Computational Anatomy from a Geometric point of view

François-Xavier Vialard

December 20, 2013

Abstract

Lecture notes for the Geometry, Mechanics and Control 8th Intl. Young Researchers Workshop 2013 held in Barcelona. This document presents the mathematical framework that has been developed in the field of computational anatomy in the last fifteen years. We will focus in particular on the use of right-invariant metrics on diffeomorphism groups and its recent developments.

1 Introduction

One goal of Computational Anatomy is to develop quantitative tools to study the statistical variability of anatomical shapes and help the practitioner with the diagnosis. Underlying this approach, there is a constitutive hypothesis which is that pathologies can be detected out of images of organs. It is known that pathologies such as Alzheimer disease entails a decay of the hippocampi (some part of the brain). Going beyond the change of global indicators (such as volume) by using the whole geometrical information might improve the results of statistical studies [LAFP11].

The mathematical modeling of shapes is far from being new and D'Arcy Thompson was probably the first to introduce the idea of studying their variability through an underlying *deformation* of a template (an average shape). Due to the increase of medical imaging data in the last twenty years, there was a need to develop quantitative methods applying such kind of ideas. To this end, Grenander laid down the fundations of pattern theory [Gre93, GM98]. It was further developed and implemented by Miller, Trouvé and Younes [Tro95, BMTY05, JM00, MTY06, YAM09]. In this note, we give a short introduction to this area from a mathematical point of view. This note is based on many sources but in particular [BH13, BMTY05, MG13, TV12]. There has been many other works in Computational Anatomy that we do not present/cite in this document. We hope that this note will give a short path to enter the field for the interested reader.

2 Diffeomorphic image matching

Even before being interested by a statistical description of shape variability, a problem of interest in medical imaging consists in registering two biomedical images. The main application is the to establish correspondences between different image acquisitions. In the rest of these notes, an image will be modeled as a scalar valued function defined on a domain (smooth domain) $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$. Finding correspondences between two images may or may not be symmetric with respect to the given images (for example, one image is noisy and the other is not so that their role in the registration may not be symmetric). Standard approaches for registration distinguish the given images: there is the source image I that will be deformed by a map $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, the deformed image is $I \circ \varphi^{-1}$. The other image J will be called the target image. For concrete applications in biomedical imaging, constraints have to be imposed on the transformation φ. A usual constraint is to require that φ is invertible and preserves the topology: it is a priori difficult to make sense of folding or collapsing tissues. In particular, diffeomorphisms of the domain satisfy this condition. Due to noise or difference in acquisition modalities, there is in general no transformation φ such that the deformed source image denoted by $\varphi \cdot I$ is equal to J. And if, by chance, there exists such a transformation, it may not be unique. From a computational viewpoint, solving for φ the equation $I \circ \varphi^{-1} = J$ is usually done via an optimization problem on the set of transformations denoted by $\mathcal{S}$:

$$\operatorname{argmin}_{\varphi \in \mathcal{S}} d(I \circ \varphi^{-1}, J)$$

where d is a function satisfying $d(I, J) = 0 \implies I = J$, such as a distance. In general, the first remark on the existence of φ implies that this minimization will not have a solution on $\mathcal{S}$ but this solution might exist on a bigger space. A practical solution for ill-posed inverse problems consists in adding a regularization term on the deformation, that penalizes "weird" deformations. The new minimization problem is then

$$\mathcal{J}(\varphi) = V(\varphi) + d(I \circ \varphi^{-1}, J)\,, \tag{1}$$

where $V(\varphi)$ is the "cost" associated with the deformation φ. With respect to the previous issues (existence and uniqueness), existence is obtained and also, most of the time, a generic uniqueness. Almost all (if not all) registration methods can be formulated as the minimization of a functional of type (1). Remark that, on purpose, we did not specify the domain for the optimization of the functional $\mathcal{J}$. Indeed, at this level of generality, one can incorporate the domain constraint in the function V, e.g. $V(\varphi)$ if $\varphi \notin \mathcal{S}$. In particular, the function V may include a prior knowledge on the space of admissible deformations, as well as statistical informations on the deformations: An idealized situation would be the following, the distribution of deformations among a population is known a priori so that it makes sense to use $V(\varphi) = -d(\varphi)$ or $V(\varphi) = -\log(d(\varphi))$, which would correspond to likelihood or log-likelihood maximisation.

In the situation where no prior information is available and the only constraint is the diffeomorphic one, different methods have been proposed to address it: generate diffeomorphisms (1) using the Lie-exponential, i.e. flows of vector fields constant in time and (2) using time-dependent vector fields.

For (1), the Lie-exponential introduces a regularization on the Lie algebra on the velocity field

$$V(\varphi) := \frac{1}{2}\|v\|_V^2$$

where V is a Hilbert space of smooth vector fields and φ is determined by v by the following autonomous ode: $\partial_t \varphi(t) = v \circ \varphi(t)$. It is well-known that the

Lie exponential is not surjective in infinite dimension: an example is given in [KW09] on $\mathrm{Diff}(S_1)$.
For (2), the regularization is given by

$$V(\varphi) := \frac{1}{2}\int_0^1 \|v_t\|_V^2 \,\mathrm{d}t$$

where $\partial_t \varphi(t) = v_t \circ \varphi(t)$ is the equation that determines φ for a given time dependent vector field v_t.

While the former approach is more computationnally demanding than the latter, yet it enjoys more mathematical properties that matter for applications. Thus, the mathematical framework for time-dependent flows will be presented in the next section. It is sometimes called LDDMM (Large Deformation Diffeomorphic Metric Mapping).

Plan of the note: After the construction of the group following Trouvé [Tro95] in SectionFlows, we present the convenient property of normal metrics on the space of shapes in Section 2.2. We then present the LDDMM registration method in the case of images in Section 2.3 and the link with EPDiff equation. We discuss the construction and the choice of inner product on V in section 3 and some developments on the similarity measure (as important as the deformation energy) in Section 4. In Section 5, we give a very brief overview of statistical tools on Riemannian manifolds and their development in the LDDMM framework. The end of the note is concerned with other models that are either new or not widely used in practical applications.

2.1 Flows and groups of diffeomorphisms

In order to have a well-defined flows of diffeomorphisms, it is convenient to work with vector fields that belong to $C^1(\Omega, \mathbb{R}^d)$. In what follows, we will denote by V a space of (sufficiently smooth) vector fields. To be precise, we define the notion of *admissibility* for a linear space of vector fields:

Definition 2.1 (Admissible space of vector fields). A separable Hilbert space of vector fields V defined on Ω is said admissible if

- for any $v \in V$, $v(x) = 0$ and $dv(x) = 0$ if $x \in \partial\Omega$.
- There exists a constant K such that, for all $v \in V$

$$\|v\|_{1,\infty} \leq K\|v\|_V \tag{2}$$

  where $\|v\|_{1,\infty}$ denotes the sup norm of v and its first derivative and $\|\cdot\|_V$ denotes the given Hilbert norm on V.

A well-known example of such a space is $H^s(\Omega, \mathbb{R}^d)$ for $s > d/2 + 1$. More generally, the Rellich-Kondrachov theorem states that $W^{j+m,p}(\Omega) \hookrightarrow C^j(\Omega)$ for $mp > d$. We refer the reader to Section 3 which will present an efficient way of constructing such spaces. We then have:

Proposition 2.2. Let $v \in L^1([0,1], V)$ be a time dependent vector field and consider the Banach space B_0 (where $B_i := C^i(\Omega, \mathbb{R}^d)$ endowed with the sup norm $\|\cdot\|_{i,\infty}$ of the first i derivatives). There exists a unique solution $\varphi \in W^{1,1}([0,1],B)$ to the flow equation

$$\begin{cases} \partial_t \varphi(t) = v(t) \circ \varphi(t) \\ \varphi(0) = \mathrm{Id} \ . \end{cases} \tag{3}$$

In addition, $\varphi(\Omega) \subset \Omega$.

We will also denote $\varphi(t)$ and $v(t)$ by φ_t and v_t.

Proof. Since V is admissible, $v \in L^1([0,1],V)$ implies $v \in L^1([0,1],B)$ and v is a L^1-Lipschitz function from $I \times B$ to B. The first part of the proposition follows from Theorem B.3.

The fact that $\varphi(\Omega) \subset \Omega$ is a direct consequence of uniqueness of solutions: if for some $t \in [0,1]$, $\varphi_t(x) \in \partial\Omega$, it implies that for all $t \in [0,1]$, $\varphi_t(x) \in \partial\Omega$. Indeed the first point of the proposition applies not only to a domain but also for any compact sets and thus for $\Omega = \{x\}$. In this case, the soluton is $\phi_t(x) = x$ since the boundary condition is $v_t(x) = 0$. □

In fact, we can improve the smoothness of φ.

Theorem 2.3. *Let $v \in L^1([0,1],V)$ be a time dependent vector field and $B := C^0(\Omega,\mathbb{R}^d)$ endowed with the sup norm. The solution $\varphi(t)$ is a continuous path in $\mathrm{Diff}^1(\Omega)$, the space of C^1 diffeomorphisms satisfying $\|D\varphi(t)\|_\infty \leq e^{\int_0^t K\|v(t)\|\,\mathrm{d}t}$.*

Proof. Let us first assume that $v \in L^1([0,1],B_2)$, then Theorem B.3 implies that φ is in $W^{1,1}([0,1],B_1)$. Looking at the equation defining $D\varphi$ by differentiating the flow equation, we have:

$$\partial_t D\varphi_t = Dv_t \circ \varphi_t \cdot D\varphi_t \tag{4}$$

This ode is again a Caratheodory differential equation on $C^0(\Omega, L(\mathbb{R}^d,\mathbb{R}^d))$ endowed with the sup norm.

We now consider an approximation of v_n of $v \in V$ by vector fields in $L^1([0,1],B_2)$. Using Proposition 2.2, we get that the solution φ_n is well-defined. Moreover, by a direct application of the dominated convergence theorem $t \to Dv_n(t) \circ \varphi_n(t)$ converges in L^1 to $Dv(t) \circ \varphi$ where φ is the flow of v. Finally, applying Proposition B.5, the solution $D\varphi_n$ uniformly converges to $D\varphi$. Since the space B_1 is complete, φ belongs to B_1.

In fact, φ is a diffeomorphism since its inverse is given by the flow of $t \to -v(1-t)$ and the last inequality is derived from equation (4) and Gronwall's lemma B.4. □

Remark 2.4. We can observe that

1. The solution is in fact absolutely continuous,
2. the previous proposition can be extended to smoother vector fields. The flow of B_k vector fields will be C^k and the inequality (2.3) can be generalized to higher derivatives.

One can relax the assumption $v \in B_i$ and replace it with $v \in L^2([0,1], H^s)$ where H^s is the Sobolev space of order s. It can be proven that for $s > d/2+1$, the flow of v is an absolutely continuous path in H^s. The case $s > d/2+2$ follows as in Proposition 2.2, and the other cases require approximation arguments. The proof of this can be found in [BV13].

We are now in position of formulating a well-posed variational problem that solves the diffeomorphic image matching problem using time dependent vector fields:

Theorem 2.5. *Let $I, J \in L^2(\Omega, \mathbb{R})$ and V be an admissible space of vector fields. The functional*

$$\mathcal{J}(v) = \int_0^1 \|v(t)\|_V^2 \,\mathrm{d}t + \|I \circ \varphi^{-1} - J\|_{L^2}^2 \,, \tag{5}$$

attains its infimum on $L^2([0,1], V)$.

