Diffeomorphic registration of discrete geometric distributions

Hsi-Wei Hsieh and Nicolas Charon

Center of Imaging Sciences
Johns Hopkins University
3400 N. Charles St, Clark Hall 317, USA
hhsieh@cis.jhu.edu, charon@cis.jhu.edu

This chapter introduces a new framework and algorithms to address the problem of diffeomorphic registration on a general class of geometric objects that can be described as discrete distributions of local direction vectors. It builds on both the large deformation diffeomorphic metric mapping (LDDMM) model and the concept of oriented varifolds introduced in previous works like [16]. Unlike previous approaches in which varifold representations are only used as surrogates to define and evaluate fidelity terms, the specificity of this paper is to derive direct deformation models and corresponding matching algorithms for discrete varifolds. We show that it gives on the one hand an alternative numerical setting for curve and surface matching but that it can also handle efficiently more general shape structures, including multi-directional objects or multi-modal images represented as distributions of unit gradient vectors.

Contents

1. Introduction

Background

Statistical shape analysis is now regarded across the board as an important area of applied mathematics as it has been and still is the source of quantities of theoretical works as well as applications to domains like computational anatomy, computer vision or robotics. Broadly speaking, one of its central aim is to provide quantitative/computational tools to analyze the variability of geometric structures in order to perform different tasks such as shape comparison or classification.

There are several specific difficulties in tackling such problems in the case of datasets involving geometric shapes. A fundamental one is the issue of defining and computing metrics on shape spaces. A now quite standard approach which was pioneered by Grenander in [13] is to compare shapes through distances based on deformation groups equipped with right-invariant metrics together with a left group action defined on the set of shapes. In this framework, the induced distance is typically obtained by solving a registration problem i.e by finding an optimal deformation mapping one object on the other one. It is thus ultimately determined by the deformation group and its metric for which many models have been proposed. In this paper, we will focus on the Large Deformation Diffeomorphic Metric Mapping (LDDMM) of [4] in which diffeomorphic transformations are generated as flows of time-dependent velocity fields.

Despite the versatility of such models, one of the other common difficulty in shape analysis is the multiple forms or modalities that shapes may take. Looking only at the applications in the field of computational anatomy, if early works have mostly considered shapes given by medical images [19,4] or manually extracted landmarks [15], the variety of geometric structures at hand has considerably increased since then, whether shapes are images acquired through multiple modalities (MRI, CT...) [3], vector or tensor fields as in Diffusion Tensor Images [5], fields of orientation distribution functions [8] or delineated objects like point clouds, curves [12], surfaces [10], fiber bundles [9]...

The intent of this paper is to make a modest step toward one possible generalized setting that could encompass a rich class of shapes including many of

the previous cases within a common representation and eventually lead to a common LDDMM matching framework. Our starting point is the set of works on curve and surface registration based on geometric distributions like measures, currents or varifolds [11, 10, 7]. In the recent article [16] for instance, an oriented curve/surface is interpreted as a directional distribution (known as oriented varifold) of its oriented tangent/normal vectors, which results in simple fidelity terms used in combination with LDDMM to formulate and solve inexact matching problems. Yet all those works so far have restricted the role of distributions' representations to intermediates for the computation of guiding terms in registration algorithms; the underlying deformation model and registration problem remains defined over point sets with meshes.

The stance we take here is to instead introduce group actions and formulate the diffeomorphic matching problem directly in spaces of geometric distributions. In this particular work, we will restrict the analysis to objects in 2D and 3D and focus on the simpler subspace of discrete distributions, i.e that write as finite sums of Dirac varifold masses: Figure 1 gives a few examples of objects naturally represented in this form. We shall consider different models of group actions and derive the corresponding optimal control problems, optimality conditions (Section 3) and registration algorithms (Section 4). This provides, on the one hand, an alternative (and theoretically equivalent) numerical framework to [16] for curve and surface matching using currents, oriented or unoriented varifolds. But the main contribution of our proposed model is that it extends LDDMM registration to the more general class of objects representable by discrete varifolds. In Section 5, we will show several examples of synthetic data besides curves or surfaces that can be treated as such, including cases like multi-directional objects or contrast-invariant images.

Related works.

A few past works share some close connections with the present paper. For instance, [5] develops an approach for registration of vector fields also within the LDDMM setting. The discrete distributions we consider here are however distinct from vector fields as they should rather be interpreted as unlabelled particles at some locations in space with orientation vectors attached (and with possibly varying number of orientation vectors at a single position) as opposed to a field of vectors defined on a fixed grid. In particular, our approach will be naturally framed in the Lagrangian setting as opposed to the Eulerian formulation of [5]. The deformation model and geodesic equations that are derived in Section 3 can be also related to the framework of [20] and [14] that introduce higher-order "jet" particles

and higher-order similarity measures. These remain defined through labelled sets of control points though, which is again different and arguably less flexible than the approach we develop here.

2. Shapes and discrete varifolds

The idea of representing shapes as distributions goes back to the many works within the field of Geometric measure theory. Those concepts have later been of great interest in the construction of simple and numerically tractable metrics between curves or surfaces for registration problems: the works of $[12, 10, 9, 7]$ are a few examples. The framework of oriented varifolds recently exploited in $[16]$ was shown to encompass all those notions into a general representation and provide a wide range of metrics on the spaces of embedded curves or surfaces. We give a brief summary of the latter work below.

In the rest of the paper, we will call an oriented varifold or, to abbreviate, a *varifold* in $\mathbb{R}^n$ (we shall here consider the cases $n = 2$ or $n = 3$) a distribution on the product $\mathbb{R}^n \times \mathbb{S}^{n-1}$. In other words, a varifold μ is by definition a linear form over a certain space W of smooth functions on $\mathbb{R}^n \times \mathbb{S}^{n-1}$, which evaluation we shall write as $\mu(\omega)$ for any test function $\omega \in W$. In all what follows, we shall restrict our focus to 'discrete' shapes and varifolds, leaving aside the analysis of the corresponding continuous models. By discrete varifold, we mean specifically that μ writes as a finite combination of Dirac masses $\mu = \sum_{i=1}^{P} r_i \delta_{(x_i, d_i)}$ with $r_i > 0$, $(x_i, d_i) \in \mathbb{R}^n \times \mathbb{S}^{n-1}$ for all i, in which case $\mu(\omega) = \sum_{i=1}^{P} r_i \omega(x_i, d_i)$ for all ω. Such a μ can be thought as a set of unit direction vectors d_i located at positions x_i with weights (or masses) equal to the r_i's. We assume by convention that the (x_i, d_i) are distinct, but not necessarily that all the positions x_i are: in other words, in our model, there can be more than a single direction vector attached to each position. In the rest of the paper, we will denote by $\mathcal{D}$ the set of all discrete varifolds. Note that in this representation and unlike the cases of landmarks and vector fields, the particles are unlabelled i.e the varifold μ is invariant to any permutation of the (x_i, d_i). One particular subset of interest that we shall denote $\mathring{\mathcal{D}} \subset \mathcal{D}$ is the space of discrete varifolds with distinct positions x_i (or equivalently, the discrete varifolds that carry a single direction vector per point position).

The relationship between shapes and varifolds relies on the fact that discrete shapes, namely curve or surface meshes, can be naturally approximated by varifolds of the previous form. As explained with more details in the aforementioned references, this is done by associating to any cell of the discrete mesh (i.e a segment for curves or a triangular face for surfaces) the weighted Dirac $r_i \delta_{(x_i, d_i)}$ as illustrated in Figure 1. In that expression, x_i is the coordinates of the center of the
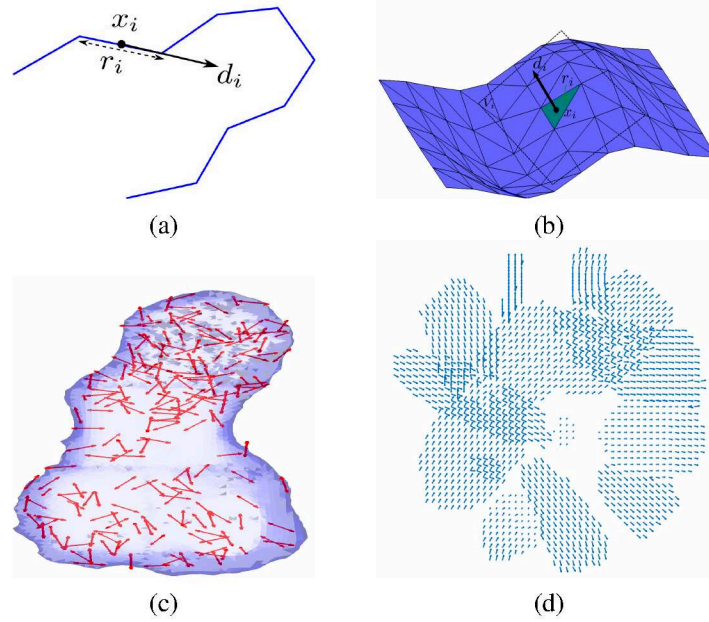
Fig. 1.   Some examples of data representable by discrete varifolds: (a) Piecewise linear curve. (b) Triangulated surface. (c) A set of cells' mitosis directions measured inside a mouse embryonic heart membrane (cf. [18]). (d) Peak diffusion directions extracted from a slice of High Angular Resolution Diffusion Imaging phantom data, note the presence of multiple directions at certain locations corresponding to fiber crossing.

cell, r_i its total length or area and d_i the direction of the tangent space represented by the unit tangent or normal orientation vector d_i. It results in a mapping $S \mapsto \mu_S$ that associates to any discrete shape S the discrete varifold $\mu_S = \sum_{i=1}^{F} r_i \delta_{(x_i, d_i)} \in \overset{\circ}{\mathcal{D}}$ obtained as the sum over all faces $i = 1, \ldots, F$ of the corresponding Diracs.

The main interest of such a representation is that it gives a convenient setting for the definition of shape similarities that are easy to compute and **do not assume pairwise correspondences between points**. With the two conditions, which are quite natural in our context, that W is a Hilbert space and that all Diracs $\delta_{(x,d)}$ for $(x,d) \in \mathbb{R}^n \times \mathbb{S}^{n-1}$ belong to the dual, W then must be a Reproducing Kernel Hilbert Space (RKHS) associated to a smooth positive definite kernel on $\mathbb{R}^n \times \mathbb{S}^{n-1}$. In particular, we will follow the construction proposed in [16] and consider separable kernels of the form $k(x,d,x',d') = \rho(|x-x'|^2)\gamma(\langle d,d' \rangle)$ where ρ and γ define positive definite kernel functions respectively on the positions between particles and the angles between their orientation vectors. The reproducing kernel metric on W then gives a dual metric on varifolds that explicitly writes, for $\mu =$

$\sum_{i=1}^{P} r_i \delta_{(x_i,d_i)}$:

$$\|\mu\|_{W^*}^2 = \sum_{i,j} r_i r_j \rho(|x_i - x_j|^2) \gamma(\langle d_i, d_j \rangle) \tag{2.1}$$

Such metrics on W^* are determined by the choice of the positive definite functions ρ and γ and provide a global measure of proximity between two discrete varifolds. One important advantage for applications to e.g registration is that the computation of a distance $\|\mu - \mu'\|_{W^*}^2$ between two distributions does not require finding correspondences between their masses but instead reduces numerically to a quadratic number of kernel evaluations. The gradients of the metric with respect to the x_i's and d_i's is also very easy to obtain by direct differentiation of (2.1). Finally, we note that the expression in (2.1) is also invariant to the action of the group of rigid motion. Namely for any rotation matrix R, translation vector h and the group action $(R,h) \cdot \mu \doteq \sum_{i=1}^{P} r_i \delta_{(Rx_i+h,Rd_i)}$, one has $\|(R,h) \cdot \mu\|_{W^*} = \|\mu\|_{W^*}$.

