

**Deformation analysis for shape and image
processing**

Laurent Younes

Contents

General notation	5
Chapter 1. Elements of Hilbert space theory	7
1. Definition and notation	7
2. Examples	8
3. Orthogonal spaces and projection	9
4. Orthonormal sequences	10
5. An indirect construction of L^2 : the Haar sequence of functions	11
6. The Riesz representation theorem	13
7. Embeddings and the duality paradox	14
8. Spline interpolation	17
9. Building V and its kernel	23
10. Weak convergence in a Hilbert space	30
Chapter 2. Ordinary differential equations and groups of diffeomorphisms	31
1. Introduction	31
2. A class of ordinary differential equations	33
3. Groups of diffeomorphisms	40
Chapter 3. Minimizing matching functionals	45
1. General principles	45
2. Matching terms and greedy descent procedures	45
3. Regularized matching	49
Chapter 4. Groups, manifolds and invariant metrics	57
1. Introduction	57
2. Differential manifolds	57
3. Submanifolds	60
4. Lie Groups	61
5. Group action	64
6. Riemannian manifolds	65
7. Gradient descent on Riemannian manifolds	72
Bibliography	75

General notation

The differential of a function $\varphi : \Omega \rightarrow \mathbb{R}^d$, Ω being an open subset of \mathbb{R}^k will be denoted $d\varphi : x \mapsto d_x\varphi$. It takes values in the space of linear operator from \mathbb{R}^k to \mathbb{R}^d so that $d_x\varphi$ can be assimilated to a $d \times k$ matrix (composed with the partial derivatives of the coordinates of φ).

Id_k is the identity matrix in \mathbb{R}^k .

$C_K^\infty(\Omega)$ is the set of infinitely differentiable functions on Ω , an open subset of \mathbb{R}^k .

Elements of Hilbert space theory

1. Definition and notation

A set H is a (real) Hilbert space if

- i) H is a vector space on \mathbb{R} (ie. addition and scalar multiplication are defined on H).
- ii) H has an inner product denoted $(h, h') \mapsto \langle h, h' \rangle_H$, for $h, h' \in H$. This inner product is a bilinear form, which is positive definite and the associated norm $\|h\|_H = \sqrt{\langle h, h \rangle_H}$.
- iii) H is a complete for the topology associated to the norm.

Converging sequences for the norm topology are simply sequences h_n for which there exists $h \in H$ such that $\|h - h_n\|_H \rightarrow 0$. Property iii) means that if a sequence $(h_n, n \geq 0)$ in H is a Cauchy sequence, ie. it collapses in the sense that, for every positive ε there exists $n_0 > 0$ such that $\|h_n - h_{n_0}\|_H \leq \varepsilon$ for $n \geq n_0$, then it necessarily has a limit: there exists $h \in H$ such that $\|h_n - h\|_H \rightarrow 0$ when n tends to infinity. If condition ii) is weakened to the fact that $\|\cdot\|_H$ is a norm (not necessarily induced by an inner product), one says that H is a Banach space.

If H satisfies i) and ii), it is called a pre-Hilbert space. On pre-Hilbert spaces, the Schwartz inequality holds

PROPOSITION 1 (Schwartz inequality). *If H is pre-Hilbert, and $h, h' \in H$, then*

$$\langle h, h' \rangle_H \leq \|h\|_H \|h'\|_H$$

The first consequence of this property is

PROPOSITION 2. *The inner product on H is continuous for the norm topology.*

PROOF. The inner product is a function $H \times H \rightarrow \mathbb{R}$. Letting $\varphi(h, h') = \langle h, h' \rangle_H$, we have, by Schwartz inequality, introducing a sequence (h_n) which converges to h

$$|\varphi(h, h') - \varphi(h_n, h')| \leq \|h - h_n\|_H \|h'\|_H \rightarrow 0$$

which proves the continuity with respect to the first coordinate, and also with respect to the second coordinate by symmetry. \square

Working with a complete normed vector space is essential when dealing with infinite sums of elements: if (h_n) is a sequence in H , and if $\|h_{n+1} + \dots + h_{n+k}\|_H$ can be made arbitrarily small for large n and any $k > 0$, then the series $\sum_{n=0}^{\infty} h_n$ has a limit in H . In particular, absolutely converging series in H converge:

$$(1) \quad \sum_{n \geq 0} \|h_n\|_H < \infty \Rightarrow \sum_{n \geq 0} h_n \text{ converges.}$$

We shall add a fourth condition to our definition of a Hilbert space

- iv. H is separable for the norm topology: there exists a denumerable subset \mathcal{S} in H such that, for any $h \in H$ and any $\varepsilon > 0$, there exists $h' \in \mathcal{S}$ such that $\|h - h'\| \leq \varepsilon$.

In the following, an Hilbert space will always be a separable Hilbert space.

A Hilbert space isometry between two Hilbert spaces H and H' is an invertible linear map $\varphi : H \rightarrow H'$ such that, for all $h, h' \in H$, $\langle \varphi(h), \varphi(h') \rangle_{H'} = \langle h, h' \rangle_H$.

A sub-Hilbert space of H is a subspace H' of H (ie. a non-empty subset stable by linear combination) which is closed for the norm topology. Closedness implies that H' is itself a Hilbert space, since Cauchy sequences in H' also are Cauchy sequences in H , hence converge in H , hence in H' since H' is closed. The next proposition shows that every finite dimensional subspace is a Hilbert subspace.

PROPOSITION 3. *If K is a finite dimensional subspace of H , then K is closed in H*

PROOF. let e_1, \dots, e_p be a basis of K . Let (h_n) be a sequence in K which converges to some $h \in H$: then h_n may be written $h_n = \sum_{k=1}^p a_{kn} e_k$ and for all $l = 1, \dots, p$:

$$\langle h_n, e_l \rangle_H = \sum_{k=1}^p \langle e_k, e_l \rangle_H a_{kn}$$

Let a_n be the vector in \mathbb{R}^p with coordinates $(a_{kn}, k = 1, \dots, p)$ and $u_n \in \mathbb{R}^p$ with coordinates $(\langle h_n, e_k \rangle_H, k = 1, \dots, p)$. Let also S be the matrix with coefficients $s_{kl} = \langle e_k, e_l \rangle_H$, so that the previous system may be written: $u_n = Sa_n$. The matrix S is invertible: if b belongs to the null space of S , a quick computation shows that

$$\left\| \sum_{i=1}^p b_i e_i \right\|_H^2 = {}^t b S b = 0$$

which is only possible when $b = 0$, since (e_1, \dots, e_n) is a basis. We therefore have $a_n = S^{-1}u_n$, and since u_n converges to u with coordinates $(\langle h, e_k \rangle_H, k = 1, \dots, p)$ (by continuity of the inner product), we obtain the fact that a_n converges to $a = S^{-1}u$. But this implies that $\sum_{k=1}^p a_{kn} e_k \rightarrow \sum_{k=1}^p a_k e_k$ and since the limit is unique, we have $h = \sum_{k=1}^p a_k e_k \in K$. \square

2. Examples

2.1. Finite dimensional Euclidean spaces. If $H = \mathbb{R}^n$ with $\langle h, h' \rangle_{\mathbb{R}^n} = \sum_{i=1}^n h_i h'_i$ is the standard example of a finite dimensional Hilbert spaces.

2.2. ℓ^2 space of real sequences. Let H be the set of real sequences $h = (h_1, h_2, \dots)$ such that $\sum_{i=1}^{\infty} h_i^2 < \infty$. Then H is a Hilbert space, the proof being left as an exercise.

2.3. L^2 space of functions. Let k and d be two integers. Let Ω be an open subset of \mathbb{R}^k . We define $L^2(\Omega, \mathbb{R}^d)$ as the set of all square integrable functions $h : \Omega \rightarrow \mathbb{R}^d$, with inner-product

$$\langle h, h' \rangle_{L^2} = \int_{\Omega} h(x) h'(x) dx$$

Integrals are taken with respect to the Lebesgue measure on Ω , and two functions which coincide everywhere except on a set of null Lebesgue measure are considered as equal. The fact that $L^2(\Omega, \mathbb{R}^d)$ is a Hilbert space is a standard result in integration theory.