Proof. The first term $\int_0^1 \|v(t)\|_V^2 \,\mathrm{d}t$ is a norm on a Hilbert space, so that it is lower semi-continuous w.r.t. the weak convergence. Since V is separable, $L^2([0,1], V)$ is also separable so that bounded balls are compact for the weak topology. The existence of a minimizer for $\mathcal{J}$ follows from the continuity of the flow on B_0 w.r.t. the weak topology, which is proven in the next lemma. Indeed, if I is a Lipschitz function, then $\|I \circ \varphi^n - I \circ \varphi\|_\infty \leq \mathrm{Lip}(I)\|\varphi^n - \varphi\|_\infty$ so that the result is clear. The more general case $I \in L^2$ follows via the application of the dominated convergence theorem using approximations by Lipschitz functions and the fact that $\mathrm{Jac}(\varphi^n)$ is bounded in $L^\infty(\Omega, \mathbb{R})$. ☐

Remark 2.6. In fact, the previous theorem is true for any similarity measure that is lower semi-continuous w.r.t. to the weak topology. Using the following lemma, are available similarity measures such as: $\sum_{i=1}^k \|\varphi(q_i) - x_i\|^2$ where q_i and x_i are two given sets of points.

Lemma 2.7. *The flow map $\Phi : L^2([0,1], V) \to B_0$ defined by $\Phi(v) = \varphi(1)$ is continuous for the weak topology on $L^2([0,1], V)$.*

Proof. For any $x \in \Omega$, we have

$$\begin{aligned} \|\varphi_t^n(x) - \varphi_t(x)\| &\leq \left\| \int_0^t v_s^n(\varphi_s^n(x)) - v_s(\varphi_s(x)) \,\mathrm{d}s \right\| \\ &\leq \int_0^t \|v_s^n(\varphi_s^n(x)) - v_s^n(\varphi_s(x))\| \,\mathrm{d}s + \left\| \int_0^t v_s^n(\varphi_s(x)) - v(\varphi_s(x)) \,\mathrm{d}s \right\| \,. \end{aligned}$$

Remark that the second term can be written as $\|m_x(v_n) - m_x(v)\|$ where

$$m_x(v) := \int_0^t v_s(\varphi_s(x)) \,\mathrm{d}s \,, \tag{6}$$

which is a continuous linear form on $L^2([0,1], V)$. In addition, the family m_x are equicontinuous w.r.t. $x \in \Omega$. Indeed,

$$\|m_x(v) - m_y(v)\| \leq \int_0^t \|v_s(\varphi_s(x)) - v_s(\varphi_s(y))\| \,\mathrm{d}s$$

$$
\begin{aligned}
&\leq \int_0^t K\|v_s\|_V \|\varphi_s(x) - \varphi_s(y)\| \, \mathrm{d}s \\
&\leq K' \int_0^t \|v_s\|_V \|x-y\| \, \mathrm{d}s \\
&\leq K' \sqrt{t} \, \|x-y\| \|v\|_{L^2([0,1],V)} \, .
\end{aligned}
$$

In particular, uniform convergence w.r.t. $x \in \Omega$ is obtained, i.e.

$$
\lim_{n \to \infty} \sup_{x \in \Omega} \left\| \int_0^t v_s^n(\varphi_s(x)) - v(\varphi_s(x)) \, \mathrm{d}s \right\| = 0 \, .
$$

Going back to the first inequality, we get, since v_n is a bounded sequence,

$$
\|\varphi_t^n(x) - \varphi_t(x)\|_\infty \leq \alpha_n(t) + \|\varphi_t^n(x) - \varphi_t(x)\|_\infty \int_0^t K\|v_n(s)\|_V \, \mathrm{d}s \, ,
$$

where $\alpha_n(t) = \sup_{x \in \Omega} \| \int_0^t v_s^n(\varphi_s(x)) - v(\varphi_s(x)) \, \mathrm{d}s \|$. The conclusion follows by application of Gronwall's lemma B.4. □

A consequence of the previous lemma is the following theorem that can be found in [Tro95] or [You08],

Theorem 2.8 (Trouvé). *The image of the map Φ (i.e. all the flows at time 1) is a group of C^1 diffeomorphisms.The distance defined on this group denoted by G_V by*

$$
d(\varphi, \psi) := \inf\{\|v\|_{L^2([0,1],V)} \mid \Phi(v) \circ \varphi = \psi\} \tag{7}
$$

makes the group a complete metric space and there exists a minimizing vector field realizing the distance between two given φ and ψ.

Proof. There is only one important point in proving that d is a distance:

$$
d(\varphi, \psi) = 0 \implies \varphi = \psi \, .
$$

Let us consider $\varphi \neq \psi$ then there exists $x \in \Omega$ such that $\varphi(x) \neq \psi(x)$ and consider a path φ_s joining φ to ψ. We thus have

$$
\|\varphi(x) - \psi(x)\| \leq \int_0^1 \|v_s(\varphi_s(x))\| \, \mathrm{d}s \leq \int_0^1 K\|v_s\|_V \, \mathrm{d}s \leq \|v\|_{L^2([0,1],V)} \tag{8}
$$

and it follows that $d(\varphi, \psi) \geq \frac{1}{K} \|\varphi(x) - \psi(x)\| > 0$.

The existence of a minimizer follows as in the previous theorem: if $v_n \in L^2([0,1],V)$ is a minimizing sequence satisfying the boundary conditions $\Phi(v_n) \circ \varphi = \psi$, by weak compactness of bounded balls, there exists a subsequence of v_n that weakly converges to v. By lower semi-continuity of the norm, v is a minimizer provided that $\Phi(v) \circ \varphi = \psi$ which is true by the previous lemma.

Let us show that G_V is a complete metric space: Consider φ_n a Cauchy sequence such that $d(\varphi_n, \varphi_{n+1}) \leq 1/2^n$, it is thus possible to concatenate the minimizing vector fields u_n between φ_n and φ_{n+1} to obtain a vector field $u_\infty \in L^2([0,1],V)$. It is easy to prove that its flow φ_∞ is the limit of φ_n. For a general Cauchy sequence, there exists a subsequence sqtisfying the above condition. This subsequence converges and consequently, the sequence as well. □

Remark 2.9. If the metric is not strong enough then this lower bound may vanish. For instance, the L^2 right-invariant metric gives a degenerate distance on the group of diffeomorphisms [MM05, BBHM12]. In fact, on S_1 the H^s right-invariant distance is known to be degenerate for $s \leq 1/2$, which is the critical index: If $s > 1/2$, the distance is not degenerate [BBHM13]. In addition, it is not degenerate for $s \geq 1$ in any dimension [MM05]. It is an open question to know what is the critical index for $\mathbb{R}^d$ where $d \geq 2$.

Remark 2.10. Following the proof of theorem (2.3), it can be proven that $G_{H^s(\mathbb{R}^d)} \subset \{\mathrm{Id} + f \mid f \in H^s(\mathbb{R}^d)\}$ for $s > d/2 + 2$. The case $s > d/2 + 1$ is more involved. In this case, the group $G_{H^s(\mathbb{R}^d)}$ is an open set of an affine space on $H^s(\mathbb{R}^d)$ and thus inherits a smooth manifold structure. It is not known if G_V has a natural differentiable structure for a general V. In the case of Gaussian kernels (see Section 3) It is probably an ILH-Lie group as defined by Omori [Omo74].

Relation to Lie groups. The group G_V constructed above is not a Lie group: Indeed, the tangent space at Id is not closed with respect to the Lie bracket. For instance, two vector fields in H^s have their Lie bracket in H^{s-1} in general. A general result by Omori [Omo78] states that:

Theorem 2.11. *If a connected Banach-Lie group G acts effectively, transitively and smoothly on a compact manifold, then G must be a finite dimensional Lie group.*

However, important structures are still available. For instance, G_{H^s} carries a smooth Riemannian structure:

Theorem 2.12. *For $s > d/2+1$, the group G_{H^s} is an infinite dimensional Riemannian manifold modelled on H^s, which is metrically complete and geodesically complete. In addition, between any two diffeomorphisms in G_{H^s}, there exists a minimizing geodesic. This minimizing geodesic is unique on a G_δ dense subset.*

The proof of this theorem can be found in [BV13]. This result is based on [EM70, TY05, Eke] and it improves a bit Theorem 9.1 in [MP10].

In what follows, we will need the definition of the Lie exponential:

Definition 2.13. The Lie-exponential map

$$\exp : V \mapsto G_V \tag{9}$$

is defined for any $v \in V$ by the flow at time 1 of the (constant) time dependent vector field $v(t) = v$.

In particular, the exponential curve $t \to \exp(tv)$ is a C^1 curve on $B_1(= C^1(\Omega, \mathbb{R}^d))$ and

$$\left.\frac{d}{dt}\right|_{t=0} \exp(tv) = v\,.$$

2.2 Right-invariant metric and left action

An important feature of the formulation (5) is that the right-invariant metric on the group descends to a Riemannian metric on the orbits of shapes. Such metrics are sometimes called normal metrics. By shape, we mean a mathematical object on which the diffeomorphism group acts in a smooth way: For instance, an embedding of S_1 or S_2 in $\mathbb{R}^3$, or even simpler a function in $L^2(S_1, \mathbb{R}^2)$ and the associated action is the composition on the left by the diffeomorphism.

Group actions. Let G_V be a group which is a manifold with a tangent space at Id denoted by V, acting from the *left* on a manifold Q. We denote the action by

$$\Phi : G \times Q \to Q\,, \quad (g, q) \mapsto g \cdot q := \Phi_g(q). \tag{10}$$

We will assume that this map is C^1. Being a left action means that Φ satisfies $g_1 \cdot (g_2 \cdot q) = (g_1 g_2) \cdot q$ and $\mathrm{Id} \cdot q = q$ for any $q \in Q$ and $g_1, g_2 \in G$.

A natural candidate for a metric on Q induced by the right-invariant metric is the following: Let $q_1, q_2 \in Q$ be two given points, the induced distance is the following

$$d(q_1, q_2) = \inf\{d(\mathrm{Id}, \varphi) \,|\, \varphi \in G_V \text{ s.t. } \varphi \cdot q_1 = q_2\}\,.$$

This distance may be degenerate, i.e. $d(q_1, q_2) = 0$ does not imply $q_1 = q_2$.

Example 2.14. Let $\varphi, \psi \in \mathrm{Diff}^\infty(\mathbb{R}^k)$, and d the right invariant metric associated to $H^1(\mathbb{R}^k)$. If $Q = \mathcal{L}_n := \{(q_1, \dots, q_n) \in [\mathbb{R}^k]^n \,|\, q_i \neq q_j \text{ for } i \neq j\}$, then the induced distance is degenerate. Note however that the distance between two diffeomorphisms does not vanish as proven in [MM05].

In what follows, we will be interested by non-degenerate distances and more than that, those that are Riemannian. Therefore, we look at the infinitesimal behaviour of the action and we will assume some smoothness hypothesis.