In all generality however, (2.1) may only yield a pseudo-metric on the set of discrete varifolds $\mathcal{D}$ since the inclusion mapping $\mathcal{D} \to W^*$ is not necessarily injective. A necessary and sufficient condition is:

Proposition 2.1: *The metric $\|\cdot\|_{W^*}$ on W^* induces a metric on $\mathcal{D}$ if and only if k is a strictly positive definite kernel on $\mathbb{R}^n \times \mathbb{S}^{n-1}$.*

The proof follows immediately from the definition of strictly positive definite kernel. This condition holds in particular if both kernels defined by ρ and γ are strictly positive definite. In the case of $\overset{\circ}{\mathcal{D}}$, one can provide different sufficient conditions which are often more convenient to satisfy in practice. These involve a density property on kernels called C_0-universality, cf [6]. A kernel on $\mathbb{R}^n$ is said to be C_0-universal if the associated RKHS is dense in $C_0(\mathbb{R}^n, \mathbb{R})$. Then one has the following

Proposition 2.2: *If the kernel defined by ρ is C_0-universal, $\gamma(1) > 0$ and $\gamma(u) < \gamma(1)$ for all $u \in [-1,1)$, then $\|\cdot\|_{W^*}$ induces a metric on $\overset{\circ}{\mathcal{D}}$.*

Proof: Let W_{pos} and W_{or} be the RKHS associated to ρ and γ. By contradiction, suppose that $\mu, \mu' \in \overset{\circ}{\mathcal{D}}$ with $\|\mu - \mu'\|_{W^*} = 0$ and $\mu \neq \mu'$ in $\overset{\circ}{\mathcal{D}}$. We can write μ, μ' in the following form:

$$\mu = \sum_{i=1}^{N} r_i \delta_{(z_i,d_i)}, \quad \mu' = \sum_{i=1}^{N} r_i' \delta_{(z_i,d_i')},$$

where $\{z_i\}$, with z_i all distinct, is the reunion of point positions from both distributions and $\max_{1 \leq i \leq N} \{r_i, r_i'\} > 0$, $\min_{1 \leq i \leq N} \{r_i, r_i'\} \geq 0$. Since μ and μ' are distinct in $\overset{\circ}{\mathcal{D}}$,

there is some i_0 such that $(d_{i_0}, r_{i_0}) \neq (d'_{i_0}, r'_{i_0})$. Without loss of generality, we may assume $r_{i_0} \geq r'_{i_0}$. Let $g(\cdot) = \gamma(\langle d_{i_0}, \cdot \rangle) \in W_{or}$ and choose $f \in C_0(\mathbb{R}^n, \mathbb{R})$ satisfying $f(z_{i_0}) = 1$ and $f(z_i) = 0$ for all $i \geq 2$. Since the kernel defined by ρ is C_0-universal, there exists $\{f_n\} \subset W_{pos}$ such that $f_n \to f$ uniformly. As $f_n \otimes g \in W$, we have that

$$0 = (\mu - \mu' | f_n \otimes g) = \sum_{i=1}^{N} f_n(z_i)(r_i g(d_i) - r'_i g(d'_i))$$

Taking the limit $n \to +\infty$, this gives:

$$0 = f(z_{i_0})(r_{i_0} g(d_{i_0}) - r'_{i_0} g(d'_{i_0})) = \underbrace{r_{i_0} \gamma(1) - r'_{i_0} \gamma(\langle d_{i_0}, d'_{i_0} \rangle)}_{A}. \qquad (2.2)$$

Since $(d_{i_0}, r_{i_0}) \neq (d'_{i_0}, r'_{i_0})$, $r_{i_0} \geq r'_{i_0}$ and $r_{i_0} > 0$, we have either $d_{i_0} \neq d'_{i_0}$ and then $A \geq r_{i_0}(\gamma(1) - \gamma(\langle d_{i_0}, d'_{i_0} \rangle)) > 0$ or $d_{i_0} = d'_{i_0}$ and $r_{i_0} > r'_{i_0}$ in which case $A = (r_{i_0} - r'_{i_0})\gamma(1) > 0$. In either case the right hand side of (2.2) is positive which is a contradiction. $\qquad \square$

Note that the C_0-universality assumption still implies that the kernel defined by ρ is strictly positive definite. However, the assumptions on γ are typically less restrictive than in Proposition 2.1.

A last subclass of varifold metrics that shall be of interest in this paper is the case of orientation-invariant kernels which amounts in choosing an even function γ in the kernel definition. This, indeed, leads to a space W^* and metric $\| \cdot \|_{W^*}$ for which Diracs $\delta_{(x,d)}$ and $\delta_{(x,-d)}$ are equal in W^* for any $(x,d) \in \mathbb{R}^n \times \mathbb{S}^{n-1}$. In other words, elements of $\mathcal{D}$ can be equivalently viewed as *unoriented* varifolds, i.e distributions on the product of $\mathbb{R}^n$ and the projective space of $\mathbb{R}^n$, similarly to the framework of [7]. In that particular situation, one obtains an induced distance under the conditions stated in the following proposition which proof is a straightforward adaptation of the one of Proposition 2.2.

Proposition 2.3: *If the kernel defined by ρ is C_0-universal, γ is an even function with $\gamma(1) > 0$ and $\gamma(u) < \gamma(1)$ for all $u \in (-1,1)$, then $\| \cdot \|_{W^*}$ induces a metric on the space $\mathring{\mathcal{D}}$ modulo the orientation.*

In Section 5 below, we will discuss more thoroughly and illustrate the effects of those kernel properties on the solutions to registration problems for different cases of discrete distributions.

3. Optimal diffeomorphic mapping of varifolds

It is essential to point out that the notion of varifold presented above contains but is also more general than curves and surfaces as it allows to model more complex

geometric structures like objects carrying multiple orientation vectors at a given position. In contrast with most previous works on diffeomorphic registration that only involve varifolds as an intermediary representation to compute fidelity terms between shapes, the purpose of this paper to derive a deformation model and registration framework on the space $\mathcal{D}$ itself.

3.1. *Group action*

A first key element is to express the way that deformations 'act' on discrete varifolds. Considering a smooth diffeomorphism $\phi \in \mathrm{Diff}(\mathbb{R}^n)$, we first intend to express how ϕ should transport a Dirac $\delta_{(x,d)}$. There is however not a canonical way to define it as the nature of the underlying data affects the deformation model itself. An important distinction to be made is on the interpretation of direction vectors d, whether they correspond for instance to a unit tangent direction to a curve or a surface in which case d is transported by the Jacobian of ϕ as $D_x\phi(d)/|D_x\phi(d)|$ or rather to a normal direction which instead requires a transport model involving the inverse of the transposed Jacobian i.e $(D_x\phi)^{-T}(d)/|(D_x\phi)^{-T}(d)|$ (see [22] chap. 10 for more thorough discussion). To keep notations more compact, we will write $D\phi \cdot d$ for a given generic action of $D\phi$ on $\mathbb{R}^n$ on either tangent or normal vector and $\overline{D\phi \cdot d}$ for the corresponding normalized vector in $\mathbb{S}^{n-1}$. That being said, we will also consider two distinct models for the action:

- $\phi_* \delta_{(x,d)} \doteq \delta_{(\phi(x),\overline{D\phi \cdot d})}$ (*normalized action*): this corresponds to transporting the Dirac mass at the new position $\phi(x)$ and transforming the orientation vector as $\overline{D\phi \cdot d}$.
- $\phi_\# \delta_{(x,d)} \doteq |D\phi \cdot d|\delta_{(\phi(x),\overline{D\phi \cdot d})}$ (*pushforward action*): the position and orientation vector are transported as previously but with a reweighting factor equal to the norm of $D\phi \cdot d$.

It is then straightforward to extend both of these definitions by linearity to any discrete varifold in $\mathcal{D}$. In both cases, we obtain a group action of diffeomorphisms on the set of discrete varifolds. However, these actions are clearly not equivalent. The normalized action operates as a pure transport of mass and rotation of the direction vector whereas the pushforward model adds a weight change corresponding to the Jacobian of ϕ along the direction d. This is a necessary term in the situation where $\mu = \mu_S$ is representing a discrete oriented curve or surface. Indeed, one can check, up to discretization errors, that under the pushforward model, we have $\phi_\# \mu_S = \mu_{\phi(S)}$; in other words the action is compatible with the usual deformation of a shape. In the result section below, we will show examples of matching based on those different group action models.

3.2. *Orbits and isotropy groups*

Although we will be focusing on special subgroups of diffeomorphisms in the next section, it will be insightful to study a little more closely the orbits of discrete varifolds under the normalized and pushforward actions of the full group $\text{Diff}(\mathbb{R}^n)$ (or similarly the equivalence classes $\mathcal{D}/\text{Diff}(\mathbb{R}^n)$). Let $\mu \in \mathcal{D}$ which we can write as $\mu = \sum_{i=1}^{N} \sum_{j=1}^{n_i} r_{i,j} \delta_{(x_i, d_{i,j})}$ where the x_i are here assumed to be distinct positions and for each $i = 1, \ldots, N$, the $(d_{i,j})_{j=1,\ldots,n_i}$ are distinct in $\mathbb{S}^{n-1}$. While it is well-known that $\text{Diff}(\mathbb{R}^n)$ acts transitively on the set of point clouds of N points in $\mathbb{R}^n$ (as $n \geq 2$), this may no longer hold when one or several direction vectors are attached to each point position.

In the case of the normalized action, we have $\phi_* \mu = \sum_{i=1}^{N} \sum_{j=1}^{n_i} r_{i,j} \delta_{(\phi(x_i), \overline{D\phi \cdot d_{i,j}})}$. We see that the orbit of μ is then given by:

$$\text{Diff}_* \mu = \left\{ \nu \in \mathcal{D} \ \ s.t \ \ \exists (y_i) \in (\mathbb{R}^n)^N, \ y_i \neq y_j \text{ for } i \neq j, \ \exists A_1, \ldots, A_N \in \text{GL}(\mathbb{R}^n), \right.$$
$$\left. \nu = \sum_{i=1}^{N} \sum_{j=1}^{n_i} r_{i,j} \delta_{(y_i, u_{i,j})} \ \text{ with } u_{i,j} = \frac{A_i d_{i,j}}{|A_i d_{i,j}|} \right\}$$

This is essentially the set of all discrete varifolds with any set of N distinct positions and for each i, a set of n_i directions obtained by a linear transformation of the $\{d_{i,j}\}_{j=1,\ldots,n_i}$ with weights $r_{i,j}$ unchanged. In particular, this imposes some constraints on the set of 'attainable' direction vectors: clearly, if the number of direction vectors at a given position exceeds the dimension i.e $n_i \geq n$, this system of vectors cannot be mapped in general to any other system of n_i vectors on the sphere by a single linear map. If we assume that the system of vectors at each position x_i forms a frame, i.e that for all i, $n_i \leq n$ and the direction vectors $d_{i,j}$ for $j = 1, \ldots, n_i$ are independent, then we see that the orbit of μ is given by the set of all discrete varifolds of the form $\sum_{i=1}^{N} \sum_{j=1}^{n_i} r_{i,j} \delta_{(y_i, u_{i,j})}$ with distinct y_i's and $(u_{i,j})$ in $\mathbb{S}^{n-1}$ such that the $(u_{i,j})_{j=1,\ldots,n_i}$ are independent for all i. In the special case of $n_i = 1$ for all i, that is $\mu \in \mathring{\mathcal{D}}$, the orbits are then entirely determined by the set of weights r_i which gives the identification of $\mathring{\mathcal{D}}/\text{Diff}(\mathbb{R}^n)$ with ordered finite sets of positive numbers.