If one is not familiar with Lebesgue integration and Lebesgue measure, one may think of the integral above as a Riemann integral. Of course the set of Riemann integrable functions is strictly included in the set of Lebesgue integrable functions, and in particular $L^2(\Omega, \mathbb{R}^d)$ would not be complete if it was restricted to Riemann integrable functions. We will see later a (less intuitive) definition of $L^2(\Omega, \mathbb{R}^k)$ which will not require Lebesgue integration.

3. Orthogonal spaces and projection

Let O be a subset of H . The orthogonal space of O is defined by

$$O^\perp = \{h : h \in H, \forall o \in O, \langle h, o \rangle_H = 0\}$$

THEOREM 1. O^\perp is a sub-Hilbert space of H .

Stability by linear combination is obvious, and closedness is a consequence of the continuity of the inner product.

When K is a sub Hilbert space of H (a closed subspace) and $h \in H$, one can define the variational problem:

$$(\mathcal{P}_K(h)) : \text{find } k \in K \text{ such that } \|k - h\|_H = \min \{\|h - k'\|_H : k' \in K\}$$

The following theorem is fundamental

THEOREM 2. If K is a closed subspace of H and $h \in H$, $(\mathcal{P}_K(h))$ has a unique solution, characterized by the property $k \in K$ and $h - k \in K^\perp$

DEFINITION 1. The solution of problem $(\mathcal{P}_K(h))$ in the previous theorem is called the orthogonal projection of h on K and will be denoted $\pi_K(h)$.

PROPOSITION 4. $\pi_K : H \rightarrow K$ is a linear, continuous function and $\pi_{K^\perp} = \text{id} - \pi_K$.

PROOF. Let $d = \inf \{\|h - k'\|_H : k' \in K\}$. The proof relies on the following construction: if $k, k' \in K$, a direct computation shows that

$$\left\| h - \frac{k + k'}{2} \right\|_H^2 + \|k - k'\|_H^2 / 4 = \left(\|h - k\|_H^2 + \|h - k'\|_H^2 \right) / 2$$

and the fact that $(k + k')/2 \in K$ implies $\left\| h - \frac{k+k'}{2} \right\|_H^2 \geq d^2$ so that

$$\|k - k'\|_H^2 \leq \left(\|h - k\|_H^2 + \|h - k'\|_H^2 \right) / 2 - d^2.$$

Now, from the definition of the infimum, one can find a sequence k_n in K such that $\|k_n - h\|_H^2 \leq d^2 + 2^{-n}$ for each n . The previous inequality implies that

$$\|k_n - k_m\|_H^2 \leq (2^{-n} + 2^{-m}) / 2$$

which implies that k_n is a Cauchy sequence, and therefore converges to a limit k which belongs to K and $\|k - h\|_H = d$. If the minimum is attained for another k' in K , we have, by the same inequality

$$\|k - k'\|_H^2 \leq (d^2 + d^2) / 2 - d^2 = 0.$$

so that $k = k'$ and uniqueness is proved, so that π_K is well defined.

Let $k = \pi_K(h)$, and $k' \in K$ and consider the function $f(t) = \|h - k - tk'\|_H^2$, which is by construction minimal for $t = 0$. We have $f(t) = \|h - k\|_H^2 - 2t\langle h - k, k' \rangle_H + t^2 \|k'\|_H^2$ and this can be minimal at 0 only if $\langle h - k, k' \rangle_H = 0$. Since this has to be true for every $k' \in K$, we obtain the fact that $h - k \in K^\perp$. Conversely, if $k \in K$ and $h - k \in K^\perp$, we have for any $k' \in K$

$$\|h - k'\|_H^2 = \|h - k\|_H^2 + \|k - k'\|_H^2 \geq \|h - k\|_H^2$$

so that $k = \pi_K(h)$.

We now prove the proposition: let $h, h' \in H$ and $\alpha, \alpha' \in \mathbb{R}$. Let $k = \pi_K(h)$, $k' = \pi_K(h')$; we want to show that $\pi_K(\alpha h + \alpha' h') = \alpha k + \alpha' k'$ for which it suffices to prove (since $\alpha k + \alpha' k' \in K$) that $\alpha h + \alpha' h' - \alpha k - \alpha' k' \in K^\perp$. But this is true since $\alpha h - \alpha k - \alpha' h' + \alpha' k' = \alpha(h - k) + \alpha'(h' - k')$, and $h - k \in K^\perp$, $h' - k' \in K^\perp$, and K^\perp is a vector space. Continuity comes from

$$\|h\|_H^2 = \|h - \pi_K(h)\|_H^2 + \|\pi_K(h)\|_H^2$$

so that $\|\pi_K(h)\|_H \leq \|h\|_H$.

Finally, if $h \in H$ and $k = \pi_K(h)$, then $k' = \pi_{K^\perp}(h)$ is characterized by $k' \in K^\perp$ and $h - k' \in (K^\perp)^\perp$. The first property is certainly true for $h - k$, and for the second, we need to show that $K \subset (K^\perp)^\perp$ which is a direct consequence of the definition of the orthogonal, and true in fact for any set O . \square

We have the interesting property

COROLLARY 1. *K is a sub-Hilbert space of H if and only if $(K^\perp)^\perp = K$*

PROOF. The \Leftarrow implication is a consequence of theorem 1, and the fact that $K \subset (K^\perp)^\perp$ is obviously true for any subset $K \subset H$, so that it suffices to show that every element of $(K^\perp)^\perp$ belongs to K . Assume that $h \in (K^\perp)^\perp$: this implies that $\pi_{K^\perp}(h) = 0$ but since $\pi_{K^\perp}(h) = h - \pi_K(h)$, this implies that $h = \pi_K(h) \in K$. \square

4. Orthonormal sequences

A sequence (e_1, e_2, \dots) in a Hilbert space H is orthonormal if and only if $\langle e_i, e_j \rangle_H = 1$ if $i = j$ and 0 otherwise. In such a case, if $\alpha = (\alpha_1, \alpha_2, \dots) \in \ell^2$, the series $\sum_{i=1}^{\infty} \alpha_i e_i$ converges in H (its partial sums form a Cauchy sequence) and if h is the limit, one has $\alpha_i = \langle h, e_i \rangle_H$.

Conversely, if $h \in H$ then the sequence $(\langle h, e_1 \rangle_H, \langle h, e_2 \rangle_H, \dots)$ belongs to ℓ^2 . Indeed, letting $h_n = \sum_{i=1}^n \langle h, e_i \rangle_H e_i$, one has $\langle h_n, h \rangle_H = \sum_{i=1}^n \langle h, e_i \rangle_H^2 = \|h_n\|_H^2$. On the other hand, one has, by Schwartz inequality $\langle h_n, h \rangle_H \leq \|h_n\|_H \|h\|_H$ which implies that $\|h_n\|_H^2 \leq \|h\|_H^2$: therefore

$$\sum_{i=1}^{\infty} \langle h, e_i \rangle_H^2 < \infty$$

Denoting by $K = \text{Hilb}(e_1, e_2, \dots)$ the smallest Hilbert subspace of H containing this sequence, one has the identity

$$(2) \quad K = \left\{ \sum_{n=1}^{\infty} \alpha_n e_n : (\alpha_1, \alpha_2, \dots) \in \ell^2 \right\}$$

This is left as an exercise. As a consequence of this, we see that $h \mapsto (\langle h, e_1 \rangle_H, \langle h, e_2 \rangle_H, \dots)$ is an isometry between $\text{Hilb}(e_1, e_2, \dots)$ and ℓ^2 . Moreover, we have, for $h \in H$

$$\pi_K(h) = \sum_{i=1}^{\infty} \langle h, e_i \rangle_H e_i$$

because $h - \pi_K(h)$ is orthogonal to every e_i .

