
Sparse Multi-Scale Diffeomorphic Registration: the Kernel Bundle Framework

Stefan Sommer · François Lauze · Mads Nielsen · Xavier Pennec

Received: date / Accepted: date

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 compactly at the relevant scales only. This paper presents the *kernel bundle* extension of the LDDMM framework that allows multiple kernels at multiple scales to be incorporated in the registration. We combine sparsity priors with the kernel bundle resulting in compact representations across scales, and we present the mathematical foundation of the framework with derivation of the KB-EPDiff evolution equations. Through examples, we illustrate the influence of the kernel scale and show that the method achieves the important property of *sparsity across scales*. In addition, we demonstrate on a dataset of annotated lung CT images how the kernel bundle framework with a compact representation reaches the same accuracy as the standard method optimally tuned with respect to scale.

Keywords kernel bundle · LDDKBM · LDDMM · diffeomorphic registration · scale space · computational anatomy

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 *kernel bundle* registration framework supports *sparsity across scales* while extending the range of deformation expressed by single-scale models. We extend the variational formulation, the evolution equations, and the momentum conservation property of the single-scale method, and we derive and test the construction to show that the cross-scale sparsity is indeed achieved; that the extra capacity of the method does not hamper generalization to test data; and that the method removes the need for classical scale selection.

1.1 Background

The LDDMM framework (Large Deformation Diffeomorphic Metric Mapping, [12,6,8,20]) 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. Recent work has shown

S. Sommer, F. Lauze, M. Nielsen
The Image Group, Department of Computer Science
University of Copenhagen, Denmark
Tel.: +4535321400
E-mail: sommer@diku.dk

M. Nielsen
BiomedIQ A/S
Copenhagen, Denmark

X. Pennec
Asclepios Project-Team
INRIA Sophia-Antipolis, France

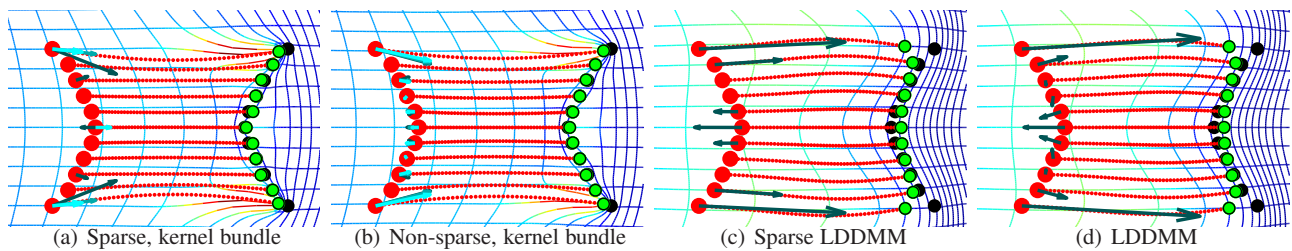


Fig. 1 Matching eleven landmarks (red) to eleven landmarks (black) and results (green) with four registration methods: (a) the proposed *kernel bundle* multi-scale method with sparse prior; (b) the kernel bundle method without sparse prior; (c) LDDMM (single-scale) with sparse prior; and (d) LDDMM (single-scale) without enforced sparsity. The arrows show the initial momentum with different colors for each of the three scales in (a) and (b). Initially square grids are shown deformed by each diffeomorphism; the grids are colored with the log-trace of Cauchy-Green strain tensor. (d) The inherent single-scale behaviour of LDDMM causes large deviation between the landmarks and results (black and green). (c) This effect is increased when adding a sparse prior; the low number of non-zero momentum vectors indicate the sparsity. (b) With the kernel bundle and multiple scales, the algorithm matches the points well through the increased capacity in the deformation description. (a) Adding a sparse prior to the kernel bundle results in a compact representation (few non-zero momentum vectors) with sparsity across scales, 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 (b).

that the infinite dimensional space of parameters for the registration can be successfully approximated using sparse, finite dimensional representations [10, 17, 9]. However, the notion of kernels, which lies at the heart of the framework, and the kernel shape and scale impose restrictions on the sparse representations, and it limits the range of deformations the model is able to express.

The key to obtain sparse representations without limiting the range of the deformation model is to increase the *capacity* of the deformation description. Locally, the capacity can be increased with higher order momenta [17] but varying the spatial extend of the deformation requires multiple scales. Enabling LDDMM to model deformation occurring at multiple scales has been the subject of several works [2, 13, 14] resulting in improved registration results. 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*. This requires a multi-scale framework designed specifically to allow sparse representations. 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. While previous approaches will represent deformation at all scales at all spatial locations, we aim for constructing a framework able to represent deformation at the appropriate scales *only*.

In order to achieve this, we introduce the *kernel bundle* framework (LDDKBM) which is designed to represent deformation individually at different scales. This in particular allows the momentum field to vanish at some scales while being non-zero at others at the same spatial locations, a property we denote *sparsity across scales*. By applying sparse priors on the momentum field at the different scales individually, we explicitly force the cross-scale sparsity, and the method therefore allows *sparse deformation description*

across space and scales. The resulting framework greatly extends the *range* of deformations expressed by the sparse models while allowing *compact* representations of deformation occurring at multiple scales.

With the sum of kernels multi-scale approach [14], the distribution of deformation over scale must be the same for all spatial locations. In contrast, the novel feature of the kernel bundle representation is the ability to allow the distribution of deformation over scale to be different for different spatial locations. The new framework therefore does not force the same scale distribution over the entire domain; instead, it adds the flexibility of using a particular set of scales in parts of the domain and different scales in other parts of the domain. This defining property furthermore adds the flexibility to incorporate priors on the scale distribution, and it allows us to enforce sparsity in the representation. The combination of the *flexible* representation and the support for *sparse priors* increases the range and compactness of the deformation model and thus distinguishes the kernel bundle from previous multi-scale approaches.

1.2 Deformation at Multiple Scales; An Example

Figure 1 shows a simple example of landmark matching. In order to register the points, movement is needed at both large and small scales, and the single-scale nature of the LDDMM algorithm limits its ability to match the points well. This fact becomes even more expressed when adding a sparsity prior. With multiple-scales, the match improves as seen by the reduced deviation between the landmarks and results. When adding a sparse prior to the kernel bundle, equivalent precision can be obtained with a compact description which exhibits sparsity across scales.

1.3 Related Work

The deformable template model pioneered by Grenander [12] and the flow approach by Christensen et al. [6] together with the theoretical contributions of Dupuis et al. and Trouvé [8, 20] 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 [21]. The review paper [23] and the book [22] 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. Bruveris et al. developed an extension of LDDMM allowing two scales through the use of semi-direct product groups [2] and Risser et al. [14] included scale in LDDMM by adding kernels of different scales. The approach of Risser et al. does not divide the deformation description across scales and enforced sparsity will occur at all scales simultaneously. This is in direct contrast to the result we seek to obtain: we search for a representation that can handle both small and large scale features independently to allow different deformation at different scales, and we wish to allow the effect of enforced sparsity to occur at different scales individually. A representation supporting this *sparsity across scales* property is the main contribution of this paper, and the fact that the kernel bundle representation supports this property fundamentally differentiates it from the approach of Risser et al. With sparse priors, optimal results in the kernel bundle framework will generally encode the deformation using different combinations of scales at different spatial locations while the approach of Risser et al. imposes the combination of scales to be the same for all spatial locations.

Outside the registration setting, the effect on the underlying Hilbert spaces when scaling the kernel have been treated in [11]. Increasing the capacity of the deformation description locally can be obtained with higher order momenta [17].

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. [10,9]. Sparsity is enforced by a $\log -L^1$ og 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.

1.4 Content and Outline

This paper combines the conference papers [18,16,19] and adds additional new material. We aim at presenting a full account of the kernel bundle framework that in the previous papers is also denoted the LDDKBM method (KB for *kernel bundle*). The new material comprises full derivation of the forwards and backwards gradient transport equations which are fundamental for computing optimal warps with the framework; additional algorithm information; discussion on the relation to other multi-scale approaches; and extended experiments section showing the obtained effect of sparsity across scales and using cross validation to tune the regularization weights for the method comparisons.

We start by discussing the variational formulation of LDDMM and the kernel bundle method before presenting the theoretical construction allowing the multi-scale representation. We relate the method to other multi-scale approaches before deriving the KB-EPDiff evolution equations. Next follows the forwards and backwards transport equations with implementation details and the extended experiments section. We end the paper with discussing future work and concluding remarks.

2 Registration: the LDDMM and Kernel Bundle Variational Formulation

The kernel bundle framework extends the single-scale LDDMM (Large Deformation Diffeomorphic Metric Mapping) framework by allowing regularization at multiple-scales in the registration. We here provide an overview of the registration problem and the variational formulation used in both frameworks.