With pushforward action, we have $\phi_\# \mu = \sum_{i=1}^{N} \sum_{j=1}^{n_i} |D\phi \cdot d_{i,j}| r_{i,j} \delta_{(\phi(x_i), \overline{D\phi \cdot d_{i,j}})}$ and the orbit writes:

$$\text{Diff}_\# \mu = \left\{ \nu \in \mathcal{D} \ \ s.t \ \ \exists (y_i) \in (\mathbb{R}^n)^N, \ y_i \neq y_j \text{ for } i \neq j, \exists A_1, \ldots, A_N \in \text{GL}(\mathbb{R}^n), \right.$$
$$\left. \nu = \sum_{i=1}^{N} \sum_{j=1}^{n_i} |A_i d_{i,j}| r_{i,j} \delta_{(y_i, u_{i,j})} \ \text{ with } u_{i,j} = \frac{A_i d_{i,j}}{|A_i d_{i,j}|} \right\}$$

In the general situation, there is again no simple characterization of the orbit.

With the additional assumptions that $n_i \leq n$ and the $(d_{i,j})_{j=1,\ldots,n_i}$ are independent vectors for each i, the orbit of μ is the set of all discrete varifolds of the form $\sum_{i=1}^{N} \sum_{j=1}^{n_i} s_{i,j} \delta_{(y_i, u_{i,j})}$ with any choice of distinct points y_i, direction vectors $(d_{i,j})$ in $\mathbb{S}^{n-1}$ such that the $(d_{i,j})_j$ are independent and weights $s_{i,j} > 0$. In particular, the action of $\mathrm{Diff}(\mathbb{R}^n)$ in the pushforward model is transitive on all subsets of $\mathring{\mathcal{D}}$ with fixed N, which implies that the equivalence classes of $\mathring{\mathcal{D}}/\mathrm{Diff}(\mathbb{R}^n)$ in that case are only determined by the number of Diracs in the discrete varifold, as we would expect.

The previous discussion thus shows that for both models and unlike the more standard cases of landmarks or discrete vector fields, the action of diffeomorphisms on discrete varifolds is in general not transitive. It is therefore necessary to formulate the registration problems in their inexact form by introducing fidelity terms like the kernel metrics introduced in Section 2.

An alternative viewpoint is to identify the previous orbits as quotients of $\mathrm{Diff}(\mathbb{R}^n)$ by their isotropy subgroups, which allows to draw some parallel with the model of jets of [14]. To simplify the presentation, let's consider $\mu = \sum_{i=1}^{N} r_i \delta_{(x_i, d_i)}$ in $\mathring{\mathcal{D}}$ (thus $x_i \neq x_j$ for $i \neq j$). Writing G_μ^* and $G_\mu^\#$ the isotropy subgroup of μ for the normalized and pushforward actions, we have the identifications $\mathrm{Diff}(\mathbb{R}^n)/G_\mu^* \simeq \mathrm{Diff}_*\mu$ and $\mathrm{Diff}(\mathbb{R}^n)/G_\mu^\# \simeq \mathrm{Diff}_\#\mu$. Now, we can describe those isotropy subgroups as follows. Denoting $\mathcal{S}_N$ the permutation group of $\{1,\ldots,N\}$, we have

$$G_\mu^* = \left\{ \phi \in \mathrm{Diff}(\mathbb{R}^n) \mid \exists \pi \in \mathcal{S}_N, \ \forall i = 1,\ldots,N, \ \phi(x_i) = x_{\pi(i)} \text{ and } \frac{d_{x_i}\phi(d_i)}{|d_{x_i}\phi(d_i)|} = d_{\pi(i)} \right\}$$

and

$$G_\mu^\# = \left\{ \phi \in \mathrm{Diff}(\mathbb{R}^n) \mid \exists \pi \in \mathcal{S}_N, \ \forall i = 1,\ldots,N, \ \phi(x_i) = x_{\pi(i)} \text{ and } r_i d_{x_i}\phi(d_i) = r_{\pi(i)} d_{\pi(i)} \right\}.$$

Interestingly, the isotropy group for the space of first-order jets on the landmark set $z = \{x_i\}_{i=1,\ldots,N}$ as derived in [14] is given by

$$G_z^{(1)} = \{ \phi \in \mathrm{Diff}(\mathbb{R}^n) \mid \forall i = 1,\ldots,N, \ \phi(x_i) = x_i \text{ and } d_{x_i}\phi = \mathrm{Id} \}.$$

We can notice that $G_z^{(1)}$ is actually a subgroup of both G_μ^* and $G_\mu^\#$, which shows again that the deformation models on discrete varifolds considered here comes with additional invariances, in particular the invariance to permutation of the Dirac varifolds in μ, in contrast with the labelled particles of [20, 14].

3.3. *Optimal control problem*

With the definitions and notations of the previous sections, we can now introduce the mathematical formulation of the diffeomorphic registration of discrete

varifolds. As mentioned in the introduction, we will rely on the LDDMM model for generating diffeomorphisms although other transformation spaces and models could be taken as well. In short, we consider a space of time dependent velocity fields $v \in L^2([0,1], V)$ such that for all $t \in [0,1]$, v_t belongs to a certain RKHS V of vector fields on $\mathbb{R}^n$. We will write $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ the vector-valued reproducing kernel of V. From v, one obtains the flow mapping ϕ_t^v at each time t as the integral of the differential equation $\partial_t \phi_t^v = v_t \circ \phi_t^v$ with $\phi_0 = \mathrm{Id}$. The deformation group G_V is then defined as the set of all flow maps ϕ_1^v for all velocity fields $v \in L^2([0,1], V)$. With the adequate assumptions on the kernel of V, this is a subgroup of $\mathrm{Diff}(\mathbb{R}^n)$ and it is naturally equipped with a *right-invariant metric* given on V by $\int_0^1 \|v_t\|_V^2 dt$. This right-invariant metric can be then shown to descend to a metric on the orbit spaces of discrete varifolds for the action of G_V, which can be described in a similar way as in the previous section. We refer to [22] for a detailed exposition of the LDDMM framework.

Now, consider two discrete varifolds $\mu_0 = \sum_{i=1}^P r_{i,0} \delta_{(x_{i,0}, d_{i,0})}$ (template) and $\tilde{\mu} = \sum_{j=1}^Q \tilde{r}_j \delta_{(\tilde{x}_j, \tilde{d}_j)}$ (target). As μ_0 and $\tilde{\mu}$ may very well not belong to the same orbit, we introduce the registration problem in its inexact formulation which consists in the variational problem:

$$\mathrm{argmin}_{v \in L^2([0,1],V)} \left\{ E(v) = \int_0^1 \|v_t\|_V^2 dt + \lambda \|\mu(1) - \tilde{\mu}\|_{W^*}^2 \right\} \qquad (3.1)$$

subject to either $\mu(t) \doteq (\phi_t^v)_* \mu_0$ in the normalized action scenario or $\mu(t) \doteq (\phi_t^v)_\# \mu_0$ for the pushforward model, and λ being a weight parameter between the regularization and fidelity terms in the energy. This is easily interpreted as an optimal control problem in which the state variable is the transported varifold $\mu(t)$, the control is the velocity field v and the cost functional is the sum of the standard LDDMM metric on the deformation and a discrepancy term between $\mu(1)$ and the target given by a varifold kernel metric as in (2.1). Those optimal control problems are well-posed in the following sense:

Proposition 3.1: *If V is continuously embedded in the space $C_0^2(\mathbb{R}^n, \mathbb{R}^n)$, or equivalently if K is of class C^2 with all derivatives up to order 2 vanishing at infinity, then there exists a global minimum to the problem* (3.1).

Proof: The result follows from an argument similar to that of the existence of minimizers in usual LDDMM registration problems. If (v^n) is a minimizing sequence in $L^2([0,1], V)$ then thanks the first term of E, we may assume that (v^n) is bounded in $L^2([0,1], V)$ and therefore that, up to extracting a subsequence, $v^n \rightharpoonup v^*$ weakly in $L^2([0,1], V)$. It then follows from the results of [22] (Chapter 8.2) that the sequence of diffeomorphisms $(\phi_1^{v^n})$ and their first-order dif-

ferentials $(d\phi_1^{v^n})$ converge uniformly on every compact respectively to $\phi_1^{v^*}$ and $d\phi_1^{v^*}$. In particular, for all $i = 1, \ldots, P$, $\phi_1^{v^n}(x_i) \to \phi_1^{v^*}(x_i)$ and $d\phi_1^{v^n}(x_i) \to d\phi_1^{v^*}(x_i)$. Then, from the expressions of the group actions and the metric (2.1), we obtain that either $\|(\phi_1^{v^n})_*\mu_0 - \tilde{\mu}\|_{W^*}^2 \xrightarrow[n\to\infty]{} \|(\phi_1^{v^*})_*\mu_0 - \tilde{\mu}\|_{W^*}^2$ or $\|(\phi_1^{v^n})_\#\mu_0 - \tilde{\mu}\|_{W^*}^2 \xrightarrow[n\to\infty]{}$ $\|(\phi_1^{v^*})_\#\mu_0 - \tilde{\mu}\|_{W^*}^2$. Finally, using the weak lower semicontinuity of the norm in $L^2([0,1], V)$, it gives in both cases:

$$E(v^*) \le \liminf_{n\to\infty} E(v^n)$$

and consequently v^* is a global minimizer of E. $\square$

3.4. *Hamiltonian dynamics*

By fixing the final time condition $\mu(1)$ and minimizing $\int_0^1 \|v_t\|_V^2 dt$ with those boundary constraints, the resulting path $t \mapsto \mu(t)$ corresponds to a *geodesic* in $\mathcal{D}$ for the metric induced by the metric on the deformation group. We can further characterize those geodesics as solutions of a Hamiltonian system. For that purpose, we follow the general setting developed in [2] for similar optimal control problems.