An orthonormal set (e_1, e_2, \dots) is said *complete* in H if $H = \text{Hilb}(e_1, e_2, \dots)$. In this case, we see that H is itself isometric to ℓ^2 , and the interesting point is that this is always true

THEOREM 3. *Every (separable) Hilbert space has a complete orthonormal sequence.*

A complete orthonormal sequence in H is also called an orthonormal basis of H .

PROOF. The proof relies on the important Schmidt orthonormalization procedure. Let f_1, f_2, \dots be a dense sequence in H . We let $e_1 = f_{k_1} / \|f_{k_1}\|_H$ where f_{k_1} is the first non vanishing element in the sequence.

Assume that an orthonormal sequence e_1, \dots, e_n have been constructed with a sequence k_1, \dots, k_n such that $e_i \in V_n = \text{vect}(f_1, \dots, f_{k_i})$ for each i . First assume that $f_k \in \text{vect}(e_1, \dots, e_n)$ for all $k > k_n$: then $H = V_n$ is finite dimensional. Indeed, H is equal to the closure of (f_1, f_2, \dots) which is included in $\text{vect}(e_1, \dots, e_n)$ which is closed, as a finite dimensional vector subspace of H .

Assume now that there exists k_{n+1} such that $f_{k_{n+1}} \notin V_n$. Then, we may set

$$e_{n+1} = \lambda (f_{k_{n+1}} - \pi_{V_n}(f_{k_{n+1}}))$$

where λ is selected so that e_{n+1} has unit norm, which is always possible.

So there are two cases: either the previous construction stops at some point, and H is finite dimensional and the theorem is true, either the process carries on indefinitely, yielding an orthonormal sequence (e_1, e_2, \dots) and an increasing sequence of integers (k_1, k_2, \dots) . But $\text{Hilb}(e_1, e_2, \dots)$ contains (f_n) which is dense in H and therefore is equal to H so that the orthonormal sequence is complete. \square

5. An indirect construction of L^2 : the Haar sequence of functions

Dyadic intervals of \mathbb{R} are intervals $I_{k,q} = [k2^{-q}, (k+1)2^{-q}[$ where k and q belong to \mathbb{Z} . We define square integrable piecewise constant functions at resolution q by

$$V_q = \left\{ h : \mathbb{R} \rightarrow \mathbb{R}, h(t) = h_k \text{ on } I_{q,k} \text{ and } \sum_{k \in \mathbb{Z}} h_k^2 < \infty \right\}$$

V_q is a Hilbert space for the L^2 inner product

$$\langle h, h' \rangle_{L^2} = \int_{\mathbb{R}} h(t)h'(t)dt$$

and if $h(t) = h_k$ on $I_{q,k}$, we have $\|h\|_{L^2}^2 = 2^q \sum_{k \in \mathbb{Z}} h_k^2$. If one lets $\varphi_{q,k}$ be the indicator function of $I_{q,k}$, then the sequence $(2^{-q/2} \varphi_{q,k}, k \in \mathbb{Z})$ is an orthogonal basis of V_q . In particular, V_0 is the set of square integrable piecewise constant functions which are constant on intervals $[k, k+1[, k \in \mathbb{Z}$.

If f is a continuous, or a piecewise continuous function on \mathbb{R} , and if f is square integrable, its “orthogonal projection” on V_q can be computed as

$$f^{(q)}(t) = 2^q \sum_{k \in \mathbb{Z}} \int_{k2^{-q}}^{(k+1)2^{-q}} f(u) du \varphi_{q,k}(t)$$

In other terms, $f^{(q)}(t)$ is the mean value of f on the interval $I_{q,k}$ to which t belongs. When f is continuous at t , $f^{(q)}(t)$ converges to $f(t)$ when t tends to infinity. In any case, $f^{(q)}$ can be seen as a discretization of f on dyadic intervals.

Consider now the variation between two such approximations: $f^{(q)}(t) - f^{(q-1)}(t)$. This variation can be expressed in function of the Haar function, given by

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1/2 \\ -1 & \text{if } 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

Indeed, letting for $j, n \in \mathbb{Z}$

$$\psi_{j,n}(t) = 2^{-j/2} \psi\left(\frac{t - 2^j n}{2^j}\right)$$

we can write

$$f^{(q+1)}(t) - f^{(q)}(t) = \sum_{k \in \mathbb{Z}} \langle f, \psi_{q,k} \rangle_2 \psi_{q,k}(t)$$

with

$$\langle f, \psi_{q,k} \rangle_2 = \int_{\mathbb{R}} \psi_{q,k}(u) f(u) du.$$

Summing these expressions, we obtain

$$f^{(q+1)}(t) - f^{(0)}(t) = \sum_{j=0}^q \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle_2 \psi_{j,k}(t).$$

And, passing to the limit when f is continuous, we have

$$f(t) - f^{(0)}(t) = \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle_2 \psi_{j,k}(t).$$

Moreover, it is not very hard to show that

$$\int_{\mathbb{R}} \psi_{j,n}(t) \psi_{j',n'}(t) dt = 0$$

unless $j = j'$ and $n = n'$ in which case the integral is 1 so that $(\psi_{j,n}, j \geq q+1, n \in \mathbb{Z})$ is an orthonormal family. Moreover, each $\psi_{j,n}$ for $j \geq 0$ is also orthogonal to V_0 . When f is continuous and square integrable, it can be shown directly that

$$\int_{\mathbb{R}} f(t)^2 dt = \int_{\mathbb{R}} f^{(0)}(t)^2 dt + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle_2^2.$$

This invites us to the following definition. Let H_0 be the Hilbert space of sequences $a = (a_{q,k}; n \geq -1, k \in \mathbb{Z})$ such that

$$\sum_{q=-1}^{\infty} \sum_{k \in \mathbb{Z}} a_{q,k}^2 < \infty$$

For $a \in H_0$, we define a function $h_a^{(q)}$ on \mathbb{R} by

$$h_a^{(q)}(t) = \sum_{k \in \mathbb{Z}} a_{-1,k} \varphi_{0,k}(t) + \sum_{j=0}^q \sum_{k \in \mathbb{Z}} a_{j,k} \psi_{j,k}(t).$$

Notice that the second sum (over n) consists in fact of only one term, since, for a given j , there is only one n such that $\psi_{j,n}(t) \neq 0$, namely the unique integer, denoted n_j which belongs to $]2^j t - 1, 2^j t]$, so that there is no problem of convergence for the series in k in the previous formula. Clearly, $h_a^{(q)} \in V_q$, and letting q tend to infinity, we let

$$h_a(t) = \lim_{q \rightarrow \infty} h_a^{(q)}(t)$$

when the limit exists and $h_a(t) = 0$ otherwise. Then, we can define $L^2(\mathbb{R}, \mathbb{R})$ as the set

$$L^2(\mathbb{R}, \mathbb{R}) = \{h_a : a \in \ell(\mathbb{Z}^2)\}$$

and set

$$\langle h_a, h_b \rangle_{L^2} = \sum_{n \in \mathbb{Z}} \sum_{j \geq -1} a_{j,n} b_{j,n}$$

If h is a continuous, square integrable, function of \mathbb{R} , we have just seen that $h \in L^2(\mathbb{R}, \mathbb{R})$ and $h = h_a$ for $a_{j,n} = \int_{\mathbb{R}} h(t) \psi_{j,n}(t) dt$, and the L^2 norm which has been defined above coincides with the usual L^2 norm of h . If h is not continuous, but simply Riemann integrable, then the function h_a coincides with h everywhere excepted on a singular set which does not affect the value of the integrals.

The construction of $L^2(\mathbb{R}^n, \mathbb{R})$ can be made similarly, essentially replacing dyadic intervals by dyadic cubes. Finally, if Ω is an open subset in \mathbb{R}^n , we may define $L^2(\Omega, \mathbb{R})$ as the set of functions h on $L^2(\Omega, \mathbb{R})$ which vanish outside $\bar{\Omega}$.

Note that, by construction, piecewise constant functions are dense (for the L^2 norm) in $L^2(\mathbb{R}^n, \mathbb{R})$. This is still true (although less obvious with our construction) on $L^2(\Omega, \mathbb{R})$ for any open set Ω in \mathbb{R}^n . Finally, any piecewise continuous function can be expressed as a limit, in the least square sense, of a sequence of C^∞ functions with compact support, so that $C_0^\infty(\Omega)$ is dense in $L^2(\Omega, \mathbb{R})$.