The *infinitesimal generator* of the action corresponding to $\xi \in V$ is the vector field on Q given by

$$\xi_Q(q) := \left.\frac{d}{dt}\right|_{t=0} \exp(t\xi) \cdot q = \xi \cdot q\,. \tag{11}$$

We will essentially use the notation $\xi \cdot q$ for the infinitesimal action. The tangent lift of Φ is defined as the action of G on TQ,

$$G \times TQ \to TQ, \quad (g, v_q) \mapsto g v_q := T\Phi_g(v_q), \tag{12}$$

with infinitesimal generator ξ_{TQ} corresponding to $\xi \in V$. Note that we have the relation

$$T\tau_Q(\xi_{TQ}(v_q)) = \xi_Q(q), \tag{13}$$

where $\tau_Q : TQ \to Q$ is the tangent bundle projection. Similarly, one defines the cotangent lifted action as

$$G \times T^*Q \to T^*Q, \quad (g, \alpha_q) \mapsto g\alpha_q := (T\Phi_{g^{-1}})^*(\alpha_q). \tag{14}$$

The momentum map $\mathbf{J} : T^*Q \to V^*$ associated with the cotangent lift of Φ is given by

$$\langle \mathbf{J}(\alpha_q), \xi \rangle_{V^* \times V} = \langle \alpha_q, \xi \cdot q \rangle_{T^*Q \times TQ}\,, \tag{15}$$

for arbitrary $\alpha_q \in T^*Q$ and $\xi \in V$. Using the definitions above, one has the important property that

$$\mathbf{J}(g\alpha_q) = \mathrm{Ad}^*_{g^{-1}} \mathbf{J}(\alpha_q)\,. \tag{16}$$

In the previous formula, we used $\mathrm{Ad}_g : V \mapsto C^0(\Omega, \mathbb{R}^d)$ defined by

$$\mathrm{Ad}_g(v) = Tg(v \circ g^{-1})\,. \tag{17}$$

This map is the differential at Id of the left action of G_V on itself defined by conjugation

$$\mathrm{Ad}_g h = ghg^{-1}\,. \tag{18}$$

The adjoint of Ad is defined by

$$\mathrm{Ad}_g(w) = Tg^*(w \circ g)\,\mathrm{Jac}(g)\,, \tag{19}$$

which satisfies

$$\langle w, \mathrm{Ad}_g(v)\rangle_{L^2} = \langle \mathrm{Ad}^*_g(w), v\rangle_{L^2}\,. \tag{20}$$

In a smooth setting, we have, if $\partial_t g(t) = v_t \circ g(t)$,

$$\partial_t \mathrm{Ad}_g(w) = \mathrm{ad}_{v_t} \mathrm{Ad}_g(w) \tag{21}$$

$$\partial_t \mathrm{Ad}^*_{g^{-1}}(w) = -\,\mathrm{ad}^*_{v_t} \mathrm{Ad}^*_{g^{-1}}(w) \tag{22}$$

and $\mathrm{ad}_v w = v \cdot w - w \cdot v = Dv(w) - Dw(v)$.

Normal metrics. We now consider an action which is *transitive* on a smooth manifold Q and consider a right-invariant Riemannian metric γ on the group G. By right-invariance, the metric g is completely determined by its evaluation at identity $\gamma(\mathrm{Id})(\cdot,\cdot)$.

$$\gamma(g)(X, Y) = \gamma(\mathrm{Id})(X \circ g^{-1}, Y \circ g^{-1})\,. \tag{23}$$

We will sometimes use the following notation, for $\xi \in V$, $\langle \xi, \xi\rangle = \|\xi\|_V^2 = \|\xi\|^2 = \gamma(\mathrm{Id})(\xi, \xi)$ and name it the norm on the Lie algebra (even it is not striclty speaking a Lie algebra). We are interested in the natural metric induced by the action of G on Q defined as follows:

Definition 2.15. For a given transitive left action $\Phi : G \times Q \to Q$ such that the map $E : (q, v) \in Q \times V \to (q, v \cdot q) \in TQ$ is a smooth (vector bundle) submersion, the *normal* metric associated with the right-invariant metric γ is the Riemannian metric on Q defined by

$$\gamma(q)(v_q, v_q) := \min\{\frac{1}{2}\|\xi\|_V^2 \,|\, \xi \in V \text{ s.t. } \xi \cdot q = v_q\}\,. \tag{24}$$

The *vertical subspace* at point q is defined as $F_q := \ker E(q, \cdot)$ which is a closed subspace of V and its orthogonal subspace $H_q := F_q^{\perp}$ is the *horizontal subspace* at point q. We denote by $L_q : T_qQ \mapsto H_q$ the pseudo-inverse of TE_q.

Proof. The minimum in the definition is indeed a minimum since for any $\xi \in V$ such that $\xi \cdot q = v_q$ the minimum of (24) is attained for the orthogonal projection of ξ on H_q. Indeed, $S_{v_q} := \{\xi \,|\, \xi \cdot q = v_q\}$ is a closed affine subspace of V of associated linear subspace F_q so that there exists an element of S_{v_q} of minimal norm by the Hilbert projection theorem. What remains to be checked is that the Riemannian structure is smooth which is the case since the map E is a submersion. □

Note that by construction, $(\ker T\pi_{q_0})^{\perp} \mapsto TQ$ is an isometry since

$$(\ker T\pi_{q_0})^{\perp} = \{\delta g \,|\, \langle \delta g g^{-1}, v\rangle = 0 \text{ for all } v \in V \text{ s.t. } v \cdot (g \cdot q_0) = 0\} = H_q\,.$$

In particular, we get

Proposition 2.16. Let $q_0 \in Q$ be a given point, then the map $\pi_{q_0} : G \to Q$ defined by $\pi_{q_0}(g) = g \cdot q$ is a Riemannian submersion.

Remark 2.17. Even if the manifold structure on the group may not be known, the normal metric might be a smooth Riemannian metric. Therefore, instead of proving a general theorem, it is sometimes a better path to write explicitely the metric in coordinates and check if it is smooth.

Definition 2.18 (*Horizontal lift* of paths)**.** Let $\pi : M \mapsto B$ be a Riemannian submersion, $a \in M$ and $x : [0,1] \mapsto B$ be a differentiable curve. There exists a unique path denoted $\tilde{x} : [0,1] \mapsto M$ such that $\tilde{x}(0) = a$ and $\pi \circ \tilde{x} = x$ defined by $\partial_t \tilde{x} = L_q(\partial_t x)$.

Remark 2.19. The lengths of $\tilde{x}$ and x are equal, this can replace the differential equation in the definition. This notion is similar to horizontal lifting on principal bundles.

Proposition 2.20. Let $q_0, q_{target} \in Q$ and d_Q a distance on Q. Minimizing

$$\mathcal{J}(v) = \int_0^1 \frac{1}{2}\|v(t)\|_V^2 \,\mathrm{d}t + d_Q(\varphi \cdot q_0, q_{target})\,, \tag{25}$$

reduces to minimizing

$$\mathcal{J}(q) = \int_0^1 \frac{1}{2}\gamma(q)(\dot{q},\dot{q}) \,\mathrm{d}t + d_Q(q(1), q_{target})\,. \tag{26}$$

The corresponding optimal solution of (25) is obtained by *horizontal lift* of the optimal solution of (26).

This property is actually very important when the dimension of Q is small. For instance, in the case of points matching or curve matching the dimensionality of the optimization problem is reduced to a few thousands. Clearly, it is of practical importance to write explicitely $\gamma(q)$.

Proposition 2.21. The cometric $\gamma(q)^{-1}$ can be written as:

$$\gamma(q)^{-1}(p,p) = \frac{1}{2}\|J(p,q)\|_{V^*}^2\,. \tag{27}$$

Proof. Exercise left to the reader. □

Applications to group of points (also called landmarks): The manifold of landmark is an open set of Ω^n defined by

$$\mathcal{L}_n := \{(q_1,\ldots,q_n) \in \Omega^n \,|\, q_i \neq q_j \text{ for } i \neq j\}\,.$$

It is a connected manifold if $d \geq 2$. It is probably the simplest case of application where the action is defined by

$$(g,(x_1,\ldots,x_n)) \mapsto (g(x_1),\ldots,g(x_n))$$

and the infinitesimal action is thus

$$v \cdot (x_1, \dots, x_n) \mapsto (v(x_1), \dots, v(x_n)) \,.$$

The momentum map is thus in coordinates

$$(x_1, \dots, x_n), (p_1, \dots, p_n) \in T^*Q \mapsto \sum_{i=1}^{n} \delta_{x_i}^{p_i} \in V^* \,. \tag{28}$$

Note that if V is an admissible space of vector fields, this application is well defined since Dirac operators belong to V^*. The tangent action is simply:

$$g \cdot (x_1, \dots, x_n), (v_1, \dots, v_n) \mapsto (g(x_1), \dots, g(x_n)) \,, (Tg_{x_1}(v_1), \dots, Tg_{x_n}(v_n)) \,.$$

and the co-tangent action is:

$$g \cdot (x_1, \dots, x_n), (p_1, \dots, p_n) \mapsto (g(x_1), \dots, g(x_n)) \,, \left(Tg_{x_1}^{-1*}(p_1), \dots, Tg_{x_n}^{-1*}(p_n)\right) \,.$$

Since V is a reproducing kernel Hilbert space (see Section 3), there exists a kernel k such that $(\delta_x^p, v)_{V^*,V} = \langle k(.,x)p, v\rangle_V$. In particular, we have, for $q = (x_1, \dots, x_n)$ and $p = (p_1, \dots, p_n)$,

$$\frac{1}{2}\|J(p,q)\|_{V^*}^2 = \frac{1}{2}\sum_{i,j=1}^{n} \langle p_i, k(x_i, x_j)p_j\rangle \,. \tag{29}$$

The optimal solutions for the minimization problem (26) satisfy the Hamiltonian equations:

$$\begin{cases} \dot{p}_i = -\partial_{q_i} H(p,q) = -\sum_{j=1}^{n} \langle p_i, \partial_1 k(q_i, q_j) p_j\rangle \\ \dot{q}_i = \partial_{p_i} H(p,q) = \sum_{j=1}^{n} k(q_i, q_j) p_j \,, \end{cases} \tag{30}$$

where

$$H(p,q) = \frac{1}{2}\sum_{i,j=1}^{n} \langle p_i, k(x_i, x_j)p_j\rangle \,.$$

Note that $\partial_p H(p,q) = \sum_{j=1}^{n} k(q_i, q_j)p_j$ which is simply the evaluation at point q_i of the vector field $v(x) = \sum_{j=1}^{n} k(x, q_j)p_j$.

Proposition 2.22. The landmark space $\mathcal{L}_n$ is a complete Riemannian manifold. In other words, solutions to System (30) are defined for all time.

Proof. Denote φ the horizontal lift of (p,q) to G_V, i.e. the flow of the time dependent vector field $K\mathbf{J}(p(t), q(t))$. Since the Hamiltonian function is constant along a solution, it means that $d^2(\mathrm{Id}, \varphi(t)) \leq cste\, t$. Using the bounds on the flow 2.3, it implies that (p,q) stays in a compact set of $T^*\mathcal{L}_n$ for t bounded. Thus the solutions can be extended for all time. □

Remark 2.23 (Peakons)**.** Note that the above result does not apply to the kernel associated with the H^1 norm $k(x,y) = e^{-\|x-y\|/\sigma}\,\mathrm{Id}$. Indeed, H^1 can not be continuously embedded in C^1 endowed with the sup norm, whatever the dimension. Solutions to System 30 for this particular kernel are called peakons [Hol09]. The Riemannian submersion approach still applies but the normal metric does not make $\mathcal{L}_n$ a complete Riemannian manifold.

Solving problem (26) is of course not reduced to an initial value problem (30). The optimization problem is reduced to:

Proposition 2.24. Solving (26) is equivalent to minimize:

$$\mathcal{J}(p(0)) = H(p(0), q(0)) + d_Q(q(1), q_{target}) \tag{31}$$

over $p(0) \in T_{q(0)}Q$ where $q(1)$ is the solution of the Hamiltonian equations (30).

Proof. Remark that the Hamiltonian function is constant along optimal solutions so that the first term in (26) can be replaced with $H(p(0), q(0))$. □

Numerically solving this problem is often called a shooting optimization method. Such an approach is developped in the case of images in [VRRC12] and the code is freely available at `https://sourceforge.net/projects/utilzreg/`.

2.3 A steepest gradient descent algorithm

The optimization problems (25) or (26) are solved via standard methods of differentiable nonlinear optimization such as steepest descent methods. In this section, we follow the lines of [BMTY05]. The code of the associated gradient descent algorithm is freely available at `https://sourceforge.net/projects/utilzreg/`. The first thing we need is the computation of the gradient of the functional w.r.t. $v \in L^2([0,1], V)$. The differentiation of the first term is straightforward whereas the differentiation of the similarity measure is more involved: It is sufficient to compute the variation of the flow $\varphi(1)$ w.r.t. $v \in L^2([0,1], V)$.