In our situation, we can describe the state $\mu(t)$ as a set of P particles each given by the triplet $(x_i(t), d_i(t), r_i(t)) \in \mathbb{R}^n \times \mathbb{S}^{n-1} \times \mathbb{R}_+^*$ representing its position, orientation vector and weight. From 3.1, we have that $x_i(t) = \phi_t^v(x_{i,0})$, $d_i(t) = \overline{D\phi_t^v \cdot d_{i,0}}$ and $r_i(t) = r_{i,0}$ for the normalized action and $r_i(t) = |D\phi_t^v \cdot d_{i,0}|r_{i,0}$ in the pushforward case. Differentiating with respect to t, the state evolution may be alternatively described by the set of ODEs

$$\begin{cases} \dot{x}_i(t) = v_t(x_i(t)) \\ \dot{d}_i(t) = P_{d_i(t)^\perp}(Dv_t \cdot d_i(t)) \\ \dot{r}_i(t) = \begin{cases} 0 \ \text{(normalized)} \\ \langle d_i(t), Dv_t \cdot d_i(t)\rangle r_i(t) \ \text{(pushforward)} \end{cases} \end{cases}$$

where $P_{d_i(t)^\perp}$ denotes the orthogonal projection on the subspace orthogonal to $d_i(t)$, $Dv_t \cdot d_i(t)$ corresponds to the infinitesimal variation of the action of $D\phi$ on vectors of $\mathbb{R}^n$ introduced in 3.1: it is given specifically by $Dv_t \cdot d_i(t) = D_{x_i(t)}v_t(d_i(t))$ in the tangent case and $Dv_t \cdot d_i(t) = -(D_{x_i(t)}v_t)^T(d_i(t))$ in the normal case. Note that other choices of transformation of the weights could be treated quite similarly by modifying accordingly the last equation in the previous system. In what follows, we detail the derivations of the optimality equations in the case of tangent direction vectors for both normalized and pushforward group action models, the situation of normal vectors being easily tackled in similar fashion.

3.4.1. *Normalized action*

In the case of normalized action, $\{r_i(t)\}_{i=1}^{P}$ are time independent as previous discussed. So we can choose the state variable of the optimal control problem to be $q := \{(x_i, d_i), \ i = 1, \cdots, P\} \in \mathbb{R}^{2dP}$ with the infinitesimal action

$$\xi_q v = \left\{ \left(v(x_i), P_{d_i^{\perp}}(D_{x_i}v(d_i)) \right), \ i = 1, \cdots, P \right\}$$

and introduce the Hamiltonian

$$H(p,q,v) = (p|\xi_q v) - \frac{1}{2}\|v\|_V^2$$

$$= \sum_{i=1}^{N} \langle p_i^{(1)}, v(x_i) \rangle + \langle P_{d_i^{\perp}}(p_i^{(2)}), D_{x_i}v(d_i) \rangle - \frac{1}{2}\|v\|_V^2,$$

where

$$p = \left\{ \left(p_i^{(1)}, p_i^{(2)} \right), \ i = 1, \cdots, P \right\} \in \mathbb{R}^{2dP}$$

is the adjoint variable of state q. We call $p_i^{(1)}$ the *spatial momentum* and $p_i^{(2)}$ the *directional momentum*. From Pontryagin's maximum principle, the Hamiltonian dynamics is given by the forward system of equations

$$
\begin{cases}
\dot{x}_i(t) &= v_t(x_i(t)) \\
\dot{d}_i(t) &= P_{d_i(t)^{\perp}}(D_{x_i(t)}v_t(d_i(t))) \\
\dot{p}_i^{(1)}(t) &= -(D_{x_i(t)}v_t)^T p_i^{(1)}(t) \\
&\quad -(D_{x_i(t)}^{(2)}v_t(\cdot, d_i(t)))^T \left(P_{d_i(t)^{\perp}}(p_i^{(2)}(t)) \right) \\
\dot{p}_i^{(2)}(t) &= -(D_{x_i(t)}v_t)^T P_{d_i(t)^{\perp}}(p_i^{(2)}(t)) \\
&\quad + \langle d_i(t), p_i^{(2)}(t) \rangle D_{x_i(t)}v_t(d_i(t)) \\
&\quad + \langle d_i(t), D_{x_i(t)}v_t(d_i(t)) \rangle p_i^{(2)}(t)
\end{cases}
\tag{3.2}
$$

and optimal vector fields v satisfy

$$\langle v_t, h \rangle_V = \left(p(t), \xi_{q(t)}h \right)$$

$$= \sum_{i=1}^{P} \langle p_i^{(1)}, h(x_i) \rangle + \langle P_{d_i(t)^{\perp}}(p_i^{(2)}(t)), D_{x_i(t)}h(d_i(t)) \rangle,$$

for any $h \in V$ and $t \in [0, 1]$. The reproducing property and reproducing property for the derivatives in a vector RKHS give [20] that $\forall x \in \mathbb{R}^n, \ z \in \mathbb{R}^n, v \in V$ and multi-index α,

$$\langle K(x, \cdot)z, v \rangle_V = (z \otimes \delta_x | v)$$
$$\langle D_1^{\alpha}K(x, \cdot), v \rangle_V = z^T D^{\alpha}v(x).$$

H-W. Hsieh, N. Charon

With the above properties, we obtain the following expression of v

$$v_t(\cdot) = \sum_{k=1}^{P} K(x_k(t), \cdot) p_k^{(1)}(t)$$
$$+ D_1 K(x_k(t), \cdot) \left(d_k(t), P_{d_k(t)^\perp}(p_k^{(2)}(t)) \right). \tag{3.3}$$

where we use the shortcut notation $D_1 K(x, \cdot)(u_1, u_2)$ for the vector $D_1(K(x, \cdot)u_2)(u_1)$. In Figure 2, we show an example of geodesic and resulting deformation for a single Dirac varifold, which is obtained as the solution of (3.2) with the initial momenta shown in the figure. It illustrates the combined effects of the spatial momentum which displaces the position of the Dirac and of the directional momentum that generates a local rotation of the direction vector.
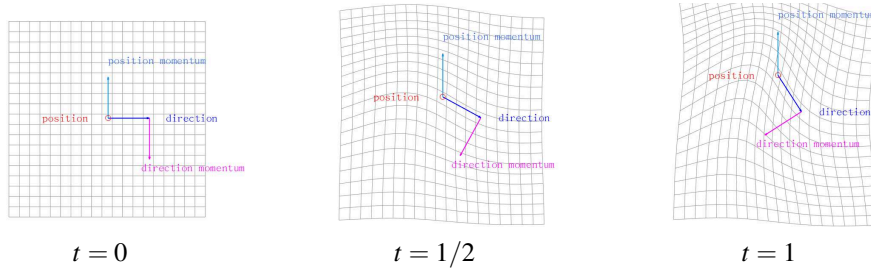

$t = 0$ $t = 1/2$ $t = 1$

Fig. 2.   Example of geodesic for a single Dirac in the normalized action case.

3.4.2. *Pushforward action*

As in the previous section, we set the state variable $q := \{(x_i, d_i, r_i), \ i = 1, \cdots, P\} \in (\mathbb{R}^n \times \mathbb{S}^{n-1} \times \mathbb{R}_+^*)^P$, the infinitesimal action

$$\xi_q v = \left\{ \left(v(x_i), P_{d_i^\perp}(d_{x_i} v(d_i)), r_i \langle d_i, d_{x_i} v(d_i) \rangle \right), \ i = 1, \cdots, P \right\}$$

and the Hamiltonian

$$H(p, q, v) = \sum_{i=1}^{P} \left\langle p_i^{(1)}, v(x_i) \right\rangle + \left\langle P_{d_i^\perp}(p_i^{(2)}), d_{x_i} v(d_i) \right\rangle + p_i^{(3)} r_i \langle d_i, d_{x_i} v(d_i) \rangle - \frac{1}{2} \|v\|_V^2,$$
$$\tag{3.4}$$

where $p = \left\{ \left(p_i^{(1)}, p_i^{(2)}, p_i^{(3)} \right), \ i = 1, \cdots, P \right\} \in \mathbb{R}^{(2d+1)P}$. Applying again Pontryagin's maximum principle, we obtain the forward system

$$
\begin{cases}
\dot{x}_i &= v_t(x_i) \\
\dot{d}_i &= P_{d_i^\perp}(d_{x_i} v(d_i)) \\
\dot{r}_i &= r_i \langle d_i, d_{x_i} v_t(d_i) \rangle \\
\dot{p}_i^{(1)} &= -(d_{x_i(t)} v_t)^T p_i^{(1)} - (d_{x_i(t)}^{(2)} v(\cdot, d_i))^T (P_{d_i^\perp}(p_i^{(2)})) - p_i^{(3)} r_i d_{x_i} v(\cdot, d_i)^T d_i \\
\dot{p}_i^{(2)} &= -d_{x_i(t)} v_t^T (p_i^{(2)}) + \left(\left\langle d_i, p_i^{(2)} \right\rangle - r_i p_i^{(3)} \right) \left[d_{x_i} v_t + d_{x_i} v_t^T \right] (d_i) \\
&\quad + \langle d_i, d_{x_i} v_t(d_i) \rangle p_i^{(2)} \\
\dot{p}_i^{(3)} &= -p_i^{(3)} \langle d_i, d_{x_i} v_t(d_i) \rangle
\end{cases}
\tag{3.5}
$$

with optimal vector field of the form

$$
v_t(x) = \sum_{k=1}^{P} K(x_k, x) p_k^{(1)} + D_1 K(x_k, x) \left(d_k, P_{d_k^\perp}(p_k^{(2)}) + p_k^{(3)} r_k d_k \right). \tag{3.6}
$$

From the forward equations (3.5), we see that $\frac{d}{dt} \langle d_i(t), d_i(t) \rangle = 0$ and $\frac{d}{dt} r_i(t) p_i^{(3)}(t) = 0$, hence $\|d_i(t)\|$ and $r_i(t) p_i^{(3)}(t)$ are constant along geodesic paths. Similarly to the normalized action case, we can use use those conservation properties to reduce the number of state and dual variables as follows.