6. The Riesz representation theorem

The dual space of a normed vector space H is the space containing all continuous linear functionals $\varphi : H \rightarrow \mathbb{R}$. It is denoted H^* , and we will often use the notation, for $\varphi \in H^*$ and $h \in H$:

$$(3) \quad \varphi(h) = (\varphi, h)$$

Thus, parentheses indicate linear forms, angles indicate inner products.

H being normed space, H^* also has a normed space structure defined by:

$$\|\varphi\|_{H^*} = \max \{(\varphi, h) : h \in H, \|h\|_H = 1\}$$

Continuity of φ is in fact equivalent to the finiteness of this norm.

When H is Hilbert, the function $\varphi_h : h' \mapsto \langle h, h' \rangle_H$ belongs to H^* , and by the Schwartz inequality $\|\varphi_h\|_{H^*} = \|h\|_H$. The Riesz-Nagy representation theorem states that there exists no other continuous linear form on H .

THEOREM 4 (Riesz). *Let H be a Hilbert space. If $\varphi \in H^*$, there exists a unique $h \in H$ such that $\varphi = \varphi_h$.*

PROOF. Uniqueness comes from the fact that if $\langle h, h' \rangle_H = 0$ for all $h' \in H$, then it is also true for $h' = h$ which yields $h = 0$. To prove existence, we use, as in finite dimension, the orthogonal of the null space of φ . So, let $K = \{h' \in H, (\varphi, h') = 0\}$; K is a closed linear subspace of H (because φ is linear and continuous). If $K^\perp = \{0\}$, then $K = (K^\perp)^\perp = H$ and $\varphi = 0 = \varphi_0$, which proves the theorem in this case. So, we assume that $K^\perp \neq 0$.

Now, let h_1 and h_2 be two non-zero elements in K^\perp . We have $\varphi(\alpha_1 h_1 + \alpha_2 h_2) = \alpha_1 \varphi(h_1) + \alpha_2 \varphi(h_2)$, so that it is always possible to find non-vanishing α_1 and α_2 such that $\varphi(\alpha_1 h_1 + \alpha_2 h_2) = 0$ since K^\perp is a vector space. We get $\alpha_1 h_1 + \alpha_2 h_2 = 0$ so that K^\perp has dimension 1. Fix $h_1 \in K^\perp$ such that, say $\|h_1\|_H = 1$. Using orthogonal projections, and the fact that K is closed, any vector in H can be written: $h = \langle h, h_1 \rangle h_1 + k$ with $k \in K$. This implies

$$(\varphi, h) = \langle h, h_1 \rangle (\varphi, h_1)$$

so that $\varphi = \varphi(\varphi, h_1) h_1$.

□

7. Embeddings and the duality paradox

7.1. Definition. Assume that H and H_0 are two Banach spaces. An embedding of H in H_0 is a continuous, one to one, linear map from H to H_0 , ie. a map $\iota : H \rightarrow H_0$ such that, for all $h \in H$,

$$(4) \quad \|\iota(h)\|_{H_0} \leq C \|h\|_H$$

The embedding is compact when the set $\{\iota(h), \|h\|_H \leq 1\}$ is compact in H_0 . In the separable case (to which we restrict), this means that for any bounded sequence $(h_n, n > 0)$ in H , there exists a subsequence of $(\iota(h_n))$ which converges in H_0 .

In all the applications we will be interested in, H and H_0 will be function spaces, and we will have a set inclusion $H \subset H_0$. For example H may be a set of smooth functions and H_0 a set of less smooth functions (see the examples of embeddings below). Then, one says that H is embedded (resp. compactly embedded) in H_0 if the canonical inclusion map: $\iota : H \rightarrow H_0$ is continuous (resp. compact).

If φ is a continuous linear map on H_0 , and $\iota : H \rightarrow H_0$ is an embedding, then one can define the form $\iota^* \varphi$ on H by $(\iota^* \varphi, h) = (\varphi, \iota(h))$, and $\iota^* \varphi$ is continuous on H . Indeed, we have, for all $h \in H$:

$$|(\iota^* \varphi, h)| = |(\varphi, \iota(h))| \leq \|\varphi\|_{H_0^*} \|\iota(h)\|_{H_0} \leq C \|\varphi\|_{H_0^*} \|h\|_H$$

where the first inequality comes from the continuity of φ and the last one from (4). This proves the continuity of $\iota^* \varphi$ as a linear form on H together with the inequality:

$$\|\iota^* \varphi\|_{H^*} \leq C \|\varphi\|_{H_0^*} .$$

This in fact proves the theorem:

THEOREM 5. *If $\iota : H \rightarrow H_0$ is a Banach space embedding, then the map $\iota^* : H_0^* \rightarrow H^*$ also is an embedding.*

7.2. Examples.

7.2.1. *Banach spaces of continuous functions.* Let Ω be an open subset of \mathbb{R}^k . The space of continuous functions on Ω with at least p continuous derivatives will be denoted $C^p(\Omega, \mathbb{R})$. If Ω is bounded, and thus $\bar{\Omega}$ is compact, $C^p(\bar{\Omega}, \mathbb{R})$ is the set of functions on Ω which are p times differentiable on Ω , each partial derivative being extended to a continuous function on $\bar{\Omega}$; $C^p(\bar{\Omega}, \mathbb{R})$ has a Banach space structure when equipped with the norm

$$\|f\|_{p,\infty} = \max_h \|h\|_\infty$$

where h varies in the set of partial derivatives of f or order lower or equal to p .

Still under the assumption that Ω is bounded, and denoting $\partial\Omega$ the boundary of Ω we let $C_0^p(\bar{\Omega}, \mathbb{R})$ be the set of functions f in $C^p(\bar{\Omega}, \mathbb{R})$ with vanishing derivatives up to order p on $\partial\Omega$.

We obviously have, almost by definition, the fact that $C^p(\bar{\Omega}, \mathbb{R})$ is embedded in $C^q(\bar{\Omega}, \mathbb{R})$ as soon as $p \leq q$. In fact the embedding is compact when $p < q$.

Compact sets in $C^0(\bar{\Omega}, \mathbb{R})$ are exactly described by Ascoli's theorem: they are bounded subsets $M \subset C^0(\bar{\Omega}, \mathbb{R})$ (for the supremum norm) which are uniformly continuous, meaning that, for any $x \in \bar{\Omega}$, for any $\varepsilon > 0$ there exists $\eta > 0$ such that,

$$\sup_{y \in \bar{\Omega}, |x-y| < \eta} \sup_{h \in M} |h(x) - h(y)| < \varepsilon.$$

Compact sets in $C^p(\bar{\Omega}, \mathbb{R})$ are bounded subsets of $C^p(\bar{\Omega}, \mathbb{R})$ over which all the p th partial derivatives are uniformly continuous.

7.2.2. *Hilbert Sobolev spaces.* Let $\Omega \subset \mathbb{R}^k$. We now define the space $H^1(\Omega, \mathbb{R})$ of functions with square integrable weak derivatives. A function u belongs to this set if and only if $u \in L^2(\Omega, \mathbb{R})$, and for each $i = 1, \dots, k$, there exists a function $u_i \in L^2(\Omega, \mathbb{R})$ such that for any function $\varphi \in C^\infty$ with compact support in Ω , one has

$$\int_{\Omega} u(x) \frac{\partial \varphi}{\partial x_i}(x) dx = - \int_{\Omega} u_i(x) \varphi(x) dx$$

The function u_i is called the (weak) directional derivative of u , and will be denoted $\frac{\partial u}{\partial x_i}$. The integration by parts formula shows that this coincides with the standard partial derivative when it exists.

$H^1(\Omega, \mathbb{R}^k)$ has a Hilbert space structure with the inner product:

$$\langle u, v \rangle_{H^1} = \langle u, v \rangle_{L^2} + \sum_{i=1}^k \left\langle \frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i} \right\rangle_{L^2}$$

(exercise: check completeness).