Lemma 2.25. *Let $u, v \in L^2([0,1], V)$ and $\varphi^{u+\varepsilon v}(t)$ be the flow at time 1 of the time dependent vector field $u + \varepsilon v$. We also denote by $\varphi(t)$ the flow of $u(t)$ and $\varphi_{s,t} = \varphi(t)\varphi(s)^{-1}$. Then,*

$$\left[\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \varphi^{u+\varepsilon v}(t) \right] \varphi(t)^{-1} = \int_0^t \mathrm{Ad}_{\varphi_{s,t}}(v(s))\,\mathrm{d}s\,. \tag{32}$$

Proof. Write the tangent linear model, with $\delta\varphi(t) = \frac{d}{d\varepsilon}\big|_{\varepsilon=0} \varphi^{u+\varepsilon v}(t)$,

$$\partial_t \delta\varphi(t) = v_t(\varphi(t)) + du_t(\varphi(t))\delta\varphi(t)\,. \tag{33}$$

Equation (33) is a Caratheodory ode on B_0 and solutions exists for all time $t \in [0,1]$. By the continuity of solutions w.r.t. pertubations (Proposition B.5), it is sufficient to prove (32) on a dense subset of vector fields. We can thus work in a smooth setting. Compose by $\varphi(t)^{-1}$ on the right of (33) to get,

$$\partial_t \delta\varphi(t)\varphi(t)^{-1} = v(t) + du(t)\delta\varphi(t)\varphi(t)^{-1}\,, \tag{34}$$

and explicit the first term on the "Lie algebra" as follows, denoting $w(t) = \delta\varphi(t)\varphi(t)^{-1}$,

$$\begin{aligned}
\partial_t \left(\delta\varphi(t)\varphi(t)^{-1}\right) &= \partial_t \delta\varphi(t)\varphi(t)^{-1} + d\delta\varphi(t)\partial_t\varphi(t)^{-1} \\
\partial_t w(t) &= v(t) + du(t)\delta\varphi(t)\varphi(t)^{-1} - d\delta\varphi(t)d\varphi(t)^{-1}u(t) \\
&= v(t) + du(t)w(t) - dw(t)u(t)
\end{aligned}$$

$$= v(t) + \mathrm{ad}_{u(t)}\, w(t)\,.$$

Using the fact that for a smoothly time dependent $w(t) \in V$,

$$\frac{d}{dt}\,\mathrm{Ad}_{\varphi(t)^{-1}}(w(t)) = \mathrm{Ad}_{\varphi(t)^{-1}}\,\partial_t w(t) - \mathrm{Ad}_{\varphi(t)^{-1}}[\mathrm{ad}_{u(t)}\, w(t)]\,, \tag{35}$$

we have

$$\left.\frac{d}{ds}\right|_{s=t} \mathrm{Ad}_{\varphi(s)^{-1}}(w(s)) = \mathrm{Ad}_{\varphi(t)^{-1}}\, v(t)\,. \tag{36}$$

Formula (32) is obtained by integration. □

Proposition 2.26. The gradient of the functional (25) is given by

$$\nabla_u \mathcal{J}(t) = u(t) + K\,\mathrm{Ad}^*_{\varphi_{t,1}}\left(\mathbf{J}(\alpha_{q(1)})\right)\,, \tag{37}$$

where $\alpha_{q(1)} := \partial_q\big|_{q=\phi_1\cdot q_0} d(q, q_{target})$ and K denotes the isomorphism from V^* to V (implicitely defined by equation (39)). Equivalently,

$$\nabla_u \mathcal{J}(t) = u(t) + K\mathbf{J}(\varphi_{1,t}\alpha_{q(1)})\,. \tag{38}$$

Proof. The first term $u(t)$ comes from the differentiation of $\frac{1}{2}\|u(t)\|^2$. The differentiation of second term with respect to $\varphi(1)$ can be written as

$$\delta_\varepsilon d(\varphi_1\cdot q_0, q_{target}) = \partial_q\big|_{q=\varphi_1\cdot q_0} d(q, q_{target})(\delta_\varepsilon \varphi(1)\cdot q_0)$$

and now write $\delta_\varepsilon\varphi(1)\cdot q_0 = \delta_\varepsilon\varphi(1)\varphi(1)^{-1}\varphi(1)\cdot q_0$ so that

$$\delta_\varepsilon d(\varphi_1\cdot q_0, q_{target}) = \partial_q\big|_{q=\varphi_1\cdot q_0} d(q, q_{target})\left(\int_0^1 \mathrm{Ad}_{\varphi_{s,1}}(v(s))\,\mathrm{d}s\cdot q(1)\right)\,.$$

We have

$$\delta_\varepsilon d(\varphi_1\cdot q_0, q_{target}) = \partial_q\big|_{q=\varphi_1\cdot q_0} d(q, q_{target})\left(\int_0^1 \mathrm{Ad}_{\varphi_{s,1}}(v(s))\,\mathrm{d}s\cdot q(1)\right)\,,$$

and finally,

$$\delta_\varepsilon d(\varphi_1\cdot q_0, q_{target}) = \int_0^1 \left\langle \mathrm{Ad}^*_{\varphi_{s,1}}\left(\mathbf{J}(\alpha_{q(1)})\right), v(s)\right\rangle_{V^*,V} \mathrm{d}s\,.$$

This concludes the proof of the first equality by putting the two the terms together and using the fact that for any $m \in V^*$ and $u \in V$,

$$\langle m, u\rangle_{V^*,V} = \langle Km, u\rangle_V\,. \tag{39}$$

Using the equivariance property (16), we get the result. □

A formal optimal control approach: In what follows, we retrieve the previous formulas using an optimal control approach, i.e. by a formal application of the Pontryagin Maximum Principle (PMP). This derivation can be made rigorous. The minimizers of Functional (25) are critical points of the augmented functional

$$\mathcal{J}(v,p,q) = \int_0^1 \frac{1}{2}\|v(t)\|_V^2 \, \mathrm{d}t + d_Q(\varphi \cdot q_0, q_{target}) + \int_0^1 (p(t), \dot{q} - v \cdot q)_{T_q^*Q, T_qQ} \, \mathrm{d}t \,, \tag{40}$$

which leads to the Euler-Lagrange equations:

$$\begin{cases} \dot{q} = v(t) \cdot q(t) \\ \dot{p} = -v^*(t) \cdot p(t) \\ v(t) = K\mathbf{J}(p(t), q(t)) \,, \end{cases} \tag{41}$$

where we used the notation $-v^*(t) \cdot p(t) := -dv^*(q(t))(p(t))$ which is the infinitesimal coadjoint action. The gradient of Functional (25) can be written as:

$$\nabla \mathcal{J}(t) = v(t) - K\mathbf{J}(p(t), q(t)) \,, \tag{42}$$

where $q(t)$ solves the forward flow equation $\dot{q} = v(t) \cdot q(t)$ with $q(0) = q_0$ and $p(t)$ solves the adjoint equation

$$\dot{p} = -v^*(t) \cdot p(t) \,, \tag{43}$$

with the boundary condition $p(1) = [\partial_q]_{q=q(1)} d_Q(q(1), q_{target})$. In particular, one has to solve the adjoint equation backward in time.

This formulation is particularly convenient for deriving optimality conditions. It is also possible to introduce additional constraints in this setting, such as constant volume for the shape.

Proposition 2.27. If $(p(t), q(t)) \in T^*Q$ is a solution to System (41), then $m(t) = \mathcal{J}(p(t), q(t))$ is a solution to the EPDiff equation

$$\partial_t m(t) + \mathrm{ad}^*_{Km(t)} m(t) = 0 \,. \tag{44}$$

Proof. By differentiation, we have,

$$\begin{aligned} \frac{d}{dt}\left(\mathbf{J}(p(t), q(t)), w\right)_{V^*,V} &= (\dot{p}, w \cdot q) + (p, w \cdot \dot{q}) \\ &= (-dv^*(q(t))(p(t)), w \cdot q) + (p, dw(v) \cdot q) \\ &= (p(t), -dv(t)(w) \cdot q(t)) + (p(t), dw(v(t)) \cdot q(t)) \\ &= \big(p(t), -\mathrm{ad}_{v(t)}(w) \cdot q(t)\big) \\ &= \Big(-\mathrm{ad}^*_{v(t)} \mathbf{J}(p(t), q(t)), w\Big)_{V^*,V} \end{aligned}$$

which gives the result. ☐

Remark 2.28 (Extension of geodesics). Note that this derivation stays at a formal level. It can be made rigorous in most of the situations we are looking at: in particular, when the action on T^*Q is well-defined. It can be proven that in such a situation, the geodesics can be extended for all time [TY05].

Remark 2.29 (Geodesic completeness). The optimal control approach may not be however the proper way to prove completeness of the group G_V (or some other actions) since the corresponding action of V on the tangent space is not sufficiently smooth. Using the smooth Riemannian structure in Theorem 2.12 may prove more useful. Unfortunately, this result is only known for Sobolev spaces.

Applications to images: In order to implement Formula (38), one needs to compute the momentum map for specified actions. We have already seen the action on landmarks in Section 2.2. We now look at the case of images, i.e. a linear space of real valued functions defined on Ω.

The natural action on points is the push-forward by the map and the corresponding action on functions is the pull-back with the inverse. Let us consider $\mathcal{I} = H^1(\Omega, \mathbb{R})$ as the space of functions, then the action is:

$$\begin{aligned}&(\varphi, I) \in G_V \times \mathcal{I} \mapsto I \circ \varphi^{-1} \in \mathcal{I} \\ &(v, I) \in V \times \mathcal{I} \mapsto -\nabla I \cdot v \in L^2(\Omega, \mathbb{R}) \,.\end{aligned}$$

Remark 2.30. Note that the infinitesimal action is not completely well-defined since $-\nabla I \cdot v \notin \mathcal{I}$. This is a very common phenomenon in analysis and this is a generic situation.

The momentum map is still well-defined by:

$$(P, I) \in L^2(\Omega, \mathbb{R}) \times H^1(\Omega, \mathbb{R}) \mapsto -P \nabla I \in V^* \,. \tag{45}$$

The action on the co-tangent space $L^2(\Omega, \mathbb{R})$ is defined by

$$(\varphi, P) \in G_V \times L^2(\Omega, \mathbb{R}) \mapsto \mathrm{Jac}(\varphi^{-1}) P \circ \varphi^{-1} \,. \tag{46}$$

Note that this co-tangent action is the transport of densities.

Corollary 2.31. *Let $I_0 \in H^1(\Omega, \mathbb{R})$ and $I_{target} \in L^2(\Omega, \mathbb{R})$ be respectively the source and target images. Then, the gradient of*

$$\mathcal{J}(v) = \int_0^1 \frac{1}{2} \|v(t)\|_V^2 \, \mathrm{d}t + \frac{1}{2} \|I_0 \circ \varphi^{-1} - I_{target}\|_{L^2}^2 \tag{47}$$

can be written as:

$$\nabla_u \mathcal{J}(t) = u(t) - K \, \mathrm{Jac}(\varphi_{t,1}) P \circ \varphi_{t,1} \nabla (I \circ \varphi_{t,1}) \,, \tag{48}$$

where $I := I_0 \circ \varphi^{-1}$ and $P := I - I_{target}$.

An important object that comes into play when implementing these equations is the operator K, the isomorphism from V^* to V. It would thus be more convenient to deal with the space V from the point of view of K, which is developed in the next section.