In the kernel bundle and LDDMM frameworks, registration is performed through the action of diffeomorphisms on geometric objects. The approach is very general and allows the frameworks to be applied to both landmarks, curves, surfaces, images, and tensors. In the case of landmarks, the action of a diffeomorphism φ takes the form $\varphi.x = \varphi(x)$, and given landmarks x_1, \dots, x_N and y_1, \dots, y_N , the registration amounts to a search for φ such that $\varphi.x_i \sim y_i$ for all $i = 1, \dots, N$. In exact matching, we wish $\varphi.x_i$ be exactly equal to y_i but, more frequently, we allow some amount of inexactness to account for noise and give smoother diffeomorphisms. This is done by defining a quality of match measure U and a regularization measure E_1 to give a combined energy

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

Here λ is a positive real representing the trade-off between regularity and goodness of fit and U is often the L^2 -error which in the landmark case takes the form $U(\varphi) = \sum_{i=1}^N \|\varphi(x_i) - y_i\|^2$.

2.1 The Regularization Energy

The formulation of the regularization energy E_1 in the kernel bundle framework is an extension of the LDDMM formulation. We here introduce notation which will lead to the LDDMM formulation before describing the extension in the next section. Let the domain Ω be a subset of \mathbb{R}^d with $d = 2, 3$ in applications, and let V denote a Hilbert space of vector fields $v : \Omega \rightarrow \mathbb{R}^d$ such that V with associated norm $\|\cdot\|_V$ is included in $L^2(\Omega, \mathbb{R}^d)$ and admissible as defined in [22, Chap. 9]. Given a time-dependent vector field $t \mapsto v_t$ with

$$E_1(v_t) = \int_0^1 \|v_t\|_V^2 dt < \infty \quad (2)$$

the associated differential equation $\partial_t \varphi_t = v_t \circ \varphi_t$ has with initial condition $\varphi_s = \varphi$ a diffeomorphism φ_s^v as unique solution. The set G_V of diffeomorphisms built from V by such differential equations is a Lie group, and V is its tangent space at each point. The inner product on V associated to the norm $\|\cdot\|_V$ makes G_V a Riemannian manifold with right-invariant metric. Setting $\varphi_{00}^v = \text{Id}_\Omega$, the map $t \mapsto \varphi_{0t}^v$ is a path from Id_Ω to φ with energy given by (2). We will use this notation throughout the paper. A critical path for the energy is a geodesic on G_V , and the LDDMM regularization energy is defined by

$$E_1(\varphi) = \min_{v_t \in V, \varphi_{01}^v = \varphi} E_1(v_t) = \min_{v_t \in V, \varphi_{01}^v = \varphi} \int_0^1 \|v_s\|_V^2 ds, \quad (3)$$

i.e., it measures the minimal energy necessary to pass from Id_Ω to φ . The energy penalizes highly varying paths and, therefore, a low value of $E_1(\varphi)$ implies that φ is regular.

The regularity is ultimately controlled by the norm on V and this norm is associated to a *reproducing kernel* $K : \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$. The kernel is often chosen to ensure rotational and translational invariance [22] and the Gaussian kernel $K(x, y) = \exp(-\frac{\|x-y\|^2}{\sigma^2}) \text{Id}_d$ is a convenient and often used choice. The scaling factor σ is not limited to Gaussian kernels and allows for many kernels to vary the amount of regularization. Larger scales lead in general to higher regularization and smoother diffeomorphisms, whereas smaller kernels penalize higher frequencies less and often gives better matches. This phenomenon is in particular apparent for objects with sparse information and images with e.g. areas of constant intensity.

3 Kernels, Momentum, and the Kernel Bundle

The kernel bundle framework extends LDDMM by equipping the diffeomorphism manifold G_V in LDDMM with vector bundles allowing deformation to be described at different scales. We start this section by discussing the relation between kernels and momentum in LDDMM before defining the kernel bundle and discussing the mathematical foundation behind the framework.

3.1 Kernel and Momentum

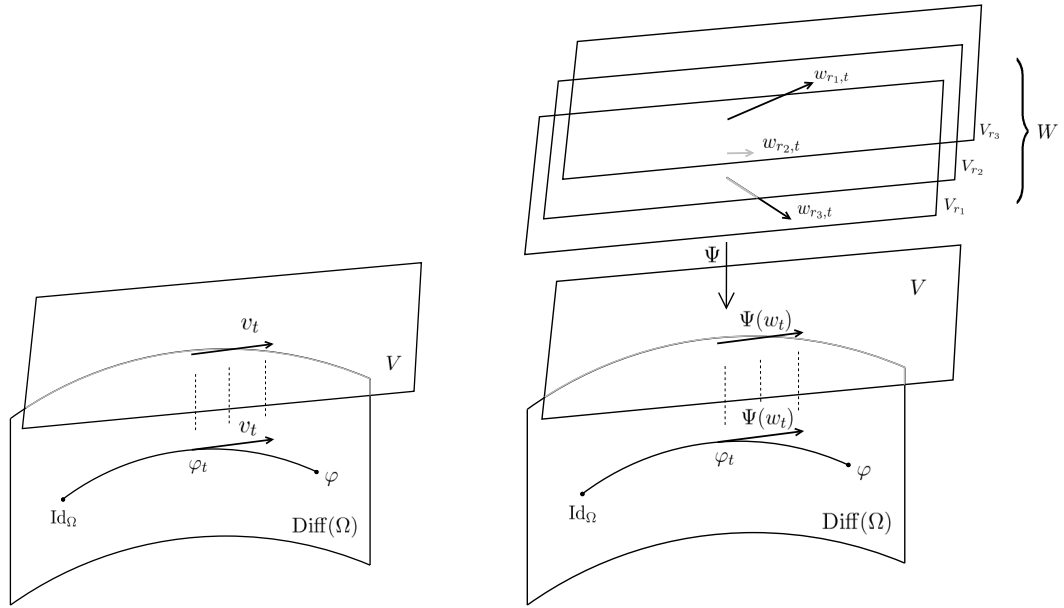
As a consequence of the assumed admissibility of V , the evaluation functionals $\delta_x : v \mapsto v(x) \in \mathbb{R}^d$ are well-defined and continuous for any $x \in \Omega$. Thus, for any $a \in \mathbb{R}^d$ the map $a \otimes \delta_x : v \mapsto a^T v(x)$ belongs to the topological dual V^* consisting of the continuous linear maps of V . This in turn implies the existence of spatially dependent matrices $K : \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$, the *kernel*, such that, for any constant vector $a \in \mathbb{R}^d$, the vector field $K(\cdot, x)a \in V$ represents $a \otimes \delta_x$ and $\langle K(\cdot, x)a, v \rangle_V = a \otimes \delta_x(v)$ for any $v \in V$, point $x \in \Omega$ and vector $a \in \mathbb{R}^d$. This latter property is denoted the reproducing property and gives V the structure of a reproducing kernel Hilbert space (RKHS). Tightly connected to the norm and kernels is the notion of *momentum* given by the linear momentum operator $L : V \rightarrow V^* \subset L^2(\Omega, \mathbb{R}^d)$ which satisfies

$$\langle Lv, w \rangle_{L^2(\Omega, \mathbb{R}^d)} = \int_\Omega (Lv(x))^T w(x) dx = \langle v, w \rangle_V \quad (4)$$

for all $v, w \in V$. The momentum operator connects the inner product on V with the inner product in $L^2(\Omega, \mathbb{R}^d)$, and the image Lv of an element $v \in V$ is denoted the momentum of v . The momentum Lv might be singular and in fact $L(K(\cdot, y)a)(x)$ is the Dirac measure $\delta_y(x)a$. Considering K as the map $a \mapsto \int_\Omega K(\cdot, x)a(x) dx$, L can be viewed as the inverse of K . Confer [22] for a thorough introduction to reproducing kernels, especially with a view towards the LDDMM framework.

3.2 The Kernel Bundle

In order to describe deformation at different scales, we extend in the following the tangent vector space V to a family of vector spaces W which will eventually lead to the bundle construction. We consider a parameter set I_W and subspaces $V_r, r \in I_W$ of the tangent space V where each V_r is equipped with a norm $\|\cdot\|_r$, corresponding kernel K_r , and momentum operator L_r . Typically, I_W will be a discrete set or a closed and bounded interval of \mathbb{R}^+ representing different scales.



(a) In LDDMM, a path on the manifold $G_V \subset \text{Diff}(\Omega)$ is generated by integrating the time varying vector field v_t in the tangent space V .

(b) With the kernel bundle, a path w_t in the vector space W , here constructed from V_{r_1} , V_{r_2} , and V_{r_3} , sum through the map Ψ to a vector field $\Psi(w_t)$, which in turn generates a path on $\text{Diff}(\Omega)$.

Fig. 2 The manifold view of LDDMM and the kernel bundle.

We then let W be the space of *Bochner integrable* functions¹ $w : I_W \rightarrow V$, $w_r \in V_r$ such that

$$\int_{I_W} \|w_r\|_r^2 dr < \infty \quad \text{and} \quad \int_{I_W} \|w_r\|_r dr < \infty.$$

The vector space structures on V_r induce a vector space structure on W , and the inner product

$$\langle v, w \rangle_W = \int_{I_W} \langle v_r, w_r \rangle_r dr, \quad v, w \in W$$

turns W into a Hilbert space [7]. With this construction, we obtain a vector bundle $G_V \times W$, the *kernel bundle*, allowing kernels of different sizes and shapes. A map $\Psi : G_V \times W \rightarrow TG_V = G_V \times V$ allows parts w_r of a bundle vector $w \in W$ at each scale r to be combined to one derivative vector in V . Ψ is defined using the Bochner integral, i.e. $\Psi(w) = \int_{I_W} w_r dr$.