The space $H^m(\Omega, \mathbb{R})$ can now be defined by induction as the set of functions $f \in H^1(\Omega, \mathbb{R})$ with all partial derivatives belonging to $H^{m-1}(\Omega, \mathbb{R})$. Partial derivatives of increasing order are defined by inductions, and the inner product on H^m is the sum of the L^2 inner products of all partial derivatives up to order m .

There are theorems which allow to obtain embeddings in classic spaces of functions. We will be interested with a special case of Morrey's theorem, which is stated below:

THEOREM 6. *Let $\Omega \subset \mathbb{R}^k$, and assume that $m - k/2 > 0$. Then, for any $j \geq 0$, $H^{j+m}(\Omega, \mathbb{R})$ is embedded in $C^j(\bar{\Omega}, \mathbb{R})$. If Ω is bounded, the embedding is compact.*

Moreover, if $\theta \in]0, m - k/2]$, and $u \in H^{j+m}(\Omega, \mathbb{R})$, then every partial derivative, h , of order j has Hölder regularity θ : for all $x, y \in \Omega$

$$|h(x) - h(y)| \leq C \|u\|_{H^{m+j}} |x - y|^\theta$$

Let Ω be bounded. As a final definition, we let $H_0^m(\Omega, \mathbb{R})$ be the completion in $H_0^m(\Omega, \mathbb{R})$ of the set of C^∞ functions with compact support in Ω : u belongs to $H_0^m(\Omega, \mathbb{R})$ if and only if $u \in H^m(\Omega, \mathbb{R})$ and there exists a sequence of functions u_n , C^∞ with compact support in Ω such that $\|u - u_n\|_{H^m}$ tends to 0. A direct application of Morrey's theorem shows that, if $m - k/2 > 0$, then, for any $j \geq 0$, $H_0^{j+m}(\Omega, \mathbb{R})$ is embedded in $C_0^j(\overline{\Omega}, \mathbb{R})$.

7.3. The duality paradox. The Riesz representation theorem allows to identify a Hilbert space H and its dual H^* . However, when H is embedded in another Hilbert space H_0 , every continuous linear form on H_0 is also continuous on H , and H_0^* is embedded in H^* (proof as exercise). We therefore have the sequence of embeddings

$$H \rightarrow H_0 \simeq H_0^* \rightarrow H^*$$

but this sequence loops since $H^* \simeq H$. This indicates that H_0 is also embedded in H . This is indeed a strange result: for example, let $H = H^1(\Omega, \mathbb{R})$, and $H_0 = L^2(\Omega, \mathbb{R})$: the embedding of H in H_0 is clear from their definition, but the converse does not seem natural, since there are more constraints in belonging to H than to H_0 . To understand this reversed embedding, we must think in terms of linear forms.

If $u \in L^2(\Omega, \mathbb{R})$, we may consider the linear form φ_u defined by

$$(\varphi_u, v) = \langle u, v \rangle_{L^2} = \int_{\Omega} u(x)v(x)dx$$

When $v \in H^1(\Omega, \mathbb{R})$, we have

$$(\varphi_u, v) \leq \|u\|_{L^2} \|v\|_{L^2} \leq \|u\|_{L^2} \|v\|_{H^1}$$

so that φ_u is continuous when seen as a linear form on $H^1(\Omega, \mathbb{R})$. The Riesz representation theorem implies that there exists a $\tilde{u} \in H^1(\Omega, \mathbb{R})$ such that, for all $v \in H^1(\Omega, \mathbb{R})$, $\langle u, v \rangle_{L^2} = \langle \tilde{u}, v \rangle_{H^1}$: the relation $u \mapsto \tilde{u}$ provides the embedding of $L^2(\Omega, \mathbb{R})$ into $H^1(\Omega, \mathbb{R})$. Let us be more specific and take $\Omega =]0, 1[$. The relation states that, for any $v \in H^1(\Omega, \mathbb{R})$,

$$\int_0^1 \tilde{u}'(t)v'(t)dt + \int_0^1 \tilde{u}(t)v(t)dt = \int_0^1 \tilde{u}(t)v(t)dt$$

Let us make the simplifying assumption that \tilde{u} has two derivatives, in order to integrate by part the first integral and obtain

$$(5) \quad u'(1)v(1) - u'(0)v(0) - \int_0^1 \tilde{u}''(t)v(t)dt + \int_0^1 \tilde{u}(t)v(t)dt = \int_0^1 \tilde{u}(t)v(t)dt$$

Such an identity can be true for every v in $H^1(\Omega, \mathbb{R})$ if and only if $u'(0) = u'(1) = 0$ and $-\tilde{u}'' + \tilde{u} = u$: \tilde{u} is thus a solution of a second order differential equation, with first order boundary conditions, and the embedding of $L^2(]0, 1[, \mathbb{R})$ into $H^1(]0, 1[, \mathbb{R})$ just shows that a unique solution exists, at least in the generalized sense of equation (5).

As seen in these examples, even if, from an abstract point of view, we have an identification between the two Hilbert spaces, the associated embeddings are of very different nature, the first one corresponding to a set inclusion (it is canonical), the

second to the solution of a differential equation in one dimension, and in fact to a partial differential equation in the general case. Another striking application of the Riesz representation theorem is presented in the next section, which will be our first incursion in the theory of deformations.

8. Spline interpolation

8.1. The scalar case. Let $\Omega \subset \mathbb{R}^k$. Consider a Hilbert space V included in $L^2(\Omega, \mathbb{R})$. We assume that elements of V are smooth enough, and require the inclusion and the canonical embedding of V in $C^0(\Omega, \mathbb{R}^k)$. For example, it suffices (from Morrey's theorem) that $V \subset H^m(\Omega, \mathbb{R}^k)$ for $m > k/2$. This implies that there exists a constant C such that, for all $v \in V$

$$\|v\|_\infty \leq C \|v\|_V .$$

We will make another assumption on V : we assume that a relation of the kind $\sum_{i=1}^N \alpha_i v(x_i) = 0$ cannot be true for every $v \in V$ unless $\alpha_1 = \dots = \alpha_N = 0$, where (x_1, \dots, x_N) is an arbitrary family of distinct points in Ω . The next section will investigate some practical methods to define such a set V .

Each x in Ω specifies a linear form δ_x defined by $(\delta_x, v) = v(x)$ for $x \in V$. We have

$$|(\delta_x, v)| \leq \|v\|_\infty \leq C \|v\|_V$$

so that δ_x is continuous on V . But this implies, by Riesz's theorem, that there exists an element K_x in V such that, for every $v \in V$, one has

$$(6) \quad v(x) = \langle K_x, v \rangle_V$$

Since K_x belongs to V , it is a continuous function $y \mapsto K_x(y)$. This also defines a function, denoted $K : \Omega \times \Omega \rightarrow \mathbb{R}$ by $K(y, x) = K_x(y)$.

This function K has several interesting properties. First, applying equation (6) to $v = K_y$ yields

$$K(x, y) = K_y(x) = \langle K_x, K_y \rangle_V$$

Since the last term is symmetric, we have $K(x, y) = K(y, x)$, and because of the obtained identity, K is called the *self-reproducing kernel* of V .

A second property is the fact that K is positive definite, in the sense that, for any family $x_1, \dots, x_N \in V$ and any sequence $\alpha_1, \dots, \alpha_N$ in \mathbb{R} , the double sum

$$\sum_{i,j=1}^N \alpha_i \alpha_j K(x_i, x_j)$$

is always positive, and vanishes if and only if all α_i equals 0. Indeed, by the self-reproducing property, this sum may be written $\left\| \sum_{i=1}^N \alpha_i K_{x_i} \right\|_V^2$ and this is positive.

If it vanishes, then $\sum_{i=1}^N \alpha_i K_{x_i} = 0$, which implies, by equation (6) that, for every $v \in V$, one has $\sum_{i=1}^N \alpha_i v(x_i) = 0$, and our original assumption on V implies that $\alpha_1 = \dots = \alpha_N = 0$.