Let the new state variable be $q = \{(x_i, u_i), \ i = 1, \cdots, P\}$ and the Hamiltonian

$$
H(p, q, v) = \sum_{i=1}^{P} \langle p_i^{(1)}, v(x_i) \rangle + \langle p_i^{(2)}, d_{x_i} v(u_i) \rangle - \frac{1}{2} \|v\|_V^2 \tag{3.7}
$$

The forward equations and optimal vector field v derived from this Hamiltonian are

$$
\begin{cases}
\dot{x}_i(t) = v_t(x_i(t)) \\
\dot{d}_i(t) = d_{x_i(t)} v_t(u_i(t)) \\
\dot{p}_i^{(1)}(t) = -(d_{x_i(t)} v_t)^T p_i^{(1)} - (d_{x_i(t)}^{(2)} v(\cdot, u_i))^T p_i^{(2)} \\
\dot{p}_i^{(2)}(t) = -(d_{x_i(t)} v_t)^T p_i^{(2)}(t)
\end{cases}
\tag{3.8}
$$

and

$$
v_t(x) = \sum_{k=1}^{P} K(x_k, x) p_k^{(1)}(t) + D_1 K(x_k, x)(u_k(t), p_k^{(2)}(t)). \tag{3.9}
$$

Then this new system is rigorously equivalent to the original one in the following sense:

Proposition 3.2: *Any solution of* (3.8) + (3.9) *is such that* $\left(x_i(t), \overline{u_i(t)}, |u_i(t)| \right)$ *is a solution of* (3.5) + (3.6). *Conversely, any solution* $(x_i(t), d_i(t), r_i(t))$ *of* (3.5)

+ (3.6) *with initial conditions satisfying* $\left\langle p_i^{(2)}(0), u_i(0) \right\rangle = r_i(0) p_i^{(3)}(0)$ *gives the solution* $(x_i(t), r_i(t) d_i(t))$ *to* (3.8) + (3.9).
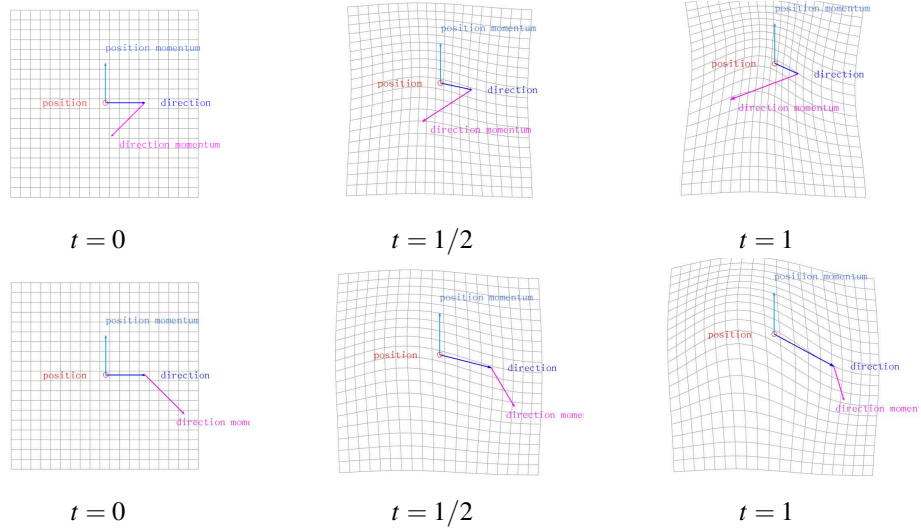

Fig. 3.   Examples of geodesics in the pushforward action case.

The proof is given in Appendix. Note that these equations can be also obtained in a more particular case as the geodesic equations on the tangent bundle of the space of landmarks, as derived for instance in [1] (Section 3.5). In what follows, we will thus replace the system (3.5) by (3.8).

Remark 3.3: We point out that there are other conserved quantities in the previous system. In particular, it's easy to see that $\left\langle p_i^{(2)}, d_i \right\rangle$ is constant along geodesics since

$$\frac{d}{dt} \left\langle p_i^{(2)}, d_i \right\rangle = - \left\langle (d_{x_i} v_t)^T p_i^{(2)}, d_i \right\rangle + \left\langle p_i^{(i)}, d_{x_i} v_t(d_i) \right\rangle = 0.$$

Figure 3 shows two geodesic trajectories of a single Dirac varifold for different initial momenta. In particular, we can again observe the effect of the directional momentum $p^{(2)}$ on the dynamics and resulting deformations. In addition to similar rotation effects as in the normalized action case, local contraction or expansion can be generated as well, depending precisely on the angle $\langle p_i^{(2)}, d_i \rangle$.

4. Registration algorithm and implementation

We now turn to the issue of numerically solving the optimization problem (3.1). We will follow the commonly used method for such problems called *geodesic shooting* (cf [21]). Indeed, from the developments of Section 3.4, we see that optimizing (3.1) with respect to vector fields v can be done equivalently by restricting to geodesics and thus by optimizing over the initial momenta variables $p_0^{(1)}$ and $p_0^{(2)}$ that completely parametrize those geodesics through the Hamiltonian equations.

4.1. *Computation of E*

Let a template and target discrete varifold be given as in Section 3.3. As mentioned above, we can rewrite the energy E as a function of the initial momenta that we will denote $p_0^{(1)} = (p_i^{(1)}(0))$ and $p_0^{(2)} = (p_i^{(2)}(0))$:

$$E(p_0^{(1)}, p_0^{(2)}) = H_r(p_0, q_0) + \lambda \underbrace{\|\mu(1) - \tilde{\mu}\|_{W^*}^2}_{:=g(q(1))} \tag{4.1}$$

where q_0 is the initial state, $\mu(1)$ is the varifold corresponding to the final time state $q(1)$ with $g(q(1))$ the resulting fidelity term between $\mu(1)$ and the target varifold, and H_r is the *reduced Hamiltonian* $H_r(p,q) = H(p,q,v)$ for the optimal v given by (3.3) or (3.9) (note that $H_r(p(t), q(t))$ is conserved along solutions of the Hamiltonian systems thus giving the above expression of the energy).

The expression of H_r as well as the resulting reduced Hamiltonian equations can be obtained in all generality by plugging the expression of v in the equations of Section 3.3. In our implementation, we actually restrict to the more particular case of radial scalar kernels for the vector fields in V, i.e we assume that $K(x,y) = h(|x-y|^2)I_n$. Then the reduced Hamiltonian for the normalized case becomes:

$$H_r(p,q) = \frac{1}{2} \langle \overline{K}_q p, p \rangle, \tag{4.2}$$

where $\overline{K}_q$ is a symmetric positive definite matrix which is defined as follows. Let

$$H = (H)_{ik} = (h_{ki})$$
$$\overline{A} = (\overline{A})_{ik} = 2\dot{h}_{ki}\langle x_k - x_i, d_k \rangle$$
$$\overline{B} = (\overline{B})_{ik} = -\left[4\ddot{h}_{ik}\langle x_k - x_i, d_i \rangle\langle x_k - x_i, d_k \rangle + 2\dot{h}_{ki}\langle d_i, d_k \rangle \right]$$

with h_{ki} being a shortcut for $h(|x_i - x_j|^2)$ and

$$P_{d^\perp}(\cdot) = \begin{pmatrix} I_d - d_1 \cdot d_1^T & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & I_d - d_N \cdot d_N^T \end{pmatrix}$$

Then we define

$$\overline{K}_q := \begin{pmatrix} I_{Pd} & 0 \\ 0 & P_{d^\perp} \end{pmatrix}^T \begin{pmatrix} H \otimes I_d & \overline{A} \otimes I_d \\ (\overline{A} \otimes I_d)^T & \overline{B} \otimes I_d \end{pmatrix} \begin{pmatrix} I_{Pd} & 0 \\ 0 & P_{d^\perp} \end{pmatrix}$$

where $\otimes$ denotes the Kronecker product. For the pushforward action case, we define H, A and B as in normalized action case with d_i and d_k replaced by u_i and u_k, then

$$H_r(p,q) = \frac{1}{2} \langle K_q p, p \rangle, \tag{4.3}$$

where

$$K_q := \begin{pmatrix} H \otimes I_d & A \otimes I_d \\ (A \otimes I_d)^T & B \otimes I_d \end{pmatrix}.$$

This gives us explicitly the first term of the energy in (4.1).

Now, the time evolution of q and p can be also rewritten equivalently in reduced Hamiltonian form, which expressions are given in full for radial scalar kernels in the Appendix. We numerically integrate those differential systems using an RK4 scheme, which we experienced to be better-adapted to these systems than the simpler Euler midpoint integrator used in [16]. Then, given initial momenta $p_0^{(1)}$ and $p_0^{(2)}$, integrating those equations forward in time produces the final state $q(1)$ and its corresponding varifold $\mu(1)$. It is then straightforward to evaluate the second term in (4.1) through the expression of the varifold norm (2.1); in the pushforward case one only needs to apply the additional intermediate operation of converting state $q(1) = (x_i(1), u_i(1))$ into $(x_i(1), \overline{u_i}(1), |u_i|(1))$. We will discuss different choices of kernels for the varifold metric in the result section.

4.2. *Computation of the gradient of* E

The second element we need is the gradient of the energy with respect to the momenta. The first term being directly a function of p_0, it can be differentiated easily and gives the following gradient:

$$\nabla_{p_0} H_r(p_0, q_0) = \overline{K}_q p_0 \quad \text{(normalized)}$$

$$\nabla_{p_0} H_r(p_0, q_0) = K_q p_0 \quad \text{(pushforward)}$$

The fidelity term $g(q(1))$ in (4.1), however, is a function of the final state $q(1)$ which is in turn a function of the momenta through the Hamiltonian system of equations. The computation of the gradient is therefore more involved due to the complicated dependency of $q(1)$ in p_0. The standard approach for optimal control problems of this form (cf for example [21] or [2]) is to introduce the *adjoint Hamiltonian system*:

$$\dot{Z}(t) = d(\partial_p H_r, \partial_q H_r)^T Z(t)$$

with $Z = (\tilde{q}, \tilde{p})^T$ the vector of the adjoint variables. Then, as detailed in [2], the gradient of $g(q(1))$ with respect to p_0 is given by $\tilde{p}(0)$ where $(\tilde{q}(t), \tilde{p}(t))^T$ is the solution of the adjoint system integrated backward in time with $\tilde{q}(1) = \nabla g(q(1))$ and $\tilde{p}(1) = 0$.

For the particular Hamiltonian equations considered here, the adjoint system is tedious to derive and to implement. We simply avoid that by approximating the differentials appearing in the adjoint system by finite difference of the forward Hamiltonian equations, following the suggestion of [2] (Section 4.1) which we refer to for details. Note that another possibility would be to take advantage of automatic differentiation methods, as used recently for some LDDMM registration problems by the authors of [17].

Lastly, the end time condition $\nabla g(q(1))$ in the previous adjoint system is computed by direct differentiation of the varifold norm (2.1) with respect to the final state variables. This is actually more direct than in previous works like [7, 16] where the gradients are computed with respect to the positions of vertices of the underlying mesh. Here, we have specifically, for the normalized model:

$$\partial_{x_i} g(q(1)) = 2 \sum_{j=1}^{P} 2 r_i r_j \rho'(|x_i(1) - x_j(1)|^2) \gamma(\langle d_i(1), d_j(1)\rangle).(x_i(1) - x_j(1)) \ - 2\dots$$

$$\partial_{d_i} g(q(1)) = 2 \sum_{j=1}^{P} r_i r_j \rho(|x_i(1) - x_j(1)|^2) \gamma'(\langle d_i(1), d_j(1)\rangle).d_j(1) \ - 2\dots$$

where the $\dots$ denote a similar term for the differential of the cross inner product $\langle \mu(1), \tilde{\mu}\rangle_{W^*}$. In the pushforward case with state variables $(x_i(1), u_i(1))$, we first compute $d_i(1) = u_i(1)/|u_i(1)|$ and $r_i(1) = |u_i(1)|$ and obtain $\partial_{x_i} g(q(1))$ with the same expression as above while $\partial_{u_i} g(q(1))$ is given by a simple chain rule.

Finally, with the above notations, the gradient of E writes:

$$\nabla_{p_0} E = \overline{K}_q p_0 + \lambda \tilde{p}(0) \tag{4.4}$$

respectively $\nabla_{p_0} E = K_q p_0 + \lambda \tilde{p}(0)$ in the pushforward case.

