, 2011. 1, 2, 4, 5

[9] U. Grenander. *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford University Press, USA, Feb. 1994. 2

[10] S. Joshi. Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. Presentation at the Geometry for Anatomy workshop, Banff, Canada, 2011. 5

[11] L. Risser, F. Vialard, R. Wolz, D. D. Holm, and D. Rueckert. Simultaneous fine and coarse diffeomorphic registration: application to atrophy measurement in alzheimer’s disease. *MICCAI 2010*, 13(Pt 2):610–617, 2010. 1

[12] L. Risser, F. X. Vialard, R. Wolz, M. Murgasova, D. D. Holm, and D. Rueckert. Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. *IEEE TMI*, 30(10):1746–1759, Oct. 2011. 1, 2, 5

[13] S. Sommer. Accelerating Multi-Scale flows for LDDKBM diffeomorphic registration. In *GPUCV workshop at ICCV 2011*. IEEE, 2011. 2

[14] S. Sommer, F. Lauze, M. Nielsen, and X. Pennec. Kernel bundle EPDiff: evolution equations for Multi-Scale diffeomorphic image registration. In *SSVM 2011*. Springer, 2011. 1, 2, 4

[15] S. Sommer, M. Nielsen, F. Lauze, and X. Pennec. A Multi-Scale kernel bundle for LDDMM: towards sparse deformation description across space and scales. In *IPMI 2011*. Springer, 2011. 1, 2, 3, 7

[16] A. Trouvé. An infinite dimensional group approach for physics based models in patterns recognition, 1995. 2

[17] M. Vaillant, M. Miller, L. Younes, and A. Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23(Supplement 1):S161–S169, 2004. 2

[18] L. Younes. *Shapes and Diffeomorphisms*. Springer, 2010. 2, 3, 6

[19] L. Younes, F. Arrate, and M. I. Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1, Supplement 1):S40–S50, Mar. 2009. 2