This kernel will help us to solve the following interpolation problem:

(S) fix a family of distinct points x_1, \dots, x_N in Ω . We want to determine a function $v \in V$ of minimal norm, subject to the constraints $v(x_i) = \lambda_i$, for some prescribed values $\lambda_1, \dots, \lambda_N \in \mathbb{R}$.

To solve this problem, define V_0 to be the set of v for which the constraints vanish

$$V_0 = \{v \in V : v(x_i) = 0, i = 1, \dots, N\}$$

Using the kernel K , we may write

$$V_0 = \{v \in V : \langle K_{x_i}, v \rangle = 0, i = 1, \dots, N\}$$

so that

$$V_0 = \text{vect} \{K_{x_1}, \dots, K_{x_N}\}^\perp$$

the orthogonal being taken for the V inner product. We have the first result

LEMMA 1. *If there exists a solution \hat{v} of problem \mathcal{S} , then $\hat{v} \in V_0^\perp = \text{vect} \{K_{x_1}, \dots, K_{x_N}\}$. Moreover, if $\hat{v} \in V_0^\perp$ is a solution of \mathcal{S} restricted to this set, then it is a solution of \mathcal{S} on V .*

PROOF. Let \hat{v} be this solution, and let v^* be its orthogonal projection on $\text{vect} \{K_{x_1}, \dots, K_{x_N}\}$. From the properties of orthogonal projections, we have $\hat{v} - v^* \in V_0$, which implies, by the definition of V_0 that $\hat{v}(x_i) = v^*(x_i)$ for $i = 1, \dots, N$. But, since $\|v^*\|_V \leq \|\hat{v}\|_V$ (by the variational definition of the projection), $\|\hat{v}\|_V \leq \|v^*\|_V$, by assumption, both norms are equal, which is only possible, when $\hat{v} = v^*$. Therefore, $\hat{v} \in V_0^\perp$ and the proof of the first assertion is complete.

Now, if \hat{v} is a solution of \mathcal{S} in which V is replaced by V_0^\perp , then, if v is any function in V which satisfies the constraints, then $v - \hat{v} \in V_0$ and $\|v\|_V^2 = \|\hat{v}\|_V^2 + \|v - \hat{v}\|_V^2$ which shows that \hat{v} is a solution of the initial problem. \square

This lemma allows us to restrict the search for a solution of \mathcal{S} to the set of linear combination of K_{x_1}, \dots, K_{x_N} which places us in a very comfortable finite dimensional situation. We look for \hat{v} under the form $\hat{v} = \sum_{i=1}^N \alpha_i K_{x_i}$, which may also be written

$$\hat{v}(x) = \sum_{i=1}^N \alpha_i K(x, x_i)$$

and we introduce the $N \times N$ matrix S such that $s_{ij} = K(x_i, x_j)$. The whole problem may now be reformulated in function of the vector $\alpha = {}^t(\alpha_1, \dots, \alpha_N)$ (a column vector) and of the matrix S . Indeed, by the self reproducing property of K , we have

$$(7) \quad \|\hat{v}\|_V^2 = \sum_{i=1}^N \alpha_i \alpha_j K(x_i, x_j) = {}^t \alpha S \alpha$$

and each constraint may be written $\lambda_i = v(x_i) = \sum_{j=1}^N \alpha_j K(x_i, x_j)$, so that, letting $\lambda = {}^t(\lambda_1, \dots, \lambda_N)$, the whole system of constraints may be expressed as $S\alpha = \lambda$.

But our hypotheses imply that S is invertible: indeed, if $S\alpha = 0$, then ${}^t \alpha S \alpha = 0$ which, by equation (7) and the positive definiteness of K , is only possible when $\alpha = 0$ (we assume that the x_i are distincts). Therefore, there is only one \hat{v} in V_0^\perp which satisfies the constraints, and it corresponds to $\alpha = S^{-1}\lambda$. These results are summarized in the next theorem.

THEOREM 7. *Problem \mathcal{S} has a unique solution in V , given by*

$$\hat{v}(x) = \sum_{i=1}^N K(x, x_i) \alpha_i$$

with

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \vdots & K(x_N, x_N) \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_N \end{pmatrix}$$

Another important variant of the same problem comes when the hard constraints $v(x_i) = \lambda_i$ are replaced by soft constraints, under the form of a penalty function added to the minimized norm. This may be expressed as the minimization of a function of the form

$$E(v) = \|v\|_V^2 + C \sum_{i=1}^N \varphi(|v(x_i) - y_i|)$$

for some increasing, convex function on $[0, +\infty[$ and $C > 0$. Since the second term of E does not depend on the projection of v on V_0 , lemma 1 remains valid, reducing again the problem to v of the kind

$$v(x) = \sum_{i=1}^N K(x, x_i) \alpha_i$$

for which

$$E(v) = \sum_{i,j=1}^N \alpha_i \alpha_j K(x_i, x_j) + C \sum_{i=1}^N \varphi \left(\left| \sum_{j=1}^N K(x_i, x_j) \alpha_j - \lambda_i \right| \right)$$

Assume, to simplify, that φ is differentiable and $\varphi'(0) = 0$. We have, letting $\psi(x) = \text{sign}(x)\varphi'(x)$,

$$\frac{\partial E}{\partial \alpha_j} = 2 \sum_{i=1}^N \alpha_i K(x_i, x_j) + C \sum_{i=1}^N K(x_i, x_j) \psi \left(\left| \sum_{l=1}^N K(x_i, x_l) \alpha_l - \lambda_i \right| \right)$$

Assuming, still, that the x_i are distincts, we can apply S^{-1} to the system $\frac{\partial E}{\partial \alpha_j} = 0, j = 1, \dots, N$, which characterizes the minimum, yielding

$$(8) \quad 2\alpha_i + C\psi \left(\left| \sum_{j=1}^N K(x_i, x_j) \alpha_j - \lambda_i \right| \right) = 0$$

This suggest an algorithm (which is a certain kind of gradient descent) to minimize E , that we can describe as follows:

ALGORITHM 1 (General Spline smoothing). step 0 *Start with an initial guess*
 α^0 for α
 step t.1 *Compute, for $i = 1, \dots, N$,*

$$\alpha_i^{t+1} = (1 - \gamma)\alpha_i^t - \frac{\gamma C}{2} \psi \left(\left| \sum_{j=1}^N K(x_i, x_j) \alpha_j - \lambda_i \right| \right)$$

where γ is a fixed, small enough, real number.

step t.2 *Use a convergence test: if positive, stop, otherwise increment t and restart step t.*

This algorithm will converge if γ is small enough. However, the particular case of $\varphi(x) = x^2$ is much simpler to solve, since in this case $\psi(x) = 2x$ and equation (8) becomes

$$2\alpha_i + 2C \left(\sum_{j=1}^N K(x_i, x_j) \alpha_j - \lambda_i \right) = 0$$

The solution of this equation is $\alpha = (S + I/C)^{-1} \lambda$, yielding a result very similar to theorem 7:

THEOREM 8. *The minimum over V of*

$$\|v\|_V^2 + C \sum_{i=1}^N |v(x_i) - \lambda_i|^2$$

is attained at

$$\hat{v}(x) = \sum_{i=1}^N K(x, x_i) \alpha_i$$

with

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = (S + I/C)^{-1} \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_N \end{pmatrix}$$

and

$$S = \begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \dots & K(x_N, x_N) \end{pmatrix}$$

8.2. The vector case. In the previous section, elements of V were functions from Ω to \mathbb{R} . When working with deformations, which our goal here, functions of interest describe displacements of points in Ω and therefore must be vector valued. This leads us to address the problem of spline approximation for vector fields in Ω , which, as will be seen, is handled very similarly to the scalar case.