We note that the parameter space I_W can be a compact interval or finite set of scalars in which case the integral reduces to just a sum. Often, it will be an interval specifying a scale range, and a practical implementation will discretize the interval into a finite set of scalars.

¹ The Bochner integral extends the Lebesgue integral to functions taking values in Banach spaces. The Banach space norm allows definition of L^p -spaces of Banach valued functions. In particular, the L^2 -spaces of functions taking values in Hilbert spaces are themselves Hilbert spaces [7].

3.3 Flows in the Bundle Setting

We use the map Ψ to get a relation between paths in W and paths in the manifold G_V that is similar to the usual connection between paths in V and paths in G_V . Given $w_t = \{w_{r,t}\}_r$, we define the path $\phi_{0t}^{\Psi(w)}$, i.e. the path starting at Id_Ω with derivative $\partial_t \phi_{0t}^{\Psi(w)} = \Psi(w_t) \circ \phi_{0t}^{\Psi(w)}$. We can measure the energy of a bundle path w_t by

$$E_1(w_t) = \int_0^1 \|w_s\|_W^2 ds, \quad (5)$$

and, using this energy, we get a new definition of the regularization energy E_1 :

$$E_1(\phi) = \min_{w_t \in W, \phi_{01}^{\Psi(w)} = \phi} E_1(w_t) = \min_{w_t \in W, \phi_{01}^{\Psi(w)} = \phi} \int_0^1 \|w_s\|_W^2 ds \quad (6)$$

Together with a quality of match measure $U(\phi)$, this defines the registration problem in the kernel bundle framework as the search for diffeomorphisms minimizing

$$E(\phi) = E_1(\phi) + \lambda U(\phi) \quad (7)$$

with E_1 given by (6). By design, paths in the kernel bundle generating the diffeomorphisms have components at each scale, and this is precisely the property that will later allow us to enforce sparsity at different scales individually. This

will be done by adding priors that affect the individual scale components of bundle vectors to (7).

The above registration energy should be compared with the LDDMM formulation (1) using the regularization (3). It is immediately clear that the kernel bundle formulation is an extension of the LDDMM regularization, since the original regularization is the special case with only one scale and hence $W = V$.

3.4 Structure of W

It is interesting to note that W possesses a structure very similar to a RKHS. On V we have for each $x \in \Omega$ and $a \in \mathbb{R}^d$ the evaluation functionals $a \otimes \delta_x(v) = a^T v(x)$. Using the integral map Ψ defined above, we define the linear maps on W

$$a \otimes \delta_x^\Psi(w) := \int_{I_W} a \otimes \delta_x(w_r) dr = \int_{I_W} a^T w_r(x) dr = a \otimes \delta_x(\Psi(w))$$

As seen from the equation, the maps evaluate w_r at each scale and integrate the results using Ψ . These maps are continuous and hence in the dual W^* . For the elements $K(\cdot, x)a = \{K(\cdot, x)_r a\}_r \in W$, we have

$$\begin{aligned} \langle K(\cdot, x)a, K(\cdot, y)b \rangle_W &= \int_{I_W} \langle K(\cdot, x)_r a, K(\cdot, y)_r b \rangle_r dr \\ &= \int_{I_W} a^T K_r(x, y) b dr = a^T \int_{I_W} K_r(x, y) b dr \\ &= a \otimes \delta_x^\Psi(K(\cdot, y)b) = a^T \Psi(K(x, y)b) \end{aligned}$$

which is similar to the reproducing property in V except for the integration performed by Ψ on the right-hand side of the equation. Also, close to the RKHS situation, we see that

$$\begin{aligned} \langle K(\cdot, x)a, w \rangle_W &= \int_{I_W} \langle K(\cdot, x)_r a, w_r \rangle_r dr \\ &= \int_{I_W} a^T w_r(x) dr = a \otimes \delta_x^\Psi(w), \quad w \in V \end{aligned}$$

again with the integration of w occurring in $a \otimes \delta_x^\Psi(w)$.

3.5 Multi-Scale Representation and Relation to other Approaches

With the kernel bundle, the momentum components can vary over scale, and any combination of small and large scale features at each spatial location can be represented. In particular, the bundle allows sparse priors to force vanishing momentum at one scale while allowing it to be non-zero at other scales at the same position. The effect is to allow representing deformation compactly with non-zero components only at the right scales.

In contrast to this, the simultaneous coarse and fine method developed by Risser et al. in [13, 14] builds a kernel by summing Gaussians of different scale. This effectively changes only the shape of the kernel and does not allow different momentum at different scales. If momenta vanish, they will vanish at all scales simultaneously, and, therefore, the ability to represent sparsity across scales that we search for here is not possible.

When not using sparse priors and when the L^2 -norm is combined linearly across scales, Bruveris et al. [3] showed that optimal deformations with the kernel bundle coincide with results obtained with the sum of kernels approach. Thus, though the kernel bundle is able to represent sparsity across scales, the cross-scale sparsity will not occur without adding more scale information to the system. This seems challenging with the approach of Risser et al. but adding such information becomes straightforward with the scale decoupled bundle representation which illustrates the descriptive power offered by the kernel bundle. Imposing sparse priors as we pursue later in this paper constitutes an example of this, and, as we will see in the experiments, optimal deformations with a prior do indeed increase the compactness of the representation and exhibit sparsity across scales. Correspondingly, incorporating scale information from the data term can guide the deformation model further towards the right mixture of scales and allow momentum-based statistics [21] to be performed across scale. This is again possible with the decoupled bundle representation, and we are currently pursuing this path.

4 Evolution Equations: Kernel Bundle EPDiff

In the single scale LDDMM case, the EPDiff equations (Euler-Poincaré for diffeomorphism groups) describes the evolution of optimal paths for the registration problem. They are most often formulated in the following continuous form: let $a_t = Lv_t$ denote the momentum at time t and assume that φ_t is a path minimizing $E_1(\varphi)$ with $\varphi_1 = \varphi$ minimizing $E(\varphi)$ and v_t is the derivative of φ_t . Then v_t satisfies the system

$$\begin{aligned} v_t &= \int_{\Omega} K(\cdot, x) a_t(x) dx, \\ \frac{d}{dt} a_t &= -Da_t v_t - a_t \nabla \cdot v_t - (Dv_t)^T a_t. \end{aligned}$$

The first equation connects the momentum a_t with the velocity v_t , and the second describes the evolution of the momentum. The EPDiff equations can be interpreted as geodesic equations on the manifold G_V , and they are important for implementations since they limit the search for optimal paths to paths satisfying the system.

As we will show in this section, there exists similar equations with the kernel bundle: if $\Psi(w_t)$ is the derivative of

the path of diffeomorphisms φ_t minimizing (7) with $\varphi = \varphi_1$ minimizing (7) then

$$\begin{aligned} w_{r,t} &= \int_{\Omega} K_r(\cdot, x) a_{r,t}(x) dx, \\ \frac{d}{dt} a_{r,t} &= \int_{I_W} -D a_{r,t} w_{s,t} - a_{r,t} \nabla \cdot w_{s,t} - (D w_{s,t})^T a_{r,t} ds. \end{aligned} \quad (8)$$

with $a_{r,t}$ being the momentum for the part $w_{r,t}$ of w_t . In essence, the standard EPDiff equations are integrated over the parameter space I_W to obtain the evolution of the momentum at each scale, and, in particular, the result will imply that the momentum conservation property of LDDMM also holds in kernel bundle case. We will derive the KB-EPDiff equations in a more general form which implies the above formulation, and, for doing this, we will follow the strategy in [22] for the LDDMM case.

4.1 Euler-Lagrange equations

For any time varying path w_t in W , we denote by $\varphi_{t_1 t_2}^{\Psi(w)}$ the diffeomorphism obtained by integrating $\Psi(w_t)$ from time t_1 to time t_2 . The end of the integrated path $\varphi_{01}^{\Psi(w)}$ is the diffeomorphism used for the registration. For the energy $E(w_t) = E_1(w_t) + \lambda U(\varphi_{01}^{\Psi(w)})$, we consider a variation $h_t \in W$ and calculate

$$\begin{aligned} \frac{d}{d\varepsilon} E(w_t + \varepsilon h_t)|_{\varepsilon=0} &= \\ 2 \int_0^1 \langle w_t, h_t \rangle_W dt + \lambda \frac{d}{d\varepsilon} U(\varphi_{01}^{\Psi(w) + \varepsilon \Psi(h)})|_{\varepsilon=0}. \end{aligned} \quad (9)$$

Following [22], we define $\text{Ad}_{\varphi} v(x) = (D\varphi v) \circ \varphi^{-1}(x)$ for $v \in V$ and get a functional Ad_{φ}^* on the dual V^* of V by $(\text{Ad}_{\varphi}^* \rho | v) = (\rho | \text{Ad}_{\varphi}(v))$.² It is shown in [22] that a variation \tilde{h}_t in V of the match functional satisfies

$$\frac{d}{d\varepsilon} U(\varphi_{01}^{v+\varepsilon \tilde{h}})|_{\varepsilon=0} = \int_0^1 (\text{Ad}_{\varphi_t}^* \bar{\partial} U(\varphi_{01}^v) | \tilde{h}_t) dt$$

with $\bar{\partial} U$ denoting the Eulerian differential of U (see [22, Chap. 10]). Inserting into (9) gives

$$\begin{aligned} \frac{d}{d\varepsilon} E(w_t + \varepsilon h_t) &= \\ 2 \int_0^1 \langle w_t, h_t \rangle_W dt + \lambda \int_0^1 (\text{Ad}_{\varphi_t}^* \bar{\partial} U(\varphi_{01}^{\Psi(w)}) | \Psi(h_t)) dt. \end{aligned} \quad (10)$$

For each r , we define the operator $\text{Ad}_{\varphi}^{T,r} v = K_r(\text{Ad}_{\varphi}^*(L_r v))$ which then satisfies $\langle \text{Ad}_{\varphi}^{T,r} v, w_r \rangle = (\text{Ad}_{\varphi}^*(L_r v) | w_r)$, and we can now derive the fundamental results [22, Prop. 11.6/Cor. 11.7] in the bundle case:

² With a functional $f \in V^*$ and a vector $v \in V$, the notation $(f|v)$ denotes f evaluated on v , i.e. $(f|v) = f(v)$.