3 Construction of admissible spaces of vector fields

This section is essentally based on [MG13]. We refer the reader to this rather complete reference. The definition of admissibility 2.1 is satisfied by a wide variety of space of vector fields. The main condition is the injection in the space of C^1 vector fields.

$$\|v\|_{1,\infty} \leq K\|v\|_V \,. \tag{49}$$

This condition implies that Dirac operators belong to V^*. This property defines reproducing kernel Hilbert space (rkhs) and the theory of such spaces have been developed since the cornerstone paper by Aronszajn.

3.1 Reproducing kernel Hilbert spaces

Definition 3.1. A reproducing kernel Hilbert space (of vector fields) is a Hilbert space V of functions from Ω to $\mathbb{R}^d$ such that the pointwise evaluation maps denoted by $\delta_x : f \in H \mapsto f(x) \in \mathbb{R}^d$ are continuous. Denoting $K : V^* \mapsto V$ the Riesz isomorphism between V^* (the dual of V) and V, the reproducing kernel associated with the space V is defined by $k(x,y) = (\delta_x, K\delta_y) \in L(\mathbb{R}^d, \mathbb{R}^d)$, where the bracket $(\cdot,\cdot)$ denotes the dual pairing.

The kernel completely specifies the reproducing kernel Hilbert space V: we refer the reader to [Sai97] for more informations on RKHS, but we can cite:

Definition 3.2. A positive kernel of dimension d on a set X is a map $k : S \times S \mapsto L(\mathbb{R}^d)$ such that

1. for all $x, y \in S$, $k(x,y) = k(y,x)^*$
2. for any $n \in \mathbb{N}$, $x_1, \ldots, x_n \in S$ and $p_1, \ldots, p_n \in \mathbb{R}^d$

$$\sum_{i,j=1}^{n} \langle p_i, k(x_i,x_j)p_j \rangle \geq 0 \,. \tag{50}$$

The kernel is said to be striclty positive if inequality (50) is strict whenever there exists $p_i \neq 0$.

Theorem 3.3. *The kernel associated to a rkhs is positive. To each positive kernel corresponds a unique rkhs of functions defined on S with values in $\mathbb{R}^d$.*

Proof. The first assertion comes from the fact that

$$\left\| \sum_{i=1}^{n} \delta_{x_i}^{p_i} \right\|_{V^*}^2 = \sum_{i,j=1}^{n} \langle p_i, k(x_i,x_j)p_j \rangle \,. \tag{51}$$

The second assertion is obtained by defining

$$V = \overline{\operatorname{Span}\left\{k(.,x)p \,|\, x \in S \text{ and } p \in \mathbb{R}^d\right\}}$$

where the closure is taken with respect to the (semi) norm defined by formula (50). Note that V is still a space of functions. Indeed, the injection

$$\operatorname{Span}\left\{k(.,x)p \,|\, x \in S \text{ and } p \in \mathbb{R}^d\right\} \mapsto \mathcal{F}(S, \mathbb{R}^d)$$

is continuous for the topology of pointwise convergence on $\mathcal{F}(S, \mathbb{R}^d)$. We now prove that the positivity of the semi-norm. Let $x \in S$ and consider $\langle f(x), p \rangle = \langle \delta_x^p, f \rangle \leq \|\delta_x^p\| \|f\| = 0$, using the Cauchy-Schwarz inequality. Therefore, $f(x) = 0$ for any $x \in S$ and thus $f = 0$. $\square$

We will be particularly interested by kernels defined on $\mathbb{R}^d$. If the kernel is smooth, then the rkhs is composed of smooth functions of the same smoothness. When the kernel is translation invariant, the kernel $k(x, y)$ is in fact a function of the variable $x - y$ that we will denote with a little abuse of notation $k(x - y)$.

Theorem 3.4 (Bochner)**.** *Let $k \in L^1(\mathbb{R}^d, \mathbb{R}^d \times \mathbb{R}^d)$ with Fourier transform $\hat{k}$ also in $L^1(\mathbb{R}^d, \mathbb{R}^d \times \mathbb{R}^d)$. Then $k(x - y)$ is a positive kernel if and only if $\hat{k}(\xi)$ is a self-adjoint positive matrix definite matrix for all $\xi \in \mathbb{R}^d$.*

We can be also interested by kernels that are invariant by rotations. A strict subset of those are of the form $k_d(x, y) = k(\|x - y\|_{\mathbb{R}^d}) \operatorname{Id}$ and among those, functions k such that k_d is a kernel for any $d \in \mathbb{N}^*$ are characterized by a theorem of Schoenberg:

Theorem 3.5 (Schoenberg)**.** *A function $k : \mathbb{R}_+ \mapsto \mathbb{R}$ defines a kernel $k(\|x - y\|_{\mathbb{R}^d}) \operatorname{Id}$ for any $d \in \mathbb{N}^*$ if and only if it satisfies one of the two equivalent properties:*

1. *$f : r \to k(\sqrt{r})$ is a completely montonic function, i.e. for any $n \in \mathbb{N}$,*
$$(-1)^n f^{(n)}(t) \geq 0\,.$$
2. *There exists a finite Radon measure μ such that*
$$k(r) = \int_0^{+\infty} e^{-r^2 u^2} \, \mathrm{d}\mu(u)\,.$$

How to build kernels? In pratice, Gaussian kernels are widely used, however a wide range of kernels are available. The set of kernels is a cone, i.e. stable under addition and multiplication by a positive scalar. It is also stable by multiplication for scalar kernels. The addition of kernels has a nice variational interpretation: Let us consider a finite set of admissible Hilbert spaces H_i with kernels k_i and Riesz isomorphisms K_i between H_i^* and H_i for $i = 1, \ldots, n$. Denoting $H = H_1 + \ldots + H_n$, the space of all functions of the form $v_1 + \ldots + v_n$ with $v_i \in H_i$, the following norm can be defined on H:

$$\|v\|_H^2 = \inf \left\{ \sum_{i=1}^n \|v_i\|_{H_i}^2 \;\Big|\; \sum_{i=1}^n v_i = v \right\}. \tag{52}$$

The minimum is achieved for a unique n-tuple of vector fields and the space H, endowed with the norm defined by (52), is complete.

Proposition 3.6. The formula (52) induces a scalar product on H which makes H a RKHS, and its associated kernel is $k := \sum_{i=1}^n k_i$, where k_i denotes the kernel of the space H_i.

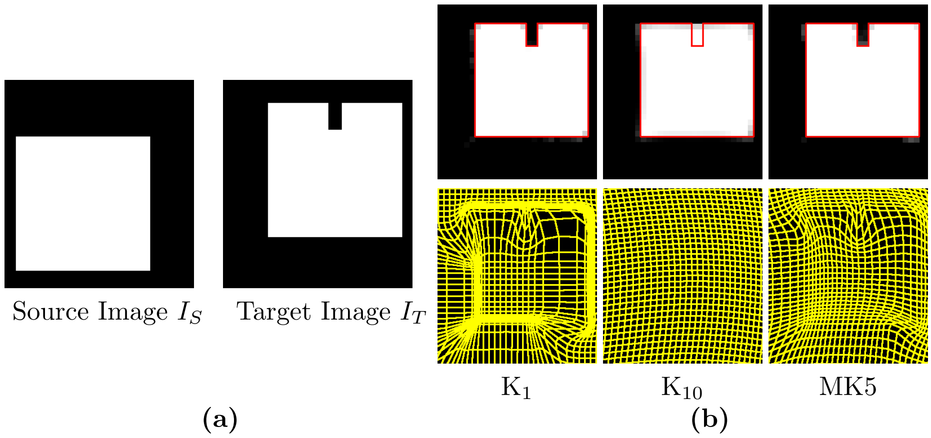

Figure 1: Influence of the smoothing kernel when registering two images containing feature differences at several scales simultaneously in the LDDMM framework. K_1 and K_{10} are Gaussian kernels of respectively 1 and 10 standard deviation, and MK5 is a sum of 5 Gaussian kernels. Better matching and plausible deformation is obtained at the cost of additional parameters with MK5.

In applications, the choice of the kernel is crucial to avoid poor local minima, as shown in [BRV12] (from which Figure 1 is taken), and also generate more plausible deformations. It is also possible to build kernels that are associated with spaces of divergence free or curl free vector fields and thus put more information in the variational functional.

In [BRV12], we also proposed a decomposition of the final deformation at different scales by interpretating the matching as standard LDDMM on a semi-direct product of groups.

4 Similarity measures

The image similarity measure has an important impact on the results of the gradient descent, i.e. change the local minimum found as shown in Figure 2[1]. Different norms can be used on the linear space of images and similarity measures such as mutual information have been applied. In general, the choice of norm should reflect the information at hand on the data and should be cheap (in comparison to the geometric distance) to compute.

Let us assume that the space of objects is the set of embeddings of S_2 in $\mathbb{R}^3$. Note however that there is no preferred parametrization that can be used to match one shape onto an other. Thus, it is natural to introduce similarity measures that are invariant under reparametrization. Such similarity measures can be developed in the framework of currents.

We denote by $\Omega_0^p(\mathbb{R}^d)$ the space of continuous differential forms of order p on $\mathbb{R}^d$ that vanish at ∞. Endowed with the sup norm, it forms a Banach space. Any compact oriented submanifold M of dimension k gives a continuous linear

[1] We thank Nicolas Charon for these pictures

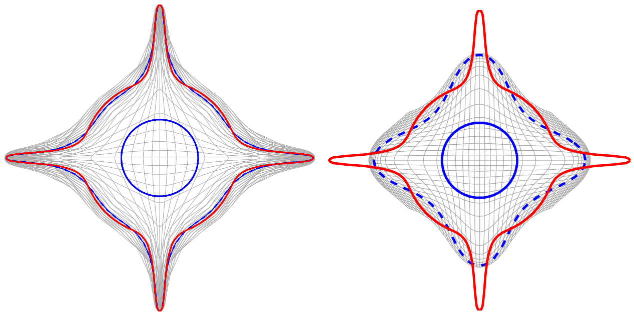

Figure 2: This figure shows a striking difference when switching from currents to varifolds to represent the curve. The red curve is the target shape and the template is the blue circle. The matching is clearly better when using varifold representation of data on this particular example where spiky shapes are present.

form on $\Omega_0^k(\mathbb{R}^d)$, by the formula

$$\omega \to \int_M \omega(x)(e_1, \ldots, e_k)\, \mathrm{d}\alpha(x)\,, \tag{53}$$

where α is the volume form of M.

Definition 4.1. The space of currents C_p of order p is the dual of $\Omega_0^p(\mathbb{R}^d)$.

Remark 4.2. The original definition given by Laurent Schwartz generalizes the approach of distributions and its Fréchet topology.

The natural left action associated with currents is the push forward since it is dual to differential forms that are acted upon by pull back.

Comparison of currents and associated norms: In order to compare such embeddings, a norm on the dual space of $\Omega_0^p(\mathbb{R}^d)$ would be sufficient, yet it should be easily computable on discrete data, for instance on a mesh. A mesh can be considered as an "unsmooth" submanifold and thus it has to be in the dual of a sufficiently smooth space of differential forms. We can assume that $W \hookrightarrow \Omega_0^p(\mathbb{R}^d)$ so that $C_p \hookrightarrow W^*$. Here again, since the pointwise evaluation is a continuous linear form, W is a rkhs and the norm on W is easy to compute so that it is natural to use the dual norm of W^* on C_p.

For instance, if $\gamma : S_1 \mapsto \mathbb{R}^2$ is a given curve, then the dual norm can be computed by

$$\|\gamma\|_{W^*}^2 = \int_{S_1^2} \langle \gamma'(t), k(\gamma(t), \gamma(s))\gamma'(s) \rangle \,\mathrm{d}t\,\mathrm{d}s\,, \tag{54}$$

where k is the underlying kernel. This is simply an average through the kernel that is invariant w.r.t. reparametrization. In general, such an approach can be extended to various type of representation of data using dual norms. We refer the reader to [CT13, CT12] for generalizations of this approach to functional currents or varifolds.