4.3. *Gradient descent algorithm*

The solution to the minimization of (4.1) is then computed by gradient descent on $p_0 = \left(p_{0,i}^{(1)}, p_{0,i}^{(2)} \right)_{i=1,...,P}$. Note that this is a non-convex optimization problem. Until convergence, each iteration consists of the following steps:

(1) Given the current estimate of p_0, integrate the Hamiltonian equations forward in time to obtain $q(1)$.

(2) Compute the gradient $\nabla g(q(1))$.

(3) Integrate the adjoint Hamiltonian system backward in time to obtain $\nabla_{p_0} E$.

(4) Update p_0: we use two separate update steps for the spatial and directional momentum which are selected, at each iteration, using a rough space search approach leading to the lowest value of E.

5. Results

We now present a few results of registration using the previous algorithm on simple and synthetic examples. Our implementation equally supports objects in 2D or 3D, we will however focus on examples in $\mathbb{R}^2$ here simply to allow for an easier visualization and interpretation of the results.

5.1. *Curve registration*

We begin with a toy example of standard curve matching to compare the result and performance of our discrete varifold LDDMM registration algorithm with the state-of-the-art LDDMM approach for curves such as the implementations of [10, 16]. The former methods share a very similar formulation to (3.1) and also make use of varifold metrics as fidelity terms, the essential difference being that the state of the optimal control problem is there the set of vertices of the deformed template curve which is only converted to a varifold for the evaluation of the fidelity term at each iteration. But the dynamics of geodesics still correspond to usual point set deformation under the LDDMM model.

We consider here the pushforward model for the action of diffeomorphisms on discrete varifolds that we have seen is compatible with the action of diffeomorphisms on curves. In this case, the two formulation and optimization problems for curve registration are theoretically equivalent up to discretization precision. We verify it with the example of Figure 4 for which both algorithms are applied with the same deformation kernel, varifold metric and optimization scheme. Note that in our approach, template and target curves are first (and only once at the begin-
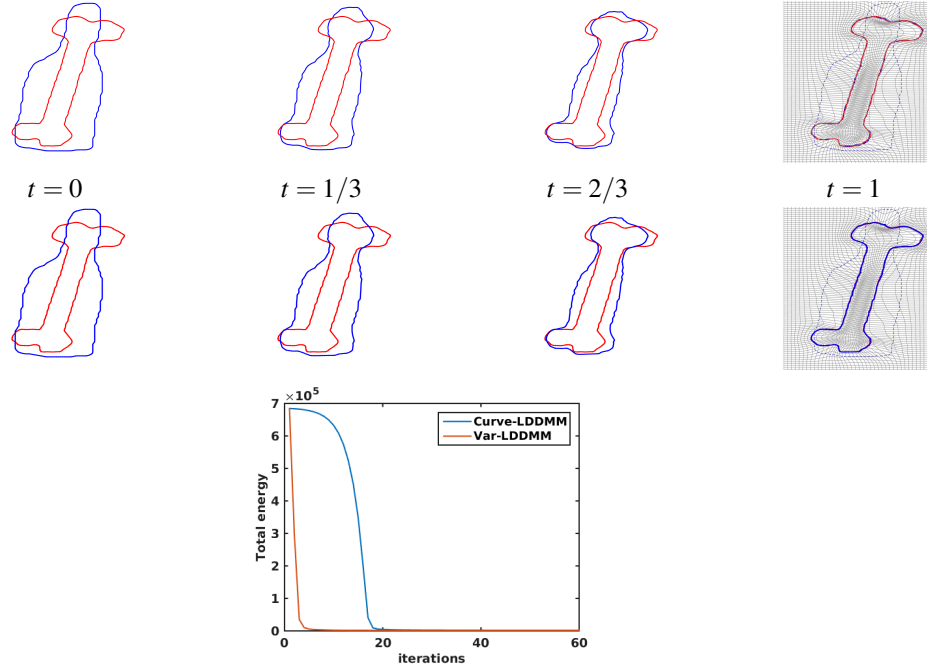
Fig. 4.   Curve registration using point-mesh LDDMM (1st row) and our proposed discrete varifold LDDMM (2nd row). On the last row is shown the evolution of the total energy across the iterations for both algorithms.

ning) converted to their discrete varifold representations as explained in Section 2.

As we can see, the resulting geodesics and deformations are consistent between the two methods. This is also corroborated by the very similar values of the energy at convergence. Interestingly however, although each iteration in our model is arguably more expansive numerically compared to standard curve-LDDMM due to the increased complexity of the Hamiltonian equations, the algorithm converges in a significantly lesser number of iterations. Whether this observation generalizes to other examples or other optimization methods will obviously require more careful examination in future work.

5.2. *Registration of directional sets*

We now turn to examples that are more specific to the framework of discrete varifolds.

Choice of the varifold metric

First, we examine more closely the effect of the metric $\|\cdot\|_{W^*}$ on the registration of discrete varifolds. The framework we propose can indeed support many choices for the kernel functions ρ and γ that define fidelity metrics $\|\cdot\|_{W^*}$ with possibly very different properties. This has been already analyzed quite extensively in [16] but only in the situation where varifolds associated to a curve or a surface. We consider here the same examples of kernels and briefly discuss what are the specific effects to expect when matching more general varifolds in $\mathcal{D}$ which may involve several orientation vectors at a given position.
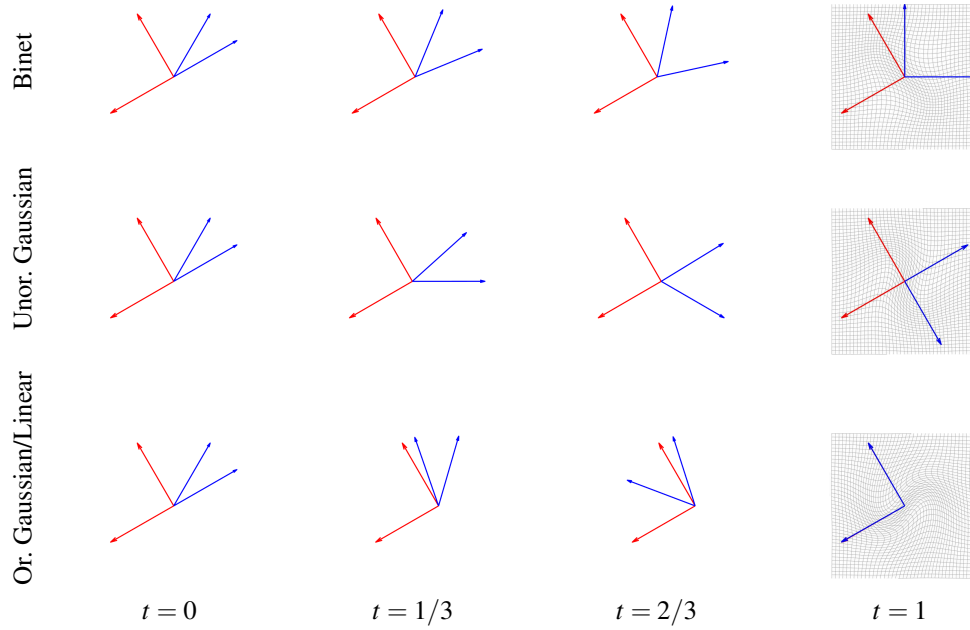

Fig. 5.   Matching of pairs of Dirac varifolds (template is in blue and target in red) under the normalized action with different choices of kernels: Binet on the first row, unoriented Gaussian ($\sigma_s = 1$) on the second and oriented Gaussian ($\sigma_s = 2$) on the last one. The linear kernel leads to the same result as the former in that particular case.

The results of Propositions 2.2 and 2.3 hold under the assumption that the kernel defined by ρ is a C_0-universal kernel on $\mathbb{R}^n$, which restricts the possible choices to a few known classes (cf [6] for a thorough analysis). Here, we will focus on the class of Gaussian kernels given by $\rho(|x-x'|^2) = e^{-\frac{|x-x'|^2}{\sigma^2}}$ with a width parameter $\sigma > 0$ that essentially provides a notion of spatial scale sensitivity to the

metric, and which must be adapted to the intrinsic sizes of shapes in each example.

In combination with ρ, as in [16], we introduce the following four kernels on $\mathbb{S}^{n-1}$:

- $\gamma(\langle d,d'\rangle) = \langle d,d'\rangle$ (linear kernel): this choice is related to the particular subclass of currents [10]. In that case, the resulting $\|\cdot\|_{W^*}$ is clearly only a pseudo-metric on $\mathcal{D}$ since the linearity implies that in W^*: $\delta_{(x,-d)} = -\delta_{(x,d)}$ and for any $d_1 \neq -d_2$, $\delta_{(x,d_1)} + \delta_{(x,d_2)} = |d_1 + d_2|\delta_{\left(x, \frac{d_1+d_2}{|d_1+d_2|}\right)}$. However, we still obtain a metric on the subspace $\mathring{\mathcal{D}}$ thanks to Proposition 2.2.

- $\gamma(\langle d,d'\rangle) = \langle d,d'\rangle^2$ (Binet kernel): γ being an even function, as discussed in Section 2, the resulting metric on W^* is invariant to the orientation of direction vectors. According to Proposition 2.3, we then have a distance on $\mathring{\mathcal{D}}$ modulo the orientation. Note however that with this particular choice, one does not obtain a metric (but only a pseudo-metric) on $\mathcal{D}$ modulo the orientation, as we will illustrate in the examples below.

- $\gamma(\langle d,d'\rangle) = e^{-\frac{2}{\sigma_s^2}(1-\langle d,d'\rangle^2)}$ (unoriented Gaussian kernel): this is another example of orientation-invariant kernel considered in [7] corresponding to a particular construction of Gaussian kernels on the projective space. In contrast with Binet kernel, it does induce a metric on $\mathcal{D}$ modulo orientation.

- $\gamma(\langle d,d'\rangle) = e^{-\frac{2}{\sigma_s^2}(1-\langle d,d'\rangle)}$ (oriented Gaussian kernel): this kernel is the restriction of the standard Gaussian kernel on $\mathbb{R}^n$ to the sphere $\mathbb{S}^{n-1}$. As such, it can be shown to be C_0-universal on $\mathbb{S}^{n-1}$ and thus, from Proposition 2.1, lead to a metric on the entire space $\mathcal{D}$.