Proposition 1 *If w_t is an optimal path for E then for almost every $r \in I_W$,*

$$w_{t,r} = \text{Ad}_{\varphi_{t1}}^{T,r} w_{1,r}$$

$$\text{with } w_{1,r} = -\frac{1}{2} \nabla^{V_r} U(\varphi_{01}^{\Psi(w)}).$$

Proof Assume instead that there exists a time varying h_t in W and $t \in [0, 1]$ such that

$$\begin{aligned} 0 &< \int_{I_W} \langle w_{t,r} - \text{Ad}_{\varphi_{t1}}^{T,r} w_{1,r}, h_{t,r} \rangle_r dr \\ &= \int_{I_W} \langle w_{t,r}, h_{t,r} \rangle_r dr - \int_{I_W} \langle \text{Ad}_{\varphi_{t1}}^{T,r} w_{1,r}, h_{t,r} \rangle_r dr \\ &= \langle w_t, h_t \rangle + \frac{1}{2} \int_{I_W} (\text{Ad}_{\varphi_{t1}}^* \bar{\partial} U(\varphi_{01}^{\Psi(w)}) | h_{t,r}) dr \\ &= \langle w_t, h_t \rangle + \frac{1}{2} (\text{Ad}_{\varphi_{t1}}^* \bar{\partial} U(\varphi_{01}^{\Psi(w)}) | \Psi(h_t)). \end{aligned}$$

But the right hand side vanishes for all t and all h_t by (10) and the fact that w_t is optimal for E , a contradiction.

Corollary 1 *Under the same conditions, for almost every $r \in I_W$,*

$$w_{t,r} = \text{Ad}_{\varphi_{t0}}^{T,r} w_{0,r}. \quad (11)$$

The proof of the corollary is identical to the proof of [22, Cor. 11.7].

4.2 Scale Conservation and KB-EPDiff

In the kernel bundle, the momentum of a path may differ across scales. For a path w_t in W , we let a_t be the bundle momentum defined by $a_{t,r} = L_r(w_{t,r})$ recalling that L_r is the momentum operator at scale r . For each t , we can consider a_t to be in the dual W^* by $(a_t | \tilde{w}) = \int_{I_W} (a_{t,r} | \tilde{w}_r) dr$ which is continuous since

$$|(a_t | \tilde{w})| \leq \left| \int_{I_W} (a_{t,r} | \tilde{w}_r) dr \right| = \left| \int_{I_W} \langle w_{t,r}, \tilde{w}_r \rangle_r dr \right| \leq \|w_t\| \|\tilde{w}\|.$$

Suppose now w_t satisfies the transport equation (11) for almost every $r \in I_W$. Then for all $\tilde{w} \in W$,

$$\begin{aligned} (a_t | \tilde{w}) &= \int_{I_W} \langle w_{t,r}, \tilde{w}_r \rangle_r dr = \int_{I_W} \left\langle \text{Ad}_{\varphi_{t0}}^{T,r} w_{0,r}, \tilde{w}_r \right\rangle_r dr \\ &= \int_{I_W} \left\langle w_{0,r}, \text{Ad}_{\varphi_{t0}}^* \tilde{w}_r \right\rangle_r dr = (a_0 | \text{Ad}_{\varphi_{t0}}^* \tilde{w}) \end{aligned} \quad (12)$$

where $\text{Ad}_{\varphi_{t0}}^* \tilde{w}$ is the element of W obtained by applying $\text{Ad}_{\varphi_{t0}}^*$ to each \tilde{w}_r . The above equation shows that the momentum at time t is completely specified by the momentum

at time 0 and thus reproduces the momentum conservation property for LDDMM. Note that since \tilde{w} can be chosen arbitrarily in (12), the momentum is conserved for each scale separately. By differentiating $\text{Ad}_{\varphi_{t_0}^{\Psi(w)}} \tilde{w}$, the momentum conservation property directly implies the equation

$$\partial_t (a_t | \tilde{w}) = - (a_t | D\Psi(w_t) \tilde{w} - D\tilde{w} \Psi(w_t)) \quad (13)$$

or, equivalently,

$$\partial_t a_t + \text{ad}_{\Psi(w_t)}^* a_t = 0$$

with $(\text{ad}_{\Psi(w_t)}^* a_t | \tilde{w}) = (a_t | D\Psi(w_t) \tilde{w} - D\tilde{w} \Psi(w_t))$. Both equations imply the system (8) and extend the EPDiff equations for LDDMM. We denote them KB-EPDiff.

An important difference from the single-scale framework relates to the energy along optimal paths. The relation to geodesics in LDDMM suggests that the norm $\|v_t\|_V$ is constant in t when v_t is optimal for $E_1(\varphi)$. This is in fact the case for LDDMM. With the kernel bundle, momentum is conserved along optimal paths of $E_1(\varphi)$ though $\|w_t\|_W$ is not constant. This occurs because the new energy is not directly related to a metric in the Riemannian sense.

4.3 KB-EPDiff for Landmarks: An Example

To give a concrete application of the KB-EPDiff equations, we redo the calculation for LDDMM landmark matching with scalar kernels to arrive at the corresponding system with the bundle. The initial momentum $a_{0,r}$ will in this case be supported at the N landmarks x_i , $i = 1 \dots, N$, i.e. $a_{0,r} = \sum_{i=1}^N a_{0,r,i} \otimes \delta_{x_i}$ with vectors $a_{0,r,i} \in \mathbb{R}^d$. We let $x_{t,i}$ denote the trajectory of the i th landmark so that $x_{t,i} = \varphi_{0t}^{\Psi(w)}(x_{0,i})$.

Letting $a_{t,r,i} = (D\varphi_{t_0}^{\Psi(w)})^T a_{0,r,i}$, we get from (12)

$$\begin{aligned} (a_{t,r} | \tilde{w}) &= \left(\text{Ad}_{\varphi_{t_0}^{\Psi(w)}}^* \left(\sum_{i=1}^N a_{0,r,i} \otimes \delta_{x_{0,i}} \right) \middle| \tilde{w} \right) \\ &= \left(\sum_{i=1}^N a_{0,r,i} \otimes \delta_{x_{0,i}} \middle| \text{Ad}_{\varphi_{t_0}^{\Psi(w)}}(\tilde{w}) \right) \\ &= \sum_{i=1}^N a_{0,r,i}^T (D\varphi_{t_0}^{\Psi(w)} \tilde{w}) \circ \varphi_{0t}^{\Psi(w)}(x_{0,i}) \\ &= \left(\sum_{i=1}^N a_{t,r,i} \otimes \delta_{x_{t,i}} \middle| \tilde{w} \right). \end{aligned}$$

Since $\frac{d}{dt} (D_{x_{t,i}} \varphi_{t_0}^{\Psi(w)})^T = -D_{x_{t,i}} \Psi(w_t)^T (D_{x_{0,i}} \varphi_{t_0}^{\Psi(w)})^T$, the derivative of the momentum satisfies

$$\frac{d}{dt} a_{t,r,i} = \frac{d}{dt} ((D\varphi_{t_0}^{\Psi(w)})^T a_{0,r,i}) = -D_{x_{t,i}} \Psi(w_t)^T a_{t,r,i}.$$

The trajectories of the landmarks and momentum evolution is therefore completely described by the system

$$\begin{aligned} \Psi(w_t) &= \int_{I_W} \sum_{l=1}^N K_r(\cdot, x_{t,l}) a_{t,r,l} dr \\ \frac{d}{dt} a_{t,r,i} &= - \left(\int_{I_W} \sum_{l=1}^N D_1(K_s(x_{t,i}, x_{t,l}) a_{t,s,l})^T ds \right) a_{t,r,i} \\ x_{t,i} &= \varphi_{0t}^{\Psi(w)}(x_{0,i}). \end{aligned} \quad (14)$$

Note that the system is finite if I_W is finite.