5 Statistical tools in a Riemannian setting

The Riemannian setting is convenient for generalizing statistical tools to manifolds. Hereafter, we describe some of them.

5.1 Fréchet/Kärcher mean

In what follows, we assume that we are interested in describing a population $(x_i)_{i=1,\ldots,n} \in M$ a Riemannian manifold. The Riemannian metric on M is associated with measuring deformations, which is already an a priori on the data.

The Riemannian metric endows the manifold with a distance: The very first interesting notion is the matrix of pairwise distance which gives some information on the population on which basic classification can be performed. A probably more interesting tool is the definition of average using the variational generalization of mean in Euclidean space:

$$\overline{x} = \arg\min_{x \in M} \sum_{i=1}^{n} d^2(x, x_i)\,. \tag{55}$$

This generalization is called Fréchet mean or Kärcher mean (probably depending on its uniqueness). Of course, other powers of the distance can be taken in this definition, the effect is then to favor or not the effect of outliers: For example, high values of p will shift the mean towards outliers. Existence of the Kärcher mean is generally achieved whereas uniqueness is lost in general.

Uniqueness of Kärcher means: From the mathematical point of view, uniqueness of the mean can be obtained on sufficiently small neighborhoods (depending on the curvature) due to the convexity of the distance [K$\ddot{7}$7]. It is not well-known however that in general, a direct consequence of [AF05] (in the spirit of [Eke]) uniqueness is obtained on a G_δ dense subset of M^n in infinite dimensions. In finite dimension, the G_δ can replaced by a set of full measure.

The application of such geometric mean in biomedical imaging have been popularized by Pennec et al. [Pen99, AFPA07].

5.2 Generalization of PCA

Once a mean shape, also called template, is computed, the population can be described by its modes of variations using PCA on the tangent space at the template or its generalization PGA (Principal Geodesic Analysis) as developped in [Fle04]. The difference between tangent PCA and PGA depends on the curvature of the space and distribution of the population. In practice, in finite dimension, a difference can be shown when there is high sectional curvature (see [SLN10]).

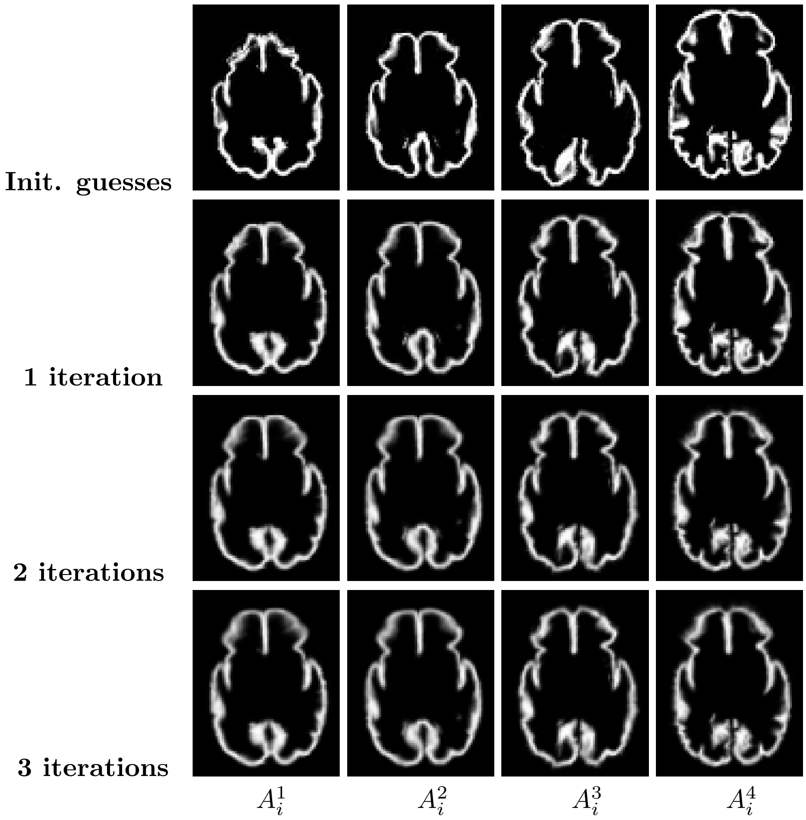

Figure 3: Average image estimates after 4 steps of a gradient descent with 4 different initializations on a population of 8 images.

5.3 Geodesic regression

The first step of PGA consists in representing the population using a geodesic and minimizing the residuals:

$$\underset{\text{x(t) is a geodesic}}{\operatorname{argmin}} \sum_{i=1}^{n} d^2(x(t), x_i)\,. \tag{56}$$

This formulation is a bit rough and one may add a regularization term to make it well posed. This representation can be used in the context of longitudinal evolutions. This approach is developped in [NHV11] in the case of images and requires the computation of Jacobi fields on the orbit of the image. Note that the adjoint equations associated with the geodesics can be written as, where $\lambda^I, \lambda^P, \lambda^v$ are the adjoint variables of I, P, v

$$\begin{cases} \lambda_t^I + \operatorname{div}(\lambda^I v + PK * \lambda^v) & = 0\,, \\ \lambda_t^P + (\nabla \lambda^P) \cdot v - \nabla I \cdot K * \lambda^v & = 0\,, \\ \lambda^v + \lambda^I \nabla I - \nabla \lambda^P P & = 0\,. \end{cases} \tag{57}$$

Note that System (57) are up to multiplication with the canonical symplectic matrix Jacobi fields equations. Numerically, these equations are not solved directly since the terms $P, I, \lambda^I, \lambda^P$ are unsmooth functions so that standard numerical schemes perform poorly. In fact, an integral equation of the system (57) based on the flow map is more efficient. This model has been extended to metamorphoses (see section 6.1). We also refer to [Fle13, BFH$^+$13] for geodesic regression.

5.4 Riemannian cubics

When looking at longitudinal evolutions, it is also natural to propose variational interpolation models such as cubics splines. Standard cubic splines in the Euclidean space has a variational definition which is acceleration minimization. This can be generalized to Riemannian manifold using covariant derivative denoted by $\frac{D}{Dt}$: On the set of curves $x : [0, T] \mapsto \mathbb{R}$ that interpolate the data, i.e. $x(t_i) = x_i$,

$$\operatorname{argmin} \int_0^1 \left\| \frac{D}{Dt} \dot{x} \right\|^2 \mathrm{d}t \,. \tag{58}$$

This has been developped in [TV12] in the context of LDDMM and this approach was originally developped in [NHP89] to interpolate camera motion and then mathematically developped in [CLP95, CL95]. More generally, the case of higher-order invariant Lagrangians has been treated in [GHM$^+$12]. In high dimension however, one should be careful that the space of geodesics is quite large as illustrated in Figure 5.4.

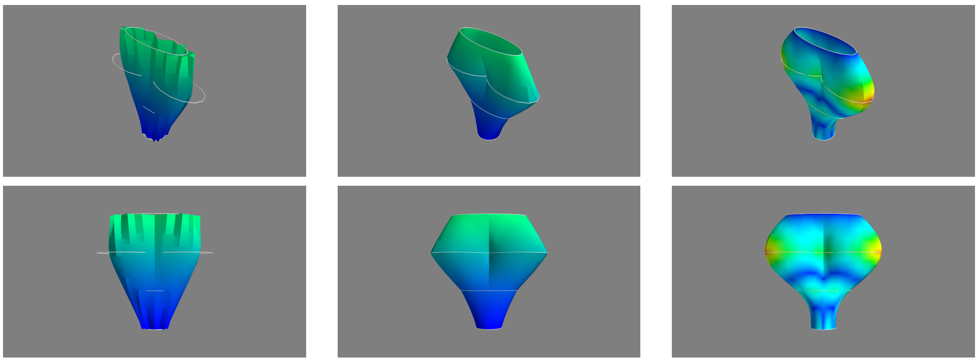

Figure 4: Geodesic regression, piecewise linear interpolation and Riemannian cubic interpolation of a sequence of 4 shapes that are deformations of a circle.

From the numerical point of view, we are of course not solving directly the Euler-Lagrange equation associated with the functional (58) but instead, we use the Hamiltonian formulation and introduce a forcing term u_t in the equations as follows:

$$\begin{cases} \dot{p}_t = -\partial_q H(p_t, q_t) + u_t \\ \dot{q}_t = \partial_p H(p_t, q_t) \end{cases} \tag{59}$$

and the minimization goes over the L^2 norm of u_t on T^*Q. However, we suggest in [TV12] to use different norms on u_t that might be related to noise on the data and the observation model.

5.5 Parallel transport

When studying longitudinal evolutions instead of static distribution of shapes, a new question appears: How to compare shape evolutions? Let us consider the problem of small evolutions such as the evolution of the hippocampus in the case of Alzheimer disease. One can model the evolution as a tangent vector on the space of shapes, i.e. TQ. A Riemannian approach is a natural way to explore and natural metrics on the tangent space are Sasaki metric [MF12] or Cheeger-gromoll metrics. Let us write the Sasaki metric:

$$\|(\dot{x}, \delta x)\|^2 = \|\dot{x}\|^2 + \left\| \frac{D}{Dt} \delta x \right\|^2 ,$$

where $\frac{D}{Dt}$ is the covariant derivative with respect to $\dot{x}$. A strict subset of geodesic equation for the Sasaki metric is

$$\nabla_{\dot{x}} w = 0 \text{ and } \nabla_{\dot{x}} \dot{x} + R(v, w)\dot{x} = 0 . \tag{60}$$

In particular, parallel transport will have a strong impact on Kärcher mean estimation when using this metric.

Another approach consists in defining a template on the space of shapes and then transport the longitudinal evolutions to the tangent space of this template. The first (historically) approaches used adjoint transport of the vector field, co-adjoint transport ot the momentum, they however suffer from issues. In particular, co-adjoint transport has been used in [FRC$^+$12] and gave promising results. More involved computationally, parallel transport under the Levi-Civita connection has been proposed in [You07, QYMC08]. Until now, there is no established transport method. Indeed, parallel transport under the Levi-Civita connection is an orthogonal map between tangent spaces but the associated Riemannian metric is somehow arbitrary and thus may bias the study. In addition, holonomy (i.e. parallel transport depends on the path) might be undesirable. Note that it is also possible to use connections that do not come a Riemannian metric, as done in [LP13].

6 Other models

Around the standard LDDMM model have been developed other models of interest.

6.1 Metamorphoses

Metamorphoses is a Riemannian metric developed to account for intensity variation of the image as well as geometric changes. The metric on the tangent space of $H^1(\Omega, \mathbb{R})$ is given by the minimization of

$$\|\delta I\|^2 = \operatorname{argmin}_{v \in V} \frac{1}{2} \|v\|_V^2 + \frac{\alpha}{2} \|\delta I + \nabla I \cdot v\|_{L^2}^2 . \tag{61}$$

Clearly, an careful analysis needs to be developed in order to prove existence of geodesics. This has been done in [TY05] and a more geometrical presentation is given in [HTY08].

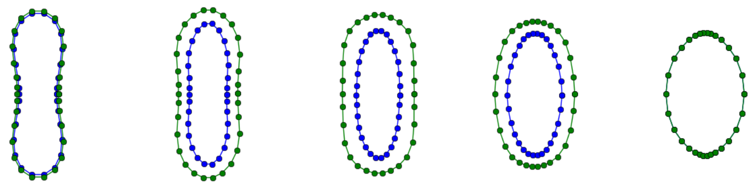

Figure 5: This figure shows snapshots of two deformations from the left-most source image to the right-most target image. The green curves show the optimal Right-LDM path (a right-geodesic), while blue curves show the optimal left path (a left-geodesic). Note that the paths are different, though both arrive at an exact match. The right-metric length of the green geodesic equals the left-metric length of the blue geodesic.