So, in this section, V is a Hilbert space, canonically embedded in $L^2(\Omega, \mathbb{R}^k)$, and in $C^0(\Omega, \mathbb{R}^k)$. Fixing $x \in \Omega$, the evaluation function $v \mapsto v(x)$ is a continuous linear map from V to \mathbb{R}^k , which implies that, for any $a \in \mathbb{R}^k$, the function $v \mapsto {}^t a v(x)$ is a continuous linear functional on V . This implies that there exists a unique element, denoted K_x^a in V such that, for any $v \in V$

$$(9) \quad \langle K_x^a, v \rangle_V = {}^t a v(x)$$

The map $a \mapsto K_x^a$ is linear from \mathbb{R}^k to V (this is because $aq \mapsto {}^t a v(x)$ is linear and because of the uniqueness of the Riesz representation), which implies that, for $y \in \Omega$, the map $a \mapsto K_x^a(y)$ is linear from \mathbb{R}^k to \mathbb{R}^k . This implies that there exists a matrix, that we will denote $K(y, x)$, such that, for $a \in \mathbb{R}^k$, $x, y \in \Omega$, $K_x^a(y) = K(y, x)a$.

Thus, the kernel K , in the case of vector fields is matrix valued. The self reproducing property, in this case, write

$$\langle K_x^a, K_y^b \rangle_V = {}^t a K_y^b(x) = {}^t a K(x, y)b$$

Since the first term is symmetrical, we obtain the fact that, for all $a, b \in \mathbb{R}^k$, ${}^t a K(x, y)b = {}^t b K(y, x)a$ which implies that $K(y, x) = {}^t K(x, y)$.

Concerning the positivity of K , we make an assumption similar to the scalar case:

ASSUMPTION 1. *If $x_1, \dots, x_N \in \Omega$ and $\alpha_1, \dots, \alpha_N \in \mathbb{R}^k$ are such that, for all $v \in V$, $\alpha_1 v(x_1) + \dots + \alpha_N v(x_N) = 0$, then $\alpha_1 = \dots = \alpha_N = 0$.*

Under this assumption, it is easy to prove (exercise) that, for all $\alpha_1, \dots, \alpha_N \in \mathbb{R}^k$,

$$\sum_{i,j=1}^N {}^t \alpha_i K(x_i, x_j) \alpha_j \geq 0$$

with equality if and only if all α_i vanish.

The interpolation problem in the vector case writes:

(\mathcal{S}_k) Given x_1, \dots, x_N in Ω , $\lambda_1, \dots, \lambda_N$ in \mathbb{R}^k , find v in V , with minimum norm such that $v(x_i) = \lambda_i$. As before, we let

$$V_0 = \{v \in V : v(x_i) = 0, i = 1, \dots, N\}$$

Then, lemma 1 remains valid:

LEMMA 2. *If there exists a solution \hat{v} of problem \mathcal{S}_k , then $\hat{v} \in V_0^\perp$. Moreover, if $\hat{v} \in V_0^\perp$ is a solution of \mathcal{S} restricted to this set, then it is a solution of \mathcal{S} on V .*

We add to this the characterization of V_0^\perp , which slightly less straightforward than in the scalar case:

LEMMA 3.

$$V_0^\perp = \left\{ v = \sum_{i=1}^N K_{x_i}^{\alpha_i}, \alpha_1, \dots, \alpha_N \in \mathbb{R}^k \right\}$$

Thus, a vector field v belongs to V_0^\perp if and only if there exists $\alpha_1, \dots, \alpha_N$ in \mathbb{R}^k such that, for all $x \in \Omega$,

$$v(x) = \sum_{i=1}^N K(x, x_i) \alpha_i$$

This expression is formally similar to the scalar case, with the difference that $K(x, x_i)$ is a matrix and α_i are vectors.

PROOF. It is clear that $w \in V_0$ is equivalent to the fact that, for any $\alpha_1, \dots, \alpha_N$, one has

$$\sum_{i=1}^N {}^t \alpha_i v(x_i) = 0$$

so that $w \in V_0$ is equivalent to $\langle v, w \rangle_V = 0$ for all v of the kind $v = \sum_{i=1}^N K_{x_i}^{\alpha_i}$. Thus

$$V_0 = \left\{ v = \sum_{i=1}^N K_{x_i}^{\alpha_i}, \alpha_1, \dots, \alpha_N \in \mathbb{R}^k \right\}^\perp$$

and since $\left\{ v = \sum_{i=1}^N K_{x_i}^{\alpha_i}, \alpha_1, \dots, \alpha_N \in \mathbb{R}^k \right\}$ is finite dimensional, hence closed, one has

$$V_0^\perp = \left\{ v = \sum_{i=1}^N K_{x_i}^{\alpha_i}, \alpha_1, \dots, \alpha_N \in \mathbb{R}^k \right\}$$

□

When $v = \sum_{j=1}^N K_{x_j}^{\alpha_j} \in V_0^\perp$, the constraint $v(x_i) = \lambda_i$ yields

$$\sum_{j=1}^N K(x_i, x_j) \alpha_j = \lambda_i$$

Since we also have, in this case,

$$\|v\|_V^2 = \sum_{i,j=1}^N {}^t \alpha_i K(x_i, x_j) \alpha_j$$

the whole problem can be rewritten quite concisely under a matricial form, introducing the notation

$$(10) \quad S = S(x_1, \dots, x_N) = \begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \dots & K(x_N, x_N) \end{pmatrix}$$

which is now a block matrix of size $Nk \times Nk$,

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix}, \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix}$$

each α_i, λ_i being considered as k dimensional column vectors, and the whole set of constraints now writes $S\alpha = \lambda$ and $\|v\|_V^2 = {}^t \alpha S \alpha$. Thus, replacing scalars by blocks, the problem has exactly the same structure as in the previous case, and we can repeat the results we have obtained:

THEOREM 9 (interpolating splines). *Problem (S_k) has a unique solution in V , given by*

$$\hat{v}(x) = \sum_{i=1}^N K(x, x_i) \alpha_i$$

with

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} K(x_1, x_1) & \dots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \dots & K(x_N, x_N) \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_N \end{pmatrix}$$

THEOREM 10 (Smoothing splines). *The minimum over V of*

$$\|v\|_V^2 + C \sum_{i=1}^N |v(x_i) - \lambda_i|^2$$

is attained at

$$\hat{v}(x) = \sum_{i=1}^N K(x, x_i) \alpha_i$$

with

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = (S + I/C)^{-1} \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_N \end{pmatrix}$$

and $S = S(x_1, \dots, x_N)$ given at equation (10).

Finally, algorithm 1 remains the same, replacing scalars by vectors. (Exercise : use the algorithm with robust penalties).

9. Building V and its kernel

9.1. Simplifying assumptions. The previous section showed that the problem of spline interpolation could reduce to a solvable, finite dimensional problem. Of course, this remains essentially abstract unless it comes with an explicit, numerically plausible, expression for the kernel K , since all the algorithms depends on it. Before providing examples of such kernels, corresponding to specific Hilbert spaces V , we review a series of natural assumptions which will significantly simplify the issue.

We assume in this section that $\Omega = \mathbb{R}^k$, and study our first assumption:

9.1.1. *Euclidean frame independence.* Changing a frame in Euclidean coordinates means replacing the variable $x \in \mathbb{R}^k$ by $Rx + T$, R being a rotation and T a vector in \mathbb{R}^k (translation). When assessing the smoothness of a function defined over \mathbb{R}^k , it is natural to require independence with respect to such deformations. This requirement is stated below:

ASSUMPTION 2 (Frame independence). *Let R be a rotation, T a vector in \mathbb{R}^k , and define the operation $A_{R,T} : V \rightarrow L^2(\Omega, \mathbb{R}^k)$ by $(A_{R,T}v)(x) = v(Rx + T)$.*

We assume that, for all R and T , $A_{R,T}$ is an isometry of V onto itself.