5 Sparse Kernel Bundle Representation

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. [10] 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 λ_{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 (7) to the energy

$$E(\varphi) = E_1(\varphi) + \lambda U(\varphi) + \lambda_{sp} \sum_{l=1}^N f(\|a_{0,l}\|) \quad (15)$$

with $f = f_{\log^c}$. 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 (4) 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 momentum take the form

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

This reduces to the number of non-zero momentum vectors $\|w\|_{L^0} = \int_{I_W} |\{L w_r \neq 0\}| dr$ 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 (15) by

$$E(\varphi) = E_1(\varphi) + \lambda U(\varphi) + \sum_{r=1}^R \lambda_{sp,r} \sum_{l=1}^N f(\|a_{0,l,r}\|) \quad (16)$$

and we require the evolution of $a_{t,l,r}$ to follow the KB-EPDiff equations. Here $\lambda_{sp,r}$ 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 (16) *sparse kernel bundle* registration.

5.1 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 (7) 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 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} [9]. Candès et al. [4] proposes several penalty functions including the function

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

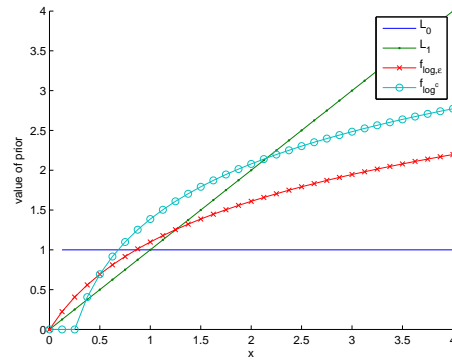


Fig. 3 The L_0 -norm for real valued data, and the approximations L_1 , f_{\log^c} , and $f_{\log,\varepsilon}$ with $\varepsilon = 0.5$ and $c = 0.25$. The L_1 -norm provides poor approximation for large values. The truncated log f_{\log^c} is not non-zero for small values.

Figure 3 illustrates the approximation of the L_0 -norm provided by the L_1 -norm, f_{\log^c} , and $f_{\log,\varepsilon}$. Both f_{\log^c} and $f_{\log,\varepsilon}$ suffer less from the poor approximation for large momenta. Both necessitates a choice of parameter, c or ε . Though $f_{\log,\varepsilon}$ may seem more natural than f_{\log^c} which is zero for small values, the gradient of $f_{\log,\varepsilon}$ 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 [10].

5.2 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 kernel bundle construction is made explicitly to allow independent velocity at the different scales, and the behaviour of sparsity across scales is allowed by this fact. As we will see in the experiments, optimal deformations computed with a sparse prior do indeed exhibit sparsity across scales.

The weights $\lambda_{sp,r}$ should ideally be chosen by cross-validation in same way the weight λ in (7) and the weighting between scales in the bundle are determined. At this point, we heuristically choose $\lambda_{sp,r}$ either constant in r or $\lambda_{sp,r} = \lambda_{sp}/r^\alpha$ for a fixed scalar λ_{sp} and exponent $\alpha \geq 0$ in order to compensate for the often larger momenta at small scales.

6 Implementation

We here describe how optimal registrations with the kernel bundle can be computed in the case of landmark matching. Extending the method to images using a control point

formulation similar to [10,9] and [17] poses no conceptual problem.

The running time will primarily be dominated by the backwards gradient transport described below. The system parallelizes well and can be implemented on GPU hardware [15]. The cost of adding a sparsity prior and computing its gradient is insignificant compared to the cost of integrating the flow equations described below. We do not experience any substantial differences in the number of iterations of the optimization procedure with and without sparse priors. The computation time is primarily a function of the number of landmarks and the number of included scales.

6.1 Algorithm

Since the evolution of the bundle momentum and velocity are required to follow the KB-EPDiff equations, we can optimize (16) using gradient based optimization strategies. A simple gradient descent scheme will given an guess for the initial momentum a_0 calculate the gradient $\nabla E(w_0) = \nabla E_1(w_0) + \lambda \nabla U(w_0)$ using $w_0 = K a_0$, add the gradient from the sparsity term if using sparse prior, and update a_0 by adding a vector proportional to the gradient. In practice, we use limited-memory BFGS updates³.

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 ϕ before flowing the gradient at time $t = 1$ backwards to obtain the gradient $\nabla E(w_0)$ at $t = 0$. The gradient $\nabla U(w_1)$ at $t = 1$ is provided by the similarity measure; 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.

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. The systems are described in detail below.

The sparse penalty functions considered here have gradients

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

If applying ∇f_{\log^c} , $\|a_{0,l,r}\|$ is considered zero if it is less than c in which case we do not add the gradient to $\nabla E(w_0)$. 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.

³ See e.g. <http://www.di.ens.fr/~mschmidt/Software/minFunc.html>.

6.2 Forward and Backward Transport

The diffeomorphism ϕ is determined by w_0 by the KB-EPDiff equations, and the forward transports integrates the KB-EPDiff system (14) to generate ϕ . The system is a non-linear ODE with w_0 and the point positions x_1, \dots, x_N as initial values.

Because w_0 through the evolution of w_t is uniquely linked to w_1 , $U(\phi)$ can in addition be considered a function of w_1 . The gradient $\nabla U(w_0)$ can be obtained by differentiating (14) and solving the transpose system backwards with $\nabla U(w_1)$ as initial condition. This approach is described in the single-scale case in [22]. With multiple scales, the gradient $\nabla E_1(w_0)$ can be computed simultaneously with $\nabla U(w_1)$ 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

$$\dot{y}_t = v_t + M_t y_t \quad (17)$$

integrated from $t = 1$ to $t = 0$. The linear component transports $\nabla U(w_t)$ while the affine component transport $\nabla E_1(w_t)$. We provide explicit form of this system below.

As in Section 4.3, we let $x_{t,i}$ denote the point positions at time t and the set of time-dependent vectors $a_{t,r,i}$ is the momentum of the flow. These components are computed from the forward integration of the KB-EPDiff equations (14). Note that the momenta have components at each scale r . By differentiating the KB-EPDiff equations we obtain the linear ODE

$$\dot{y}_t = B_t y_t. \quad (18)$$

The matrix M_t in the backwards equations (17) arises as the transpose of the matrix B_t . Both systems (18) and (17) have components coding the variation in point positions and momentum, respectively. We denote these components $b_{t,k}^x$ and $b_{t,k,r}^a$ for (18) and $m_{t,k}^x$ and $m_{t,k,r}^a$ for (17). Here $k = 1, \dots, N$ and we consider the case of a finite set of scales R so that $r = 1, \dots, R$. We assume the kernel K is scalar $K(x, y) = \gamma(\|x - y\|^2) \text{Id}_d$ with a real function γ and write $\gamma_{t,kl}$ for $\gamma(\|x_{t,k} - x_{t,l}\|^2)$. Differentiating (14) then provides the components of (18):

$$\begin{aligned} b_k^x &= \sum_{r=1}^R \sum_{l=1}^N \gamma_{kl}^r b_{l,r}^a + 2 \sum_{r=1}^R \sum_{l=1}^N \gamma_{kl}^r (x_k - x_l)^T (m_k^x - m_l^x) a_{l,r} \\ b_{k,s}^a &= -2 \sum_{r=1}^R \sum_{l=1}^N \gamma_{kl}^r (a_{k,r}^T m_{l,r}^a + a_{l,r}^T m_{k,s}^a) (x_k - x_l) \\ &\quad - 2 \sum_{r=1}^R \sum_{l=1}^N \gamma_{kl}^r a_{k,s}^T a_{l,r} (m_k^x - m_l^x) \\ &\quad - 4 \sum_{r=1}^R \sum_{l=1}^N \gamma_{kl}^r a_{k,s}^T a_{l,r} (x_k - x_l)^T (m_k^x - m_l^x) (x_k - x_l) \end{aligned}$$

where we omitted the time dependence of all terms to keep the notation compact. By transposing B_t , we get M_t and hence the linear parts of (17). This is in components

$$\begin{aligned} m_k^x &= -2 \sum_{r=1}^R \sum_{l=1}^N \dot{\gamma}_{kl}^r (a_{k,r}^T m_l^x + a_{l,r}^T m_k^x) (x_k - x_l) \\ &+ 2 \sum_{r,r'=1}^R \sum_{l=1}^N \dot{\gamma}_{kl}^{r'} (a_{k,r}^T a_{l,r'} m_{k,r}^a - a_{k,r'}^T a_{l,r} m_{l,r}^a) \\ &+ 4 \sum_{r,r'=1}^R \sum_{l=1}^N \dot{\gamma}_{kl}^{r'} (x_k - x_l)^T (a_{k,r}^T a_{l,r'} m_{k,r}^a - a_{k,r'}^T a_{l,r} m_{l,r}^a) (x_k - x_l) \\ m_{k,s}^a &= - \sum_{l=1}^N \gamma_{kl}^s m_l^x + 2 \sum_{r=1}^R \sum_{l=1}^N (x_k - x_l)^T (\dot{\gamma}_{kl}^r m_{k,r}^a - \dot{\gamma}_{kl}^s m_{l,r}^a)_{l,r}. \end{aligned}$$

The simpler affine term has components

$$\begin{aligned} m_k^x &= -4 \sum_{r=1}^R \sum_{l=1}^N \dot{\gamma}_{kl}^r a_{k,r}^T a_{l,r} (x_k - x_l) \\ m_{k,s}^a &= - \sum_{l=1}^N 2 \gamma_{kl}^s m_{l,s}^a. \end{aligned}$$

Letting $m_{1,k}^x$ equal the k th component of $U(w_1)$ and setting $m_{1,k,r}^a$ to zero provides the initial conditions for the system. After backwards integration, the components $m_{0,k,r}^a$ contain $\nabla E(a_0)$ providing $\nabla E(w_0)$ using $w_0 = K a_0$.