6.2 Left-invariant metrics

The right-invariant point of view on group of diffeomorphisms is analog to an Euler point of view on fluids. Therefore, kernels that are used in practice are translation and rotation invariant, not only to save computational cost. Indeed, operators that are used in practice are separable in Fourier space. In theory however, it is possible to use kernel that are spatially varying. Such an approach is fully justified when switching from right- to left- invariant metrics on the group of diffeomorphisms as shown in [SRV13].

The model is simply the minimization of the functional (5) under the *convective velocity* constraint:

$$\partial_t \phi(t) = d\phi(t) \cdot v(t) \,. \tag{62}$$

The inversion map on a Lie group is an isometry when switching from left- to right-invariant metric so that the final optimal map $\varphi(1)$ will be the same. However the optimal path will be different as shown in Figure 5.

Of course, what is lost using a left-invariant together with a left action is the induced metric on the space of shapes (in the presence of isotropy). However, this opens the possibility of tuning spatially varying metrics in situations of interest.

6.3 Other constraints

In construction.

A Bochner integral

In this section, we give a brief introduction to the Bochner integral based on [SG05] which is a generalization of Lebesgue integral on real valued functions to Banach valued functions.

Let B denote a real Banach space and $U \subset B$ be an open set. If $A \subset B$ is a set, the notation $\mathbf{1}_A$ is the indicator function of A defined by $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise.

Definition A.1. Let $(\Omega, \mathcal{A}, \mu)$ be a complete $\sigma-$finite measure space. A function $f : \Omega \to E$ is called a *step function* if it can be written as

$$f(x) = \sum_{i=1}^{n} \mathbf{1}_{E_i}(x)\, a_i\,, \tag{63}$$

where $n \geq 1$ and $E_i \in \mathcal{A}$ are measurable sets of finite measure and $a_i \in B$. The integral of f is defined by

$$\int_\Omega f \,\mathrm{d}\mu := \sum_{i=1}^{n} \mu(E_i)\, a_i$$

Definition A.2. A function $f : \Omega \to B$ is called $\mu-$*measurable* if it is almost everywhere ($\mu-$a.e.) the pointwise limit of step functions, i.e. there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of step functions such that $\mu-$a.e. $\lim\limits_{n\to\infty} \|f_n(t) - f(t)\| = 0$.

Note that this definition implies the convergence $\mu-$a.e. of the norms by using the triangle inequality $\lim\limits_{n\to\infty} |\|f_n(t)\| - \|f(t)\|| = 0$. Since $\|f_n(x)\| = \sum_{i=1}^{n} \mathbf{1}_{E_i}(x)\, \|a_i\|$ for a step function, this implies that $\|f_n\|$ is Lebesgue measurable. In particular, using Fatou's lemma, we have $\int_\Omega \|f\| \,\mathrm{d}\mu \leq \liminf\limits_{n\to\infty} \int \|f_n\| \,\mathrm{d}\mu$.

Remark A.3. This definition implies that for every $e \in B^*$ the dual space of B, the real function $e(f)$ is measurable.

Definition A.4. A measurable function $f : \Omega \to B$ is called $\mu-$*integrable* if there exists a sequence of step functions g_n converging $\mu-$a.e. on Ω to f such that $\lim\limits_{n\to\infty} \int_\Omega \|g_n - f\| \,\mathrm{d}\mu = 0$. The integral of f is defined by

$$\int_\Omega f \,\mathrm{d}\mu := \lim_{n\to\infty} \int_\Omega g_n \,\mathrm{d}\mu\,.$$

It can be checked easily that the limit does not depend on the chosen converging sequence g_n. Let us prove the following equivalence,

Proposition A.5. A measurable function f is $\mu-$*integrable* if and only if $\int_\Omega \|f\| \,\mathrm{d}\mu < \infty$.

Proof. Let us prove the first implication: By the triangle inequality, the definition implies that $| \int_\Omega \|f_m\| \,\mathrm{d}\mu - \int_\Omega \|f_n\| \,\mathrm{d}\mu| \leq \int_\Omega \|f_m - f\| \,\mathrm{d}\mu + \int_\Omega \|f_n - f\| \,\mathrm{d}\mu$, which shows that $\int_\Omega \|f_m\| \,\mathrm{d}\mu$ is a Cauchy sequence. Its limit is $\int_\Omega \|f\| \,\mathrm{d}\mu$ (again by triangle inequality).

We now prove the reverse implication in the case $\mu(\Omega) < \infty$ (the case where μ is only $\sigma-$finite follows easily). In addition, by restriction to the measurable set $\{x \in \Omega \,|\, \|f(x)\| > 0\}$, we can assume $\|f(x)\| > 0$ for almost all $x \in \Omega$.

Let f be a measurable function such that $\int_\Omega \|f\| \,\mathrm{d}\mu < \infty$. Since f is measurable, there exists a sequence f_n of step functions converging pointwisely to f. For such a given sequence, it is not true a priori that $\lim\limits_{n\to\infty} \int_\Omega \|f - f_n\| \,\mathrm{d}\mu = 0$, but this is true for a modification of f_n. Consider $\varepsilon > 0$,

$$A_{n,\varepsilon} := \{x \in \Omega \,|\, \|f_n(x) - f(x)\| > \varepsilon \|f(x)\|\}$$

and $B_{N,\varepsilon} = \cup_{n \geq N} A_{n,\varepsilon}$. By pointwise convergence of f_n, $\lim_{n\to\infty} \mu(B_{N,\varepsilon}) = 0$. Thus, for $\varepsilon = 1/k$, there exists $N(k)$ such that $\mu(B_{N(k),1/k}) \leq 1/k^2$. Let us define the following step function

$$g_k := \frac{1}{1+1/k} \mathbf{1}_{\Omega \setminus B_{N(k),1/k}} f_{N(k)} \,.$$

First, remark that $\|g_k\| \leq \|f\|$ since on $\Omega \setminus B_{N(k),1/k}$, we have

$$\|f_{N(k)}(x)\| - \|f(x)\| \leq \frac{1}{k} \|f(x)\| \,,$$

so that

$$\|f_{N(k)}(x)\| \leq (1 + \frac{1}{k}) \|f(x)\| \,.$$

In addition, $g_k(x)$ converges to $f(x)$ for almost all $x \in \Omega$: Indeed, we have

$$\|f(x) - g_k(x)\| \leq \|f(x) - f_{N(k)}(x)\| + (1 - \frac{1}{1+n}) \|f_{N(k)}(x)\| \leq \frac{2}{k}$$

on $\Omega \setminus \cup_{k \geq n} B_{N(k),1/k}$. Since $\mu(\cup_{k \geq n} B_{N(k),1/k}) \leq \sum_{k=n}^{\infty} \frac{1}{k^2} \underset{n\to\infty}{\longrightarrow} 0$, the result ensues.

□

Definition A.6. Let $p \geq 1$, the space $L^p(\Omega, \mu, B)$ is the Banach space of $\mu-$measurable functions (equivalence classes for the equivalence relation $f \sim g$ if $g = f$ $\mu-$a.e.) such that $\|f\| \in L^p(\Omega, \mu)$.

Hereafter are some direct consequences and other useful remarks:

- Continuous functions on an interval I, $C^0(I, B)$ are Bochner integrable.
- Composition of continuous maps are Bochner integrable.
- Let $I \subset \mathbb{R}$ be an interval and H a Hilbert space, then $L^2(I, H)$ is a Hilbert space.
- For $p \geq 1$, $W^{1,p}(I, H)$ defined as primitive integrals of elements in $L^p(I, B)$ is a Banach space. Notet that of $W^{1,p}(I, B) \subset C^0(I, B)$.
- $W^{1,2}(I, H)$ is a Hilbert space, and it is separable if H is separable.

We end this section with a proposition that will be used in the appendix on ODE:

Proposition A.7. Let I be an interval and $f : I \times B \to B$ a function that is Bochner measurable when the second variable is fixed and continuous in the second variable for almost every $t \in I$. Then, if $x : I \to B$ is continuous, the composition $t \to f(t, x(t))$ is Bochner measurable.

Proof. The proof is left to the reader as exercise. □

B ODE on Banach spaces

This section follows [O'R97].

B.1 Integration of ODE

In this section, we are concerned with the initial value problem of ordinary differential equations (ODE):
Let I be an interval, B be a Banach space and $f : I \times B \to B$ be a function. Find a function $x : I \to B$ such that

$$\begin{cases} x'(t) = f(t, x(t)) \\ x(0) = a \in B\,. \end{cases} \tag{64}$$

In what follows, we define the analytical framework suitable for our purpose. The class of functions f we want to work with is called *Caratheodory functions.*

Definition B.1. A function $f : I \times B \to B$ is an L^p-Caratheodory function if

1. The map $y \to f(t, y)$ is continuous for almost all $t \in I$,
2. For all $y \in B$, the map $t \to f(t, y)$ is Bochner measurable,
3. For every $r > 0$, there exists $h_r \in L^p(I, \mathbb{R})$ such that if $\|y\| \leq r$ then $|f(t, y)| \leq h_r(t)$ a.e. on I.

Definition B.2. A function $f : I \times B \to B$ is said to be L^p-Lipschitz if there exists $\alpha \in L^p(I, \mathbb{R})$ such that for all $x, y \in B$

$$\|f(t, x) - f(t, y)\| \leq \alpha(t)\|x - y\|\,,$$

for almost all $t \in I$.

Theorem B.3. *Let $f : [0, T] \times B \to B$ be a L^p-Caratheodory function and L^p-Lipschitz. Then, there exists a unique $x \in W^{1,p}([0, T], B)$ solving* (64).

Proof. We denote $I := [0, T]$. Define $A(t) = \int_0^t \alpha(s)\,\mathrm{d}s$, so that a.e. $A'(t) = \alpha(t)$. We introduce the norm on $C^0(I, B)$, $\|y\|_A = \sup_{x \in I} \|y(t)e^{-A(t)}\|$. This norm is equivalent to the standard sup norm since $e^{-A(t)}$ is bounded below and above by positive real numbers. Thus the space $(C^0(I, B), \|\cdot\|_A)$ is a Banach space.