We illustrate the aforementioned properties on a very simple registration example between pairs of Dirac varifolds located at the same position x i.e $\delta_{(x,d_1)} + \delta_{(x,d_2)}$ and $\delta_{(x,d_1')} + \delta_{(x,d_2')}$. In Figure 5, the template and target pairs of Diracs are matched based on the normalized action model. The estimated matching and deformations clearly differ with the choice of kernel but each of these result is in fact perfectly consistent with the different invariances of those kernels. Indeed the two Diracs are exactly matched to the target using the oriented Gaussian kernel since $\|\cdot\|_{W^*}$ is in that case a metric on the entire space $\mathcal{D}$. They are however matched to the opposite vectors with the unoriented Gaussian kernel which is indeed insensitive to orientation. In the case of Binet kernel, in addition to orientation-invariance, there exists other pairs of Diracs which are distinct in $\mathcal{D}$ but coincide in W^*. For example, it can be easily verified that all discrete varifolds of the form $\delta_{(x,d_1)} + \delta_{(x,d_2)}$ with orthogonal vectors d_1 and d_2 are equal in W^*,
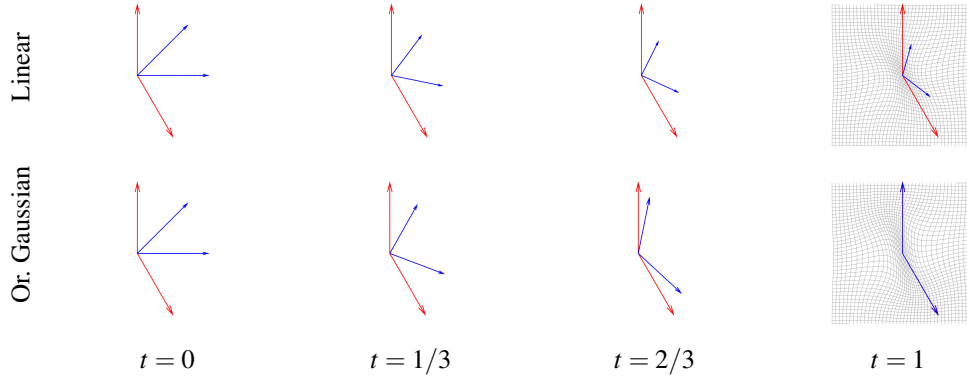
Fig. 6.   Registration of pairs of Dirac varifolds with the pushforward model for both the linear and oriented Gaussian kernel.

which is reflected by the result in Figure 5.

We emphasize the difference of behavior between linear and oriented Gaussian kernels with the example of Figure 6 associated this time to the pushforward action model. The result shown in the first row is a consequence of the fact that fidelity terms derived from the linear kernel only constrains the sums $d_1 + d_2$ and $d'_1 + d'_2$ to match.

Multi-directional varifold matching

Finally, Figure 7 shows an example of matching on more general discrete varifolds that involve varying number of directions at different spatial locations. This is computed with the normalized action using an oriented Gaussian kernel for the fidelity term. Although purely synthetic, it illustrates the potentialities of the proposed approach to register data with complex directional patterns.
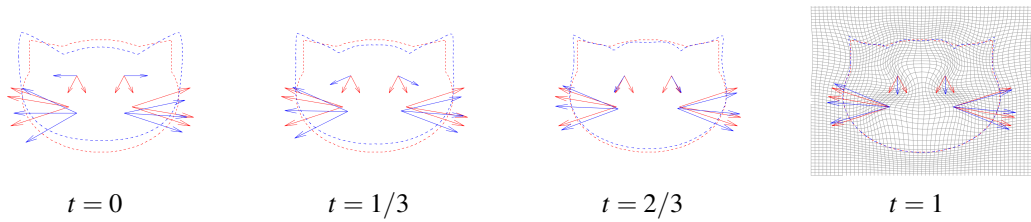

Fig. 7.   Registration of multi-directional sets. The lengths of vectors correspond to the weights of the Dirac varifolds.

5.3. *Contrast-invariant image registration*

A last possible application worth mentioning is the registration of images with varying contrast. Indeed, an image I modulo all contrast changes is equivalently represented by its unit gradient vector field $\frac{\nabla I}{|\nabla I|}$. Note that this may in fact be only defined at isolated pixels in the image, specifically the ones where the gradient is non vanishing. Within the setting of this work, it is thus natural to associate to I the discrete varifold

$$\mu_I = \sum_{\nabla I(x_i) \neq 0} \delta_{\left(x_i, \frac{\nabla I}{|\nabla I|}(x_i)\right)} \in \mathscr{D}$$

It is straightforward that μ_I is invariant to increasing contrast changes. It also becomes invariant to decreasing ones by quotienting out the orientation of the unit gradient vectors, which in our framework is simply done by selecting an orientation-invariant kernel $\gamma(\langle d, d' \rangle)$ to define $\| \cdot \|_{W^*}$. In Figure 8, this approach is used to map two oppositely contrasted synthetic phantom brain images. We show both the alignment of the discrete varifolds as well as the full deformation applied to the image itself. Note that these images have no noise and a simple structure with relatively low number of non-vanishing gradients. There will be clearly the need for more validation to be done in the future in order to evaluate the practicality and robustness of this method for real multi-modal medical images.

6. Conclusion and future work

We have proposed, in this paper, a framework for large deformation inexact registration between discrete varifolds. It relies on the LDDMM setting for diffeomorphisms and include different models of group action on the space of varifolds. In each case, we derived the corresponding optimal control problems and the associated geodesic equations in Hamiltonian form. By combining those with the use of kernel-based fidelity metrics on varifolds, we proposed a geodesic shooting algorithm to numerically tackle the optimization problems. We finally illustrated the versatility and properties of this approach through examples of various natures which go beyond the classical cases of curves or surfaces.

Several improvements or extensions of this work could be considered for future work. From a theoretical standpoint, it would be for instance important to derive a more general 'continuous' varifold matching model i.e with more general distributions than Dirac sums. Besides, higher dimensional varifolds could be possibly introduced within our model, although this would involve dealing with direction elements in Grassmann manifolds as in [7] instead of the simpler $\mathbb{S}^{n-1}$. Lastly, future work will also include adapting the existing fast GPU implementa-
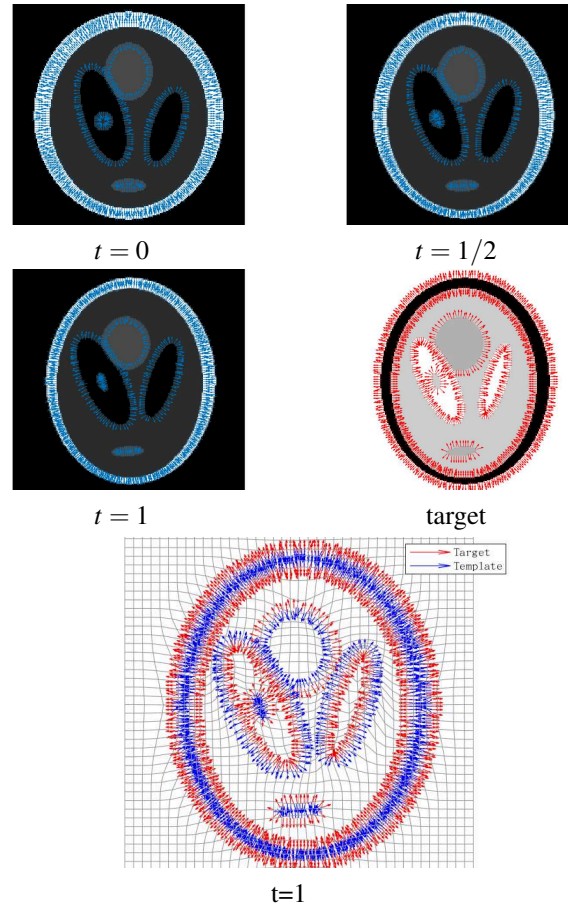
Fig. 8.   Registration of images modulo contrast changes. The matching is computed between the discrete varifolds associated to both images with the normalized action model and unoriented Gaussian fidelity term. The estimated deformation can be then applied to the template image.

tions for LDDMM to the new dynamical systems appearing here, with the objective of making the whole approach more scalable to real data applications.

Acknowledgements

Diffeomorphic registration of discrete geometric distributions   27

References

1. S. Arguillere, *The general setting of Shape Analysis*, preprint (2015).

2. S. Arguillere, E. Trélat, A. Trouvé, and L. Younes, *Shape deformation analysis from the optimal control viewpoint*, Journal de Mathmatiques Pures et Appliques **104** (2015), no. 1, 139–178.

3. B. Avants, C. Epstein, M. Grossman, and J. Gee, *Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain*, Medical Image Analysis **12** (2008), no. 1, 26 – 41.

4. M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, International journal of computer vision **61** (2005), no. 139-157.

5. Y. Cao, M. Miller, R. Winslow, and L. Younes, *Large deformation diffeomorphic metric mapping of vector fields*, IEEE Transactions on Medical Imaging **24** (2005), no. 9, 1216–1230.

6. C. Carmeli, E. De Vito, A. Toigo, and V. Umanita, *Vector valued reproducing kernel Hilbert spaces and universality*, Analysis and Applications **8** (2010), no. 01, 19–61.

7. N. Charon and A. Trouvé, *The varifold representation of non-oriented shapes for diffeomorphic registration*, SIAM journal of Imaging Sciences **6** (2013), no. 4, 2547–2580.

8. J. Du, A. Goh, and A. Qiu, *Diffeomorphic Metric Mapping of High Angular Resolution Diffusion Imaging Based on Riemannian Structure of Orientation Distribution Functions*, IEEE Transactions on Medical Imaging **31** (2012), no. 5, 1021–1033.

9. S. Durrleman, P. Fillard, X. Pennec, Alain Trouvé, and Nicholas Ayache, *Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents*, NeuroImage **55** (2010), no. 3, 1073–1090.

10. J. Glaunès, A. Qiu, M.I. Miller, and L. Younes, *Large deformation diffeomorphic metric curve mapping*, Int J Comput Vis **80** (2008), no. 3, 317336.

11. J. Glaunès, A. Trouvé, and L. Younes, *Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching*, CVPR **2** (2004), 712–718.

12. J. Glaunès and M. Vaillant, *Surface matching via currents*, Proceedings of Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science **3565** (2006), no. 381-392.

13. U. Grenander, *General pattern theory: A mathematical study of regular structures*, Clarendon Press Oxford, 1993.

14. H. Jacobs and S. Sommer, *Higher-order spatial accuracy in diffeomorphic image registration*, Geometry, Imaging and Computing **1** (2014), no. 4, 447–484.

15. S. Joshi and M. Miller, *Landmark matching via large deformation diffeomorphisms*, IEEE Transactions on Image Processing **9** (2000), no. 8, 1357–1370.

16. I. Kaltenmark, B. Charlier, and N. Charon, *A general framework for curve and surface comparison and registration with oriented varifolds*, Computer Vision and Pattern Recognition (CVPR) (2017).

17. L. Kühnel and S. Sommer, *Computational Anatomy in Theano*, Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics: proceedings of the MFCA Workshop (2017), 164–176.

28   *H-W. Hsieh, N. Charon*

18. C. Ragni, N. Diguet, J-F. Le-Garrec, et al., *Amotl1 mediates sequestration of the Hippo effector Yap1 downstream of Fat4 to restrict heart growth*, Nature communications **8** (2017).

19. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, *Nonrigid registration using free-form deformations: application to breast MR images*, IEEE Transactions on Medical Imaging **18** (1999), no. 8, 712–721.

20. S. Sommer, M. Nielsen, S. Darkner, and X. Pennec, *Higher-Order Momentum Distributions and Locally Affine LDDMM Registration*, SIAM Journal on Imaging Sciences **6** (2013), no. 1, 341–367.

21. F-X. Vialard, L. Risser, D. Rueckert, and C.J. Cotter, *Diffeomorphic 3D Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation*, International Journal of Computer Vision **97** (2012), no. 2, 229–241.