7 Experiments

We perform four sets of experiments to illustrate and test the behaviour of the kernel bundle method and its ability to support sparsity across scales. We start with a simple synthetic example to visually illustrate the differences between the single and multi-scale, sparse and non-sparse methods. In particular, will see that sparsity is achieved at the different scales individually. We then present landmark based examples of matching hand outlines to test the methods ability to represent both small and large scale features, and we illustrate the differences in the evolution in the diffeomorphism manifold when matching with LDDMM and the kernel bundle with sparse prior. Finally, we apply the method to register annotated lung CT scans, and we show that the extra capacity of the method does not affect its ability to generalize to test data; that manual scale selection is not necessary with the multi-scale method; and that we can control the sparsity across scales by varying the weight of the sparse prior.

7.1 Synthetic Example

Figure 1 presents a simple example which illustrates the effect of fusing sparsity and multiple scales. In the figure, we show the results of matching two sets of 11 points using

the sparse kernel bundle method and the kernel bundle without sparse prior together with results when using sparse and non-sparse LDDMM 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. However, the fixed kernel scale has 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 with the kernel bundle algorithm without sparse prior but all momenta at all scales are non-zero as shown by the non-zero momentum vectors and the representation is far from compact. The kernel bundle method with sparse prior obtains the best of both worlds: even with vanishing momenta 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.

7.2 Hand Outlines

We consider the hand outlines shown in Figure 4. Using the landmarks (red dots) on the moving hand image, we wish to compute the kernel bundle match against the landmarks on the fixed image (black dots). The match is computed with three scales ($\sigma = 8, 4, 2$ in units of the grid overlaid the figures). Figure 4 shows the results of computing the match with the kernel bundle together with results obtained with single-scale LDDMM with each of the three scales separately. For LDDMM with the largest scale, the match is poor and the sharp bend of the thumb is especially badly modelled. The situation improves for the middle scale though the bend of the thumb is still not sufficiently sharp and the match is bad for the middle fingers. For the smallest scales, the thumb is correctly matched but now the smaller scale is not able to model the even movement of the index finger. The kernel bundle method is by including all scales able to correctly register all the critical areas, and, at the same time, it gives the best match of the landmarks.

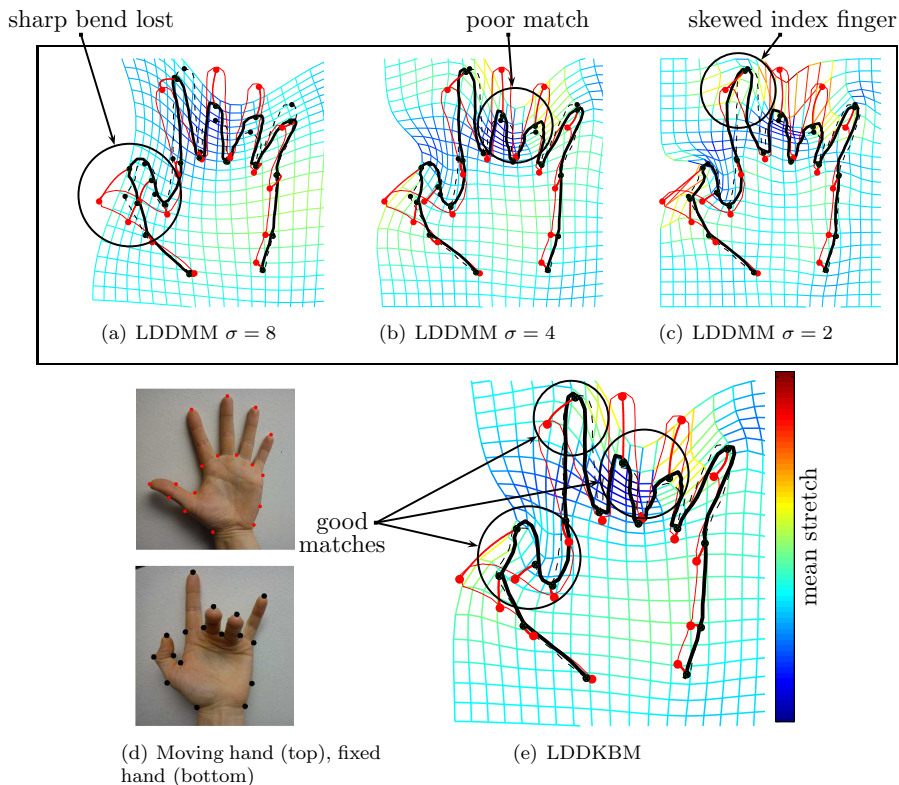


Fig. 4 Matching hands with single- and multiple scales. (d) The moving and fixed hands; (e) result of matching the hands with the kernel bundle method; (a)-(c) results of matching the hands with the single-scale LDDMM method with three different scales separately ($\sigma = 8, 4, 2$). The red landmarks of the moving hand are matched against the black landmarks of the fixed hand. The outline of the moving hand (red line) is transported to the black outline and should be compared with the outline of the fixed image (black dashed). The kernel bundle method is by incorporating movement at the multiple scales able to correctly match the critical areas on which LDDMM fails.

7.3 KB-EPDiff Across Scales

To illustrate the difference in the evolution of critical paths with the kernel bundle and LDDMM, we match in Figure 5 again eleven points (red) against eleven points (black) with results (green) using both LDDMM and kernel bundle method with two scales and enforced sparsity. In the figure, the results of the two registrations are visible in row 1 and 2 right-most, and the evolution of the critical paths generated by the EPDiff and KB-EPDiff equations are shown with time increasing across columns. The lower rows display the deformation obtained with the kernel bundle separated for each scale. We see how most of the transport occurs at the largest scale while the lowest scale perform almost no horizontal movement but takes care of the fine adjustment allowing the kernel bundle method to obtain a good match. The sparse prior forces compactness in the representation and spatial locality of the fine scale movement.

7.4 Annotated Lung CT Registration

We now test the sparse kernel bundle on the publicly available DIR [5] dataset of lung CT images and manually an-

notated landmarks. We aim to show that the extra capacity of the method does not affect its ability to generalize to test data; that manual scale selection is not necessary with the multi-scale method; and that we can control the sparsity across scales by varying the weight of the sparse prior. We note important differences between the experiments performed in the conference papers [18, 19]: we test with sparse priors, and we use the fast algorithm with the backwards gradient transport developed in this paper to allow cross-validation tuning of the regularization term λ in the energies (1) and (7). Thus, we are able to remove the influence of λ on the experiments. In addition, we use isotropic kernels and include more points in the experiments resulting in markedly lower test errors and more robust evaluation.

The dataset consists of five cases of CT images for different stages of the inhale and exhale phases and annotated landmarks for the maximum inhale and exhale phases, respectively. The images and landmarks are available on grids with voxel size varying slightly between the cases but close to $0.6 \times 0.6 \times 2.5$ 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