This assumptions states two facts: first, if $v \in V$, its transform $A_{R,T}v$ also belongs to V . Second, it states that, if v and w belong to V ,

$$\langle A_{R,T}v, A_{R,T}w \rangle_V = \langle v, w \rangle_V$$

This has important consequences on the kernel K . Indeed, take $A_{R,T}v = K_x^\alpha$ in the previous formula. It yields, noting that $A_{R,T}^{-1} = A_{{}^tR, -{}^tRT}$

$$\langle K_x^\alpha, A_{R,T}w \rangle_V = \langle A_{{}^tR, -{}^tRT}K_x^\alpha, w \rangle_V$$

But the left hand term is, by definition of K_x^α also equal to ${}^t\alpha(A_{R,T}w)(x) = {}^t\alpha w(Rx + T) = \langle K_{Rx+T}^\alpha, w \rangle_V$. Thus for all $w \in V$

$$\langle A_{{}^tR, -{}^tRT}K_x^\alpha, w \rangle_V = \langle K_{Rx+T}^\alpha, w \rangle_V$$

which means that, for all $y \in \Omega, x \in \Omega, \alpha \in \mathbb{R}^k$, $K({}^tR(y-T), x)\alpha = K(y, Rx+T)\alpha$, thus $K({}^tR(y-T), x) = K(y, Rx+T)$. Taking $T = y$ and $R = \text{Id}$ yields $K(0, x) = K(y, x+y)$: K is translation invariant, and letting $G(x) = K(0, x)$, we obtain the fact that $K(y, z) = G(z-y)$. Since $K(y, z) = K(Ry, Rz)$, we obtain the fact that $G(x) = G(Rx)$ for all $x \in \mathbb{R}^k$, which is only possible when G is a function $\Gamma(|x|)$. We thus have the result:

THEOREM 11. *If V satisfies the frame independence assumption, its kernel K is necessarily such that there exists a function $\Gamma : [0, +\infty[\rightarrow \mathcal{M}_k(\mathbb{R})$ with*

$$K(x, y) = \Gamma(|x - y|)$$

Note that Γ must be a symmetric matrix, since $K(x, y) = {}^tK(y, x)$.

9.1.2. *Isotropy.* The assumption we make now is that the smoothness of a vector field remains unchanged when it is transformed by a constant rotation. This is natural, if there is no reason to privilege one direction rather than the other. It is stated as

ASSUMPTION 3 (Isotropy). *V is stable by the operation $v \mapsto Rv$ (R being an arbitrary rotation) and this operation is an isometry.*

Note that we do not need $\Omega = \mathbb{R}^k$ for this assumption. This again has strong consequences on the kernel. Indeed it implies that

$$\langle K_x^\alpha, Rw \rangle_V = \langle {}^t R K_x^\alpha, w \rangle_V$$

But, since $\langle K_x^\alpha, Rw \rangle_V = {}^t \alpha R w(x) = \langle K_x^{t R \alpha}, w \rangle_V$, we obtain the fact that $K_x^{t R \alpha} = {}^t R K_x^\alpha$ which yields the fact that, for every $x, y \in \Omega$, ${}^t R K(y, x) = K(y, x) {}^t R$, and that $K(y, x)$ commutes with all rotations. It is a well-known fact that the only matrices which commute with all the rotations are the homotheties, which means that there exists a function $\chi : \Omega \times \Omega \rightarrow \mathbb{R}$ such that $K(y, x) = \chi(y, x) \text{Id}$. We state this in the:

THEOREM 12. *If V satisfies the isotropy conditions, its kernel K takes the form*

$$K(x, y) = \chi(x, y) \text{Id}$$

where χ is a positive definite kernel on \mathbb{R} .

Although isotropy cannot be considered as a universally valid condition, it is a natural one in a large variety of situations. For this reason, and for simplicity, we shall restrict the discussion which follows to scalar kernels.

9.1.3. *Radial kernels.* When isotropy is combined with frame independence, the kernel χ must be radial, ie:

$$\chi(x, y) = \gamma(|x - y|)$$

for $x, y \in \mathbb{R}^k$. Here γ is a function defined on $[0, +\infty[$, with values in $]0, +\infty[$. There is a nice characterization of functions γ which provide positive kernels on \mathbb{R}^k for any k . This is partially addressed in the next proposition.

PROPOSITION 5. *Assume that there exists a positive function f on $]0, +\infty[$ such that*

$$\gamma(t) = \int_0^{+\infty} e^{-t^2 u} f(u) du$$

Then, whatever the dimension k , the kernel $\chi(x, y) = \gamma(|x - y|)$ is positive definite.

The interesting fact is that this sufficient condition is almost necessary. The necessary and sufficient condition, which requires Lebesgue integration, is that there exists a positive measure μ on $[0, +\infty[$ such that

$$\gamma(t) = \int_0^{+\infty} e^{-t^2 u} d\mu(u).$$

An important application is when μ is a Dirac measure $\mu = \delta_{\sigma^{-2}}$ which provides $\gamma(t) = e^{-\frac{t^2}{\sigma^2}}$. The kernel

$$\chi(x, y) = e^{-\frac{|x-y|^2}{\sigma^2}}$$

is called the *Gaussian kernel* on \mathbb{R}^k and is probably the most commonly used for spline smoothing. But proposition 5 can be used to generate other kernels, for example, $f(u) = e^{-u}$ provides the kernel

$$\chi(x, y) = \frac{1}{1 + |x - y|^2}$$

9.1.4. *Translation invariant kernels.* The translation invariant kernels (not necessarily radial), of the kind $\chi(x, y) = \Gamma(x - y)$ can be characterized in a similar way.

PROPOSITION 6. *Assume that there exists a positive, even function F on \mathbb{R}^k such that*

$$\Gamma(x) = \int_{\mathbb{R}^k} e^{-i^t x u} F(u) du$$

Then the kernel $\chi(x, y) = \Gamma(x - y)$ is positive definite.

9.2. From kernels to Hilbert spaces.

9.2.1. *Mercer's theorem.* The interest of the above discussion is that it is possible to reconstruct the Hilbert space V used in spline representation from a positive definite kernel. One of the most efficient way to achieve this is by the use of Mercer's theorem

THEOREM 13. *We take $\Omega = \mathbb{R}^k$. Let $K : \Omega \times \Omega \rightarrow \mathbb{R}$ be a continuous, positive definite kernel, such that*

$$\int_{\Omega \times \Omega} K(x, y)^2 dx dy < \infty$$

Then, there exists an orthonormal sequence of functions in $L^2(\Omega, \mathbb{R})$, $\varphi_1, \varphi_2, \dots$ and a decreasing sequence (ρ_n) which tends to 0 when n tends to ∞ such that

$$K(x, y) = \sum_{n=1}^{\infty} \rho_n \varphi_n(x) \varphi_n(y)$$

Define the Hilbert space V by

$$V = \left\{ v \in L^2(\Omega, \mathbb{R}) : \sum_{n=1}^{\infty} \langle v, \varphi_n \rangle_{L^2}^2 < \infty \right\}$$

and define, for $v, w \in V$:

$$\langle v, w \rangle_V = \sum_{n=1}^{\infty} \rho_n^{-1} \langle v, \varphi_n \rangle_{L^2} \langle w, \varphi_n \rangle_{L^2}$$

Note that, for $v \in V$, there is a pointwise convergence of the series

$$v(x) = \sum_{n=1}^{\infty} \langle v, \varphi_n \rangle_{L^2} \varphi_n(x)$$

since

$$\left(\sum_{n=p}^m \langle v, \varphi_n \rangle_{L^2} \varphi_n(x) \right)^2 \leq \sum_{n=p}^m \rho_n^{-1} \langle v, \varphi_n \rangle_{L^2}^2 \sum_{n=p}^m \rho_n (\varphi_n(x))^2$$

and both terms in the upper-bound can be made arbitrarily small. Similarly,

$$v(x) - v(y) = \sum_{n=1}^{\infty} \langle v, \varphi_n \rangle_{L^2} (\varphi_n(x) - \varphi_n(y))$$

so that

$$\begin{aligned} (v(x) - v(y))^2 &\leq \sum_{n=1}^{\infty} \rho_n^{-1} \langle v, \varphi_n \rangle_{L^2}^2 \sum_{n=1}^{\infty} \rho_n (\varphi_n(x) - \varphi_n(y))^2 \\ &= \|v\|_V^2 (K(x, x) - 2K(x, y) + K(y, y)) \end{aligned}$$

so that v is continuous.

Then, letting $K_x = K(\cdot, x)$, one has

$$\langle \varphi_m