The map $F : C^0(I, B) \to C^0(I, B)$ defined by

$$F(y)(t) = \int_0^t f(s, y(s))\,\mathrm{d}s\,,$$

is well-defined since $s \to f(s, y(s))$ is Bochner measurable by proposition A.7 and it is integrable by Proposition A.5: its norm is integrable using the Lipschitz property. Using standard property of Bochner integral and the fact that f is L^p-Lipschitz, we get

$$\begin{aligned} \|F(x)(t) - F(y)(t)\| &\leq \int_0^t \|f(s, x(s)) - f(s, y(s))\|\,\mathrm{d}s \\ &\leq \int_0^t \alpha(s)\|x(s) - y(s)\|\,\mathrm{d}s\,. \end{aligned}$$

Multiplying the previous inequality by $e^{-A(t)}$, we get:

$$e^{-A(t)}\|F(x)(t) - F(y)(t)\| \leq e^{-A(t)} \int_0^t \|f(s, x(s)) - f(s, y(s))\|\,\mathrm{d}s$$

$$
\begin{aligned}
&\leq e^{-A(t)} \int_0^t \alpha(s) e^{A(s)} e^{-A(s)} \| f(s, x(s)) - f(s, y(s)) \| \, \mathrm{d}s \\
&\leq e^{-A(t)} \| x - y \|_A \int_0^T \alpha(s) e^{A(s)} \, \mathrm{d}s \\
&\leq (1 - e^{-A(t)}) \| x - y \|_A \,.
\end{aligned}
$$

Noting that $k := \sup_{t \in I} 1 - e^{-A(t)} < 1$, it implies that F is a contraction

$$
\| F(x) - F(y) \|_A \leq k \| x - y \|_A \,, \tag{65}
$$

and existence and uniqueness of the solution in $C^0(I, B)$. Let x be this continuous solution. Since f is L^p-Lipschitz, the map $t \to f(t, x(t))$ is $L^p(I, B)$, so that $x \in W^{1,p}(I, B)$. $\square$

B.2 Continuity of solutions w.r.t parameters

We also state without proof Gronwall's lemma (see [You08] for a proof):

Lemma B.4 (Gronwall's lemma)**.** *Let f, a, b be three measurable positive real functions defined on the interval $[0, T]$ for $T > 0$. If*

$$
f(t) \leq a(t) + \int_0^t b(s) f(s) \, ds \,, \tag{66}
$$

then, for all $t \in [0, T]$

$$
f(t) \leq a(t) + \int_0^t a(s) b(s) e^{\int_0^s b(u) du} \, ds \,. \tag{67}
$$

A consequence of this lemma is the continuity of the solutions with respect to the initial condition. Another interesting perturbation that will be used in Section 2.1 is the following:

Proposition B.5. Let f_n be a bounded sequence of L^1-Lipschitz functions such that on any bounded sets in B, there exists a sequence of L^1 functions α_n s.t. $\| f_n(t, x) - f(t, x) \| \leq \alpha_n(t)$ with $\lim_{n \to \infty} \| \alpha_n \|_{L^1} = 0$ ($(f_n)_{n \in \mathbb{N}}$ converges uniformly to f on bounded sets) then $\lim_{n \to \infty} \sup_{t \in [0,1]} \| x_n(t) - x(t) \| = 0$.

Proof. First, remark that all the solutions $x_n(t)$ and $x(t)$ are bounded on $[0, 1]$. We have,

$$
\begin{aligned}
\| x_n(t) - x(t) \| &\leq \int_0^t \| f_n(s, x_n) - f(s, x) \| \, \mathrm{d}s \\
&\leq \int_0^t \| f_n(s, x_n(s)) - f(s, x_n(s)) \| \, \mathrm{d}s + \int_0^t \| f(s, x_n(s)) - f(s, x(s)) \| \, \mathrm{d}s \\
&\leq \int_0^t \alpha_n(s) \, \mathrm{d}s + \int_0^t M(s) \| x_n(s) - x(s) \| \, \mathrm{d}s \,.
\end{aligned}
$$

Applying Gronwall's lemma, we obtain, denoting $\int_0^t \alpha_n(s) \, \mathrm{d}s = A_n(t)$

$$
\| x_n(t) - x(t) \| \leq A_n(t) + \int_0^t A_n(s) M(s) e^{\int_0^s M(u) \, \mathrm{d}u} \, \mathrm{d}s
$$

$$\leq \sup_{t\in[0,1]} A_n(t) \left(1 + \int_0^t M(s) e^{\int_0^s M(u)\,\mathrm{d}u} \mathrm{d}s\right) .$$

This gives the result since $\|\alpha_n\|_{L^1} \to_{n\to\infty} 0$. □

References

[AF05] Daniel Azagra and Juan Ferrera. Proximal calculus on riemannian manifolds. *Mediterranean Journal of Mathematics*, 2(4):437–450, 2005.

[AFPA07] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.

[BBHM12] Martin Bauer, Martins Bruveris, Philipp Harms, and Peter W. Michor. Vanishing geodesic distance for the Riemannian metric with geodesic equation the KdV-equation. *Ann. Global Anal. Geom.*, 41(4):461–472, 2012.

[BBHM13] Martin Bauer, Martins Bruveris, Philipp Harms, and Peter W. Michor. Geodesic distance for right invariant Sobolev metrics of fractional order on the diffeomorphism group. *Ann. Global Anal. Geom.*, 44(1):5–21, 2013.

[BFH+13] Benjamin Berkels, P. Thomas Fletcher, Behrend Heeren, Martin Rumpf, and Benedikt Wirth. Discrete geodesic regression in shape space. In *EMMCVPR*, pages 108–122, 2013.

[BH13] M. Bruveris and D. D. Holm. Geometry of Image Registration: The Diffeomorphism Group and Momentum Maps. *ArXiv e-prints*, June 2013.

[BMTY05] M. Faisal Beg, Michael I. Miller, Alain Trouv, and Laurent Younes. Computing large deformation metric mappings via geodesic flow of diffeomorphisms. *International Journal of Computer Vision*, 61:139–157, 2005.

[BRV12] Martins Bruveris, Laurent Risser, and François-Xavier Vialard. Mixture of kernels and iterated semidirect product of diffeomorphisms groups. *Multiscale Model. Simul.*, 10(4):1344–1368, 2012.

[BV13] Martins Bruveris and François-Xavier Vialard. On groups of diffeomorphisms. *Preprint*, 2013.

[CL95] P. Crouch and F. Silva Leite. The dynamic interpolation problem: On Riemannian manifold, Lie groups and symmetric spaces. *Journal of dynamical & Control Systems*, 1:177–202, 1995.

[CLP95] M. Camarinha, F. Silva Leite, and P.Crouch. Splines of class $\mathcal{C}^k$ on non-euclidean spaces. *IMA Journal of Mathematical Control & Information*, 12:399–410, 1995.

[CT12] N. Charon and A. Trouvé. Functional Currents : a new mathematical tool to model and analyse functional shapes. *ArXiv e-prints*, June 2012.

[CT13] Nicolas Charon and Alain Trouvé. The varifold representation of non-oriented shapes for diffeomorphic registration. *CoRR*, abs/1304.6108, 2013.

[Eke] Ivar Ekeland. The hopf-rinow theorem in infinite dimension. *Journal of Differential Geometry*, 13.

[EM70] David G. Ebin and Jerrold Marsden. Groups of diffeomorphisms and the motion of an incompressible fluid. *Ann. of Math. (2)*, 92:102–163, 1970.

[Fle04] Thomas Fletcher. *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*. PhD thesis, University of North Carolina, 2004.

[Fle13] P. Thomas Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International Journal of Computer Vision*, 105(2):171–185, 2013.

[FRC+12] Jean-Baptiste Fiot, Laurent Risser, LaurentD. Cohen, Jurgen Fripp, and Franois-Xavier Vialard. Local vs global descriptors of hippocampus shape evolution for alzheimers longitudinal population analysis. In Stanley Durrleman, Tom Fletcher, Guido Gerig, and Marc Niethammer, editors, *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data*, volume 7570 of *Lecture Notes in Computer Science*, pages 13–24. Springer Berlin Heidelberg, 2012.

[GHM+12] F. Gay-Balmaz, D. D. Holm, D. M. Meier, T. S. Ratiu, and F.-X. Vialard. Invariant Higher-Order Variational Problems. *Communications in Mathematical Physics*, 309:413–458, January 2012.

[GM98] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, LVI(4):617–694, 1998.

[Gre93] Ulf Grenander. *General Pattern Theory.* Oxford Science Publications, 1993.

[Hol09] D. D Holm. Peakons. *ArXiv e-prints*, August 2009.

[HTY08] Darryl D. Holm, Alain Trouve, and Laurent Younes. The euler-poincare theory of metamorphosis. 2008.

[JM00] S.C. Joshi and M.I. Miller. Landmark matching via large deformation diffeomorphisms. *International Journal of Computer Vision*, 2000.

[K̈77] H. Kärcher. Riemannian center of mass and mollier smoothing. *Communications on Pure and Applied Mathematics*, pages 509–541, 1977.

[KW09] Boris A. Khesin and Robert Wendt. *The Geometry of Infinite-Dimensional Groups*, volume 51. Springer-Verlag, 2009.

[LAFP11] Marco Lorenzi, Nicholas Ayache, Giovanni B. Frisoni, and Xavier Pennec. Mapping the effects of Aβ_{1-42} levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 6892 of *LNCS*, pages 663–670. Springer, 2011.

[LP13] Marco Lorenzi and Xavier Pennec. Efficient parallel transport of deformations in time series of images: from schild's to pole ladder. *Journal of Mathematical Imaging and Vision*, October 2013.

[MF12] Prasanna Muralidharan and P. Thomas Fletcher. Sasaki metrics for analysis of longitudinal data on manifolds. In *CVPR*, pages 1027–1034, 2012.

[MG13] M. Micheli and J. A. Glaunès. Matrix-valued Kernels for Shape Deformation Analysis. *ArXiv e-prints*, August 2013.

[MM05] Peter W. Michor and David Mumford. Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. *Doc. Math.*, 10:217–245, 2005.

[MP10] Gerard Misioek and StephenC. Preston. Fredholm properties of riemannian exponential maps ondiffeomorphism groups. *Inventiones mathematicae*, 179(1):191–227, 2010.

[MTY06] Michael I. Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.*, 24(2):209–228, 2006.

[NHP89] L. Noakes, G. Heinzinger, and B. Paden. Cubic splines on curved spaces. *IMA Journal of Mathematical Control & Information*, 6:465–473, 1989.

[NHV11] M. Niethammer, Y Huang, and F.-X. Vialard. Geodesic regression for image time-series. In *MICCAI*, 2011.

[Omo74] Hideki Omori. *Infinite dimensional Lie transformation groups.* Lecture Notes in Mathematics, Vol. 427. Springer-Verlag, Berlin, 1974.

[Omo78] Hideki Omori. On Banach-Lie groups acting on finite dimensional manifolds. *Tôhoku Math. J.*, 30(2):223–250, 1978.

[O'R97] Donald. O'Regan. *Existence Theory for Nonlinear Ordinary Differential Equations.* Springer, 1997.

[Pen99] Xavier Pennec. Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In A.E. Cetin, L. Akarun, A. Ertuzun, M.N. Gurcan, and Y. Yardimci, editors, *Proc. of Nonlinear Signal and Image Processing (NSIP'99)*, volume 1, pages 194–198, June 20-23, Antalya, Turkey, 1999. IEEE-EURASIP.

[QYMC08] A Qiu, L Younes, M I Miller, and J G Csernansky. Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the alzheimer's type. *NeuroImage*, 40:68–76, 2008.

[Sai97] S. Saitoh. *Integral transforms, reproducing kernels and their applications.* Pitman Research Notes in Mathematics, 1997.

[SG05] S. Schwabik and Y. Guoju. *Topics in Banach Space Integration.* World Scientific, 2005.

[SLN10] Stefan Sommer, François Lauze, and Mads Nielsen. The differential of the exponential map, jacobi fields and exact principal geodesic analysis. *CoRR*, abs/1008.1902, 2010.

[SRV13] Tanya Schmah, Laurent Risser, and François-Xavier Vialard. Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In *MICCAI (1)*, pages 203–210, 2013.

[Tro95] Alain Trouvé. Action de groupe de dimension infinie et reconnaissance de formes. (infinite dimensional group action and pattern recognition). 1995.

[TV12] Alain Trouvé and Francçois-Xavier Vialard. Shape splines and stochastic shape evolutions: A second order point of view. *Quarterly of Applied Mathematics*, 2012.

[TY05] A. Trouvé and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59, 2005.

[VRRC12] François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and ColinJ. Cotter. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 97(2):229–241, 2012.

[YAM09] Laurent Younes, Felipe Arrate, and Michael I. Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1, Supplement 1):S40 – S50, 2009. Mathematics in Brain Imaging.

[You07] L. Younes. Jacobi fields in groups of diffeomorphisms and applications. *Quart. Appl. Math.*, 65:113–134, 2007.

[You08] Laurent Younes. *Shapes and Diffeomorphisms*. Springer, 2008.