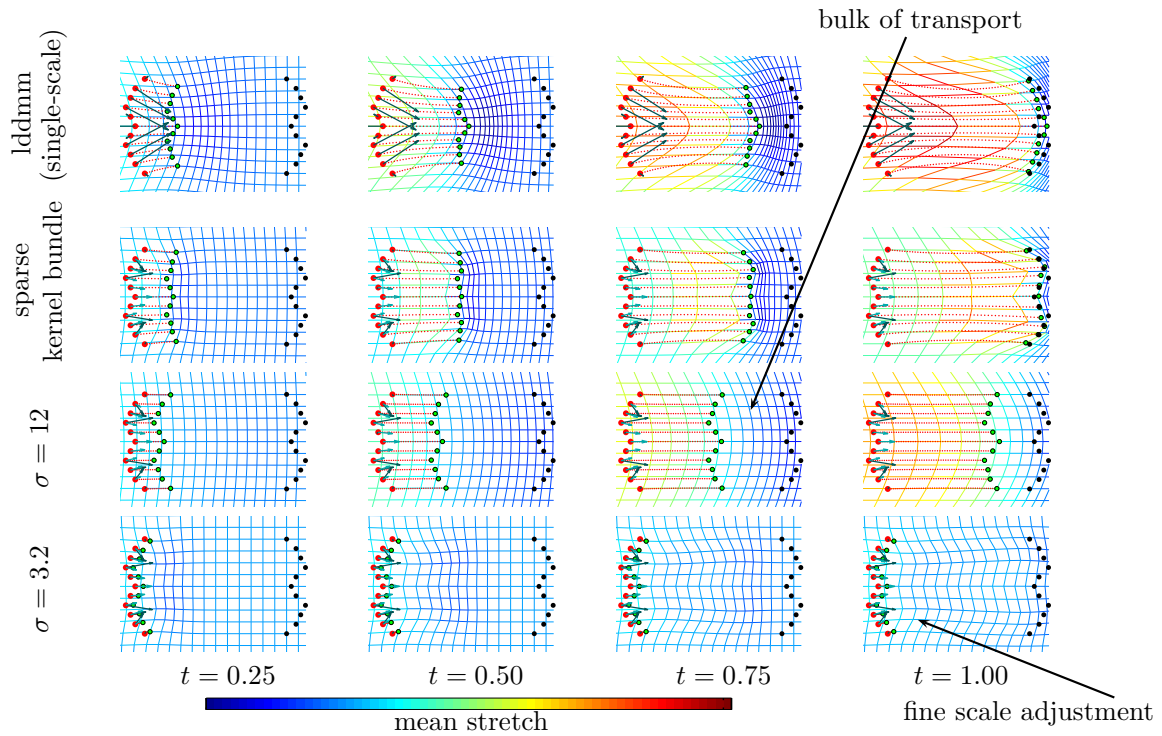


Fig. 5 LDDMM and sparse kernel bundle match of landmarks (red) to landmarks (black) with results (green). The kernels are Gaussian with the kernel bundle applying two scales. Time points of the critical paths are shown along the horizontal axis with the rightmost subfigures displaying the final deformation. (top row) Critical path determined by EPDiff equations with LDDMM (single scale); (row 2) critical path determined by KB-EPDiff equations with the sparse kernel bundle method; (row 3-4) individual contribution of each of the bundle scales (scale σ in grid units). Initially square grids are shown deformed by the diffeomorphism, and the grids are colored with the trace of Cauchy-Green strain tensor indicative of the mean stretch (log-scale for each row individually). With the sparse kernel bundle method, the largest scale contribute to most of the transport movement with smooth deformation while the smallest scale performs fine adjustment of the trajectories to obtain a good match. The sparse prior forces compactness in the representation and spatial locality of the fine scale movement.

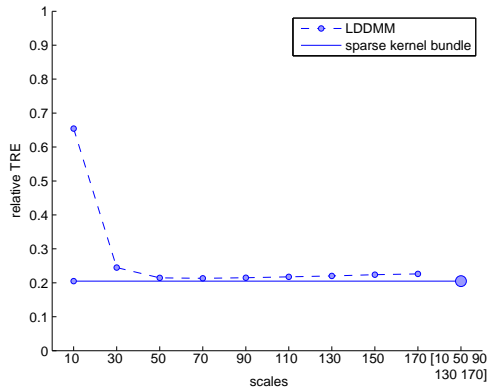
will drive the registration using random subsets of these landmarks, and evaluate the computed match using the target registration error (TRE) measured on the landmarks not used to drive the registration.

To compare single scale LDDMM and the sparse kernel bundle method, we choose random subsets of 200 landmarks to drive the registration, and for each such choice of subset S and each of the five patient cases, we compute the TRE $(\sum_{j \notin S} \|\varphi(x_j^I) - x_j^E\|^2)^{1/2}$. We find the relative size of the TRE against the value before the match (relative TRE), and average over the patients and different choices of subsets. This setup is performed for LDDMM with Gaussian kernels with scale ranging between 10mm and 170mm and with the kernel bundle method with five scales evenly distributed in the same range ($\sigma = 10, 50, 90, 130, 170$ mm.).

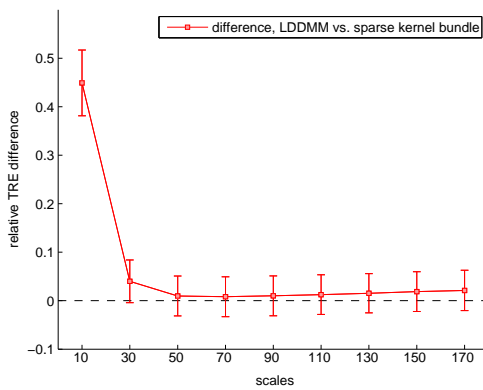
For each choice of random subset S , we tune the regularization term λ used in the energies (1) and (7) using cross-validation on further subsets of S . This ensures that possible variation in the effect of λ on the single- and multi-scale methods does not influence the presented results. For the kernel bundle, we let each scale contribute with equal weight. For this experiment, we let the sparsity weight be

$\lambda_{sp} = 0.02$, and we let the prior vary over scale by $\lambda_{sp,r} = \lambda_{sp}/r$. The result of the experiment is not affected when λ_{sp} varies within a reasonable range of the chosen value, and we further explore the choices of λ_{sp} and $\lambda_{sp,r}$ below.

In Figure 6(a), we see that with single-scale LDDMM the TRE decreases with increasing scale up to a scale of 70mm after which it starts increasing. This indicates that a kernel scale of 70mm will be appropriate for LDDMM. As displayed in Figure 6(b), the sparse kernel bundle method attains an error lower than but within one standard deviation of the best LDDMM result. Without tuning for scale, the sparse kernel bundle method is thus as good as LDDMM, and classical scale selection by cross validation is not needed with the multi-scale method. Furthermore, the results indicate that the same quality of match can be reached with less data since we potentially could use the entire dataset to drive the registration with the kernel bundle. Manual scale selection will allow only a part of the data as input for the registration as the rest is needed to select the kernel scale. The experiment shows in addition that the extra capacity and additional degrees of freedom of the kernel bundle do



(a) LDDMM and sparse kernel bundle relative errors



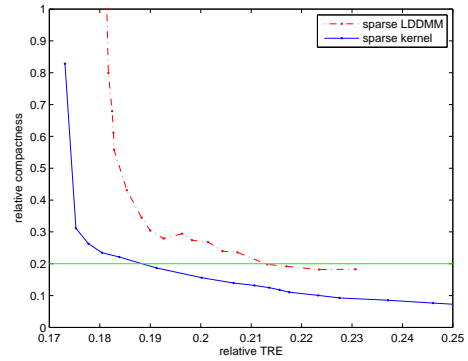
(b) Error differences

Fig. 6 LDDMM and the sparse kernel bundle method: (a) average relative TRE for different kernel scales (LDDMM) and the sparse kernel bundle method (horizontal line and rightmost). Zero relative error indicates perfect match and a relative error of one indicates no error reduction. Labels on the horizontal axis are kernel scale in mm (9 different scales for LDDMM and the interval [10 170] discretized in five scales for the sparse kernel bundle). (b) relative TRE for LDDMM (scale in mm on horizontal axis) subtracted the respective relative errors with the sparse kernel bundle again matching with the scale interval [10 170] discretized in five scales. Positive values show superior performance of the kernel bundle method. Error bars show standard deviation of the results.

not reduce the ability of the method to generalize to the test data.

The computation time of each registration with a non-optimized CPU-only implementation of the algorithm increases from approximately 5 minutes with one scale to approximately 50 minutes with five scales. This increase can, however, be removed using GPU acceleration as shown in [15]. We further discuss computation time in Section 8.

To evaluate the effect of applying sparse priors with both single and multiple scales, we compare LDDMM and the kernel bundle both with sparse priors. We fix the regularization term to $\lambda = 8$ and average over all 5 patients and several randomly selected subsets of points to drive the match. We provide LDDMM with the advantage by selecting an already



(a) Relative compactness as a function of relative TRE

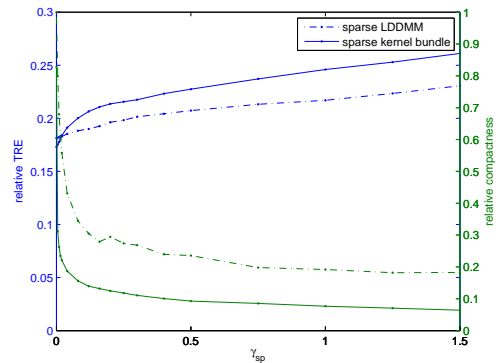
(b) Relative TRE and relative compactness as a function of the sparsity weight γ_{sp}

Fig. 7 Sparse LDDMM and the sparse kernel bundle method: (a) relative TRE (horizontal axis) versus relative compactness (vertical axis) when the varying sparsity weight λ_{sp} ; (b) relative TRE increase (left axis, blue) and relative decrease in non-zero momenta (right axis, green) as function of λ_{sp} . Results averaged over 5 patients. With a factor 5 reduction in non-zero parameters (horizontal line, top), relative TRE for sparse kernel bundle registration is 0.188 in contrast to 0.213 for sparse LDDMM. Though the kernel bundle method achieves the lowest relative TRE, the single scale method obtains the lowest total number of parameters for a given relative TRE.

tuned best possible scale of $\sigma = 50$ mm with these parameters, and we test against the kernel bundle method using five scales in the range 10mm to 170mm as in the previous experiment ($\sigma = 10, 50, 90, 130, 170$ mm.). Again, we let the scale parameters for the sparse prior vary by $\lambda_{sp,r} = \lambda_{sp}/r$. The value of $\lambda_{sp,50} = \lambda_{sp}/50$ is used for LDDMM in accordance with the choice scale.