22. L. Younes, *Shapes and diffeomorphisms*, Springer, 2010.

APPENDIX

Proof of Proposition 3.2

Let $\left(X_i(t), u_i(t), P_i^{(1)}(t), P_i^{(2)}(t) \right)$ and $V_t(\cdot)$ satisfy equations (3.8) and (3.9), then it's straightforward to verify that

$$\left(x_i(t), d_i(t), r_i(t), p_i^{(1)}(t), p_i^{(2)}(t), p_i^{(3)}(t) \right)$$
$$:= \left(X_i(t), \overline{u_i}(t), |u_i(t)|, P_i^{(1)}(t), |u_i(t)| P_i^{(2)}(t), \left\langle P_i^{(2)}(t), \overline{u_i}(t) \right\rangle \right)$$

is a solution of equation (3.5) for V with the initial conditions

$$(X_i(0), \overline{u_i}(0), |u_i(0)|, P_i^{(1)}(0), |u_i(0)| P_i^{(2)}(0), \langle P_i^{(2)}(0), \overline{u_i}(0) \rangle).$$

Moreover, we see that

$$\left\langle p_i^{(2)}(t), d_i(t) \right\rangle = \left\langle u_i(t), P_i^{(2)}(t) \right\rangle = r_i(t) p_i^{(3)}(t), \ \forall t \tag{6.1}$$

which leads to V_t being equal to the vector field v_t defined in (3.6) and therefore to a solution for the system (3.5).

    Conversely, let $\left(x_i(t), d_i(t), r_i(t), p_i^{(1)}(t), p_i^{(2)}(t), p_i^{(3)}(t) \right))$ and $v_t(\cdot)$ satisfying (3.5) and (3.6) with initial conditions such that

$$\left\langle p_i^{(2)}(0), d_i(0) \right\rangle = r_i(0) p_i^{(3)}(0). \tag{6.2}$$

Now let $\left(X_i(t), u_i(t), P_i^{(1)}(t), P_i^{(2)}(t) \right)$ be the solution of (3.8) with the initial condition

$$\left(x_i(0), r_i(0) d_i(0), p_i^{(1)}(0), p_i^{(2)}(0) \right)$$

and vector field $v_t(\cdot)$. We define $V_t(\cdot)$ as in (3.9), then as in previous discussion, we see that

$$\left(X_i(t), \overline{u_i}(t), |u_i(t)|, P_i^{(1)}(t), |u_i(t)|P_i^{(2)}(t), \left\langle P_i^{(2)}(t), \overline{u_i}(t) \right\rangle \right)$$

is the solution for (3.5) with initial value

$$\left(X_i(0), \overline{u_i}(0), |u_i(0)|, P_i^{(1)}(0), |u_i(0)|P_i^{(2)}(0), \left\langle P_i^{(2)}(0), \overline{u_i}(0) \right\rangle \right)$$
$$= \left(x_i(0), d_i(0), r_i(0), p_i^{(1)}(0), p_i^{(2)}(0), p_i^{(3)}(0) \right).$$

Since $\left(x_i(t), d_i(t), r_i(t), p_i^{(1)}(t), p_i^{(2)}(t), p_i^{(3)}(t) \right))$ is a solution for the same initial value problem, by uniqueness of ODE, we obtain

$$\left(x_i(t), d_i(t), r_i(t), p_i^{(1)}(t), p_i^{(2)}(t), p_i^{(3)}(t) \right)$$
$$= \left(X_i(t), \overline{u_i}(t), |u_i(t)|, P_i^{(1)}(t), |u_i(t)|P_i^{(2)}(t), \left\langle P_i^{(2)}(t), \overline{u_i}(t) \right\rangle \right), \ \forall t \in [0,1].$$

Also, we have equation (6.1), and from this equation we have

$$P_{d_k^\perp}(p_k^{(2)}) + p_k^{(3)} r_k d_k = p_k^{(2)} + \left(\left\langle p_k^{(2)}, d_k \right\rangle - p_k^{(3)} r_k \right) d_k = p_k^{(2)}$$

and hence

$$v_t(\cdot) = \sum_{k=1}^{P} K(x_k, \cdot) p_k^{(1)} + D_1 K(x_k, \cdot)(d_k, p_k^{(2)})$$
$$= \sum_{k=1}^{P} K(X_k, \cdot) P_k^{(1)} + D_1 K(X_k, \cdot)(u_k, P_k^{(2)}) = V_t(\cdot).$$

Reduced Hamiltonian equations

For convenience, let us denote $f(|x_k - x_i|^2)$ by f_{ki} for any function f. Then the reduced Hamiltonian equations for the normalized action can be shown to be

$$
\begin{cases}
\dot{x}_i &= \sum_{k=1}^{P} h_{ki} p_k^{(1)} + 2\dot{h}_{ki} \langle x_k - x_i, d_k \rangle P_{d_k^\perp}(p_k^{(2)}) \\
\dot{d}_i &= \sum_{k=1}^{P} -2\dot{h}_{ki} \langle x_k - x_i, d_i \rangle P_{d_k^\perp}\left(p_k^{(1)}\right) \\
&\quad - \left[4\ddot{h}_{ki} \langle x_k - x_i, d_k \rangle \langle x_k - x_i, d_i \rangle + 2\dot{h}_{ki} \langle d_k, d_i \rangle \right] P_{d_k^\perp}(p_k^{(2)}) \\
\dot{p}_i^{(1)} &= \sum_{k=1}^{P} \Bigg\{ \left[2\dot{h}_{ki} \langle p_k^{(1)}, p_i^{(1)} \rangle + 4\ddot{h}_{ki} \langle x_k - x_i, d_k \rangle \langle P_{d_k^\perp}(p_k^{(2)}), p_i^{(1)} \rangle \right] \\
&\quad - \Bigg[4\ddot{h}_{ki} \langle x_k - x_i, d_i \rangle \langle p_k^{(1)}, P_{d_i^\perp}(p_i^{(2)}) \rangle \\
&\quad + \left(8h_{ki}^{(3)} \langle x_k - x_i, d_k \rangle \langle x_k - x_i, d_i \rangle + 4\ddot{h}_{ki} \langle d_k, d_i \rangle \right) \langle P_{d_k^\perp}(p_k^{(2)}), P_{d_i^\perp}(p_i^{(2)}) \rangle \Bigg] \Bigg\} (x_k - x_i) \\
&\quad + \left[2\dot{h}_{ki} \langle P_{d_k^\perp}(p_k^{(2)}), p_i^{(1)} \rangle - 4\ddot{h}_{ki} \langle x_k - x_i, d_i \rangle \langle P_{d_k^\perp}(p_k^{(2)}), P_{d_i^\perp}(p_i^{(2)}) \rangle \right] d_k \\
&\quad - \left[2\dot{h}_{ki} \langle p_k^{(1)}, P_{d_i^\perp}(p_i^{(2)}) \rangle + 4\ddot{h}_{ki} \langle x_k - x_i, d_k \rangle \langle P_{d_k^\perp}(p_k^{(2)}), P_{d_i^\perp}(p_i^{(2)}) \rangle \right] d_i \\
\dot{p}_i^{(2)} &= \sum_{i=1}^{P} \left[2\dot{h}_{ki} \langle p_k^{(1)}, P_{d_i^\perp}(p_i^{(2)}) \rangle + 4\ddot{h}_{ki} \langle x_k - x_i, d_k \rangle \left\langle P_{d_k^\perp}(p_k^{(2)}), P_{d_i^\perp}(p_i^{(2)}) \right\rangle \right] (x_k - x_i) \\
&\quad + 2\dot{h}_{ki} \left\langle P_{d_k^\perp}(p_k^{(2)}), P_{d_i^\perp}(p_i^{(2)}) \right\rangle d_k \\
&\quad - \left\langle d_i, p_i^{(2)} \right\rangle \Big\{ 2\dot{h}_{ki} \langle x_k - x_i, d_i \rangle p_k^{(1)} + \Big[4\ddot{h}_{ki} \langle x_k - x_i, d_k \rangle \langle x_k - x_i, d_i \rangle \\
&\quad + 2\dot{h}_{ki} \langle d_k, d_i \rangle \Big] P_{d_k^\perp}(p_k^{(2)}) \Big\} \\
&\quad - \Big\{ 2\dot{h}_{ki} \langle x_k - x_i, d_i \rangle \langle p_k^{(1)}, d_i \rangle \\
&\quad + \left[4\ddot{h}_{ki} \langle x_k - x_i, d_k \rangle \langle x_k - x_i, d_i \rangle + 2\dot{h}_{ki} \langle d_k, d_i \rangle \right] \langle P_{d_k^\perp}(p_k^{(2)}), d_i \rangle \Big\} p_i^{(2)}
\end{cases}
$$

In the pushforward action case, these equations are:

$$
\begin{cases}
\dot{x}_i &= \sum_{k=1}^{P} h_{ki} p_k^{(1)} + 2\dot{h}_{ki} \langle x_k - x_i, u_k \rangle p_k^{(2)} \\
\dot{u}_i &= \sum_{k=1}^{P} -2\dot{h}_{ki} \langle x_k - x_i, u_i \rangle p_k^{(1)} - \left[4\ddot{h}_{ki} \langle x_k - x_i, u_i \rangle \langle x_k - x_i, u_k \rangle + 2\dot{h}_{ki} \langle u_i, u_k \rangle \right] p_k^{(2)} \\
\dot{p}_i^{(1)} &= \sum_{k=1}^{P} \Bigg\{ \left[2\dot{h}_{ki} \left\langle p_k^{(1)}, p_i^{(1)} \right\rangle + 4\ddot{h}_{ki} \langle x_k - x_i, u_k \rangle \left\langle p_k^{(2)}, p_i^{(1)} \right\rangle \right] \\
&\quad - \left[4\ddot{h}_{ki} \langle x_k - x_i, u_i \rangle \langle p_k^{(1)}, p_i^{(2)} \rangle \right. \\
&\quad + \left(8h_{ki}^{(3)} \langle x_k - x_i, u_k \rangle \langle x_k - x_i, u_i \rangle + 4\ddot{h}_{ki} \langle d_k, d_i \rangle \right) \langle p_k^{(2)}, p_i^{(2)} \rangle \Big] \Bigg\} (x_k - x_i) \\
&\quad + \left[2\dot{h}_{ki} \langle p_k^{(2)}, p_i^{(1)} \rangle - 4\ddot{h}_{ki} \langle x_k - x_i, u_i \rangle \langle p_k^{(2)}, p_i^{(2)} \rangle \right] u_k \\
&\quad - \left[2\dot{h}_{ki} \langle p_k^{(1)}, p_i^{(2)} \rangle + 4\ddot{h}_{ki} \langle x_k - x_i, u_k \rangle \langle p_k^{(2)}, p_i^{(2)} \rangle \right] u_i \\
\dot{p}_i^{(2)} &= \sum_{k=1}^{P} \left[2\dot{h}_{ki} \left\langle p_k^{(1)}, p_i^{(2)} \right\rangle + 4\ddot{h}_{ki} \langle x_k - x_i, u_k \rangle \left\langle p_k^{(2)}, p_i^{(2)} \right\rangle \right] (x_k - x_i) \\
&\quad + 2\dot{h}_{ki} \left\langle p_k^{(2)}, p_i^{(2)} \right\rangle u_k
\end{cases}
$$