With this setup, Figure 7 shows the connection between relative TRE, the sparsity weight λ_{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 kernel bundle with only slightly increasing TRE. The sparse kernel bundle method obtains the largest reduction of non-zero parameters for a given increase in relative TRE. Though the kernel bundle

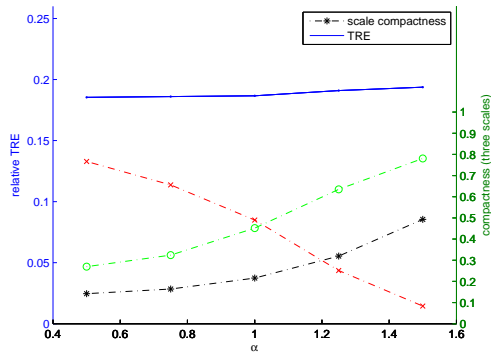


Fig. 8 Sparsity across scales: relative TRE (left axis, blue) and relative decrease in non-zero momenta for each of three scales (right axis, red (x)/green (o)/black (*)) as function of the scale weighting α averaged over the 5 patients. The red (x) marked curve shows compactness for the smallest scale ($\sigma = 10$), 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. In particular, the figure shows that sparsity across scales is achieved.

method achieves the lowest relative TRE, sparse LDDMM still provides the smallest number of total parameters for a fixed relative TRE. This fact should be viewed in the light that the sparse LDDMM method is already tuned to the best scale, and that the kernel bundle has more degrees of freedom than LDDMM. 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 λ_{sp} .

The weighting of the sparsity parameter across scales can be controlled by letting $\lambda_{sp,r} = \lambda_{sp}/r^\alpha$ and varying α . To explore this and the resulting cross-scale effects, we select $\lambda_{sp} = 0.05$, and plot the relative TRE against α in Figure 8. In addition, the figure shows how the distribution of non-zero parameters at the different scales varies with α . To give a clear visual picture of this effect, we use only three scales ($\sigma = 10, 50, 90$ mm.). The increased penalty at small scales for $\alpha > 1$ and corresponding increased penalty for large scales for $\alpha < 1$ is clearly visible. Indeed, the difference in the number of non-zero parameters at the different scales shows that sparsity across scales is achieved.

To illustrate the result of one lung registration with the sparse kernel bundle method, Figure 9 shows the energy of the initial velocity field for three bundle scales separately. 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.

8 Future Work

As previously discussed, when sparse priors are not used and when the L^2 -norm is combined linearly across scales, the cross-scale sparsity will not occur without adding more information to the system. In this paper, we used sparse priors to address this issue and get the cross-scale sparsity but the scale information may also be obtained more directly by incorporating scale information from the data term. The decoupled kernel bundle representation will support such a construction, and we are currently working on including e.g. image scale information in the framework.

Another way of obtaining scale information is to learn compact models from datasets of populations of images. We expect the sparse kernel bundle method to be particularly powerful when applied to population analysis of e.g. atrophy during Alzheimer’s disease. The single-scale analysis in [10,9] has already showed that the spatial distribution of non-zero momenta can be learned across populations using sparse priors. Learning also the scale distribution with the kernel bundle method can potentially lead to even more compact models. After a learning phase, the extracted low-dimensional model can then be used when registering new data. This approach may both increase statistical power due to the reduced dimensionality and increase the speed of the registration process.

Though the execution time increases when multiple scales are used in the registration, the increased computational load can be offset using e.g. GPU acceleration [15]. In addition, if a very compact model is learned from population datasets, the running time should decrease correspondingly due to the reduced size of the differential systems. This could potentially lead to a registration algorithm that is faster than single-scale LDDMM.

9 Conclusion

The multi-scale kernel bundle framework extends the LDDMM framework by incorporating multiple kernels at multiple scales in the registration. The method allows representing deformation at multiple scales at different spatial locations, and it thereby increases the capacity of the deformation description while supporting application of sparse priors that ensure compact representation. Since the priors are applied independently across the parts of the bundle, the algorithm allows sparsity across scales, and the multiple scales extend the range of deformation the algorithm is able to model significantly. 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.

We visually illustrate the method on synthetic data and show the obtained sparsity across scales. We show the multi-

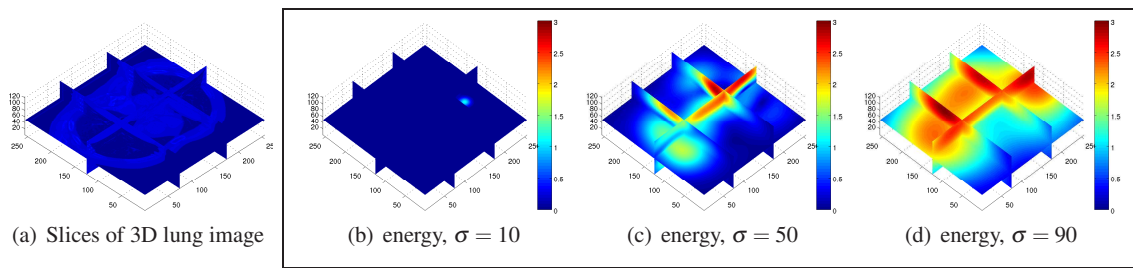


Fig. 9 Slices of 3D lung image and the multi-scale initial vector field at three scales that when combined generate the sparse kernel bundle registration. Left to right: (a) slices of CT image, (b)-(d) squared L^2 -norm of the components at each of the three scales $\sigma = 10, 50, 90$ which in combination make up the multi-scale bundle vector w_0 generating φ 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.

scale effects and cross-scale evolution on additional examples. In addition, when applying the method to a dataset of annotated lung CT images, we demonstrate that the method removes the need for classical scale selection; that sparsity across scales is achieved; that the sparsity may be achieved with only minor increase in registration error; and that the extra capacity of the algorithm does not affect generalization ability.

In addition to the applications demonstrated in this paper, we expect the sparse kernel bundle method to be particularly powerful when applied to population analysis. From both a theoretical and a practical point of view, the sparse kernel bundle framework provides a compact representation of deformation across scales.

References

1. Beg, M.F., Miller, M.I., Trouné, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* **61**(2), 139–157 (2005)
2. Bruveris, M., Gay-Balmaz, F., Holm, D.D., Ratiu, T.S.: The momentum map representation of images. *Journal of Nonlinear Science* **21**(1), 115–150 (2011)
3. Bruveris, M., Risser, L., Vialard, F.: Mixture of kernels and iterated Semi-Direct product of diffeomorphism groups. *Multiscale Model. Simul.* **10**(4), 1344–1368 (2011)
4. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications* **14**, 877–905 (2008)
5. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology* **54**(7), 1849–1870 (2009)
6. Christensen, G., Rabbitt, R., Miller, M.: Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on* **5**(10) (2002)
7. Diestel, J., Uhl Jr., J.J.: Vector measures. American Mathematical Society, Providence, R.I. (1977). *Mathematical Surveys*, No. 15
8. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching (1998)
9. Durrleman, S., Allasonnière, S., Joshi, S.: Sparse adaptive parameterization of variability in image ensembles. *International Journal of Computer Vision* pp. 1–23 (2012)
10. Durrleman, S., Prastawa, M., Gerig, G., Joshi, S.: Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. In: *IPMI 2011*, vol. 22, pp. 123–134 (2011)
11. Fasshauer, G.E., Ye, Q.: Reproducing kernels of generalized sobolev spaces via a green function approach with distributional operators. *Numerische Mathematik* (2011)
12. Grenander, U.: *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford University Press, USA (1994)
13. Risser, L., Vialard, F., Wolz, R., Holm, D.D., Rueckert, D.: Simultaneous fine and coarse diffeomorphic registration: application to atrophy measurement in alzheimer’s disease. *MICCAI 2010* **13**(Pt 2), 610–617 (2010)
14. 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)
15. Sommer, S.: Accelerating Multi-Scale flows for LDDKBM diffeomorphic registration. In: *GPUCV workshop at ICCV 2011*. IEEE (2011)
16. Sommer, S., Lauze, F., Nielsen, M., Pennec, X.: Kernel bundle EPDiff: evolution equations for Multi-Scale diffeomorphic image registration. In: *SSVM 2011*. Springer (2011)
17. Sommer, S., Nielsen, M., Darkner, S., Pennec, X.: Higher-order momentum distributions and locally affine LDDMM registration. To appear in *SIAM Journal on Imaging Sciences*.
18. Sommer, S., Nielsen, M., Lauze, F., Pennec, X.: A Multi-Scale kernel bundle for LDDMM: towards sparse deformation description across space and scales. In: *IPMI 2011*. Springer (2011)
19. Sommer, S., Nielsen, M., Pennec, X.: Sparsity and scale: Compact representations of deformation for diffeomorphic registration. In: *MMBIA at WACV 2012* (2012)
20. Trouné, A.: An infinite dimensional group approach for physics based models in patterns recognition (1995)
21. Vaillant, M., Miller, M., Younes, L., Trouné, A.: Statistics on diffeomorphisms via tangent space representations. *NeuroImage* **23**(Supplement 1), S161–S169 (2004). DOI 10.1016/j.neuroimage.2004.07.023
22. Younes, L.: *Shapes and Diffeomorphisms*. Springer (2010)
23. Younes, L., Arrate, F., Miller, M.I.: Evolution equations in computational anatomy. *NeuroImage* **45**(1, Supplement 1), S40–S50 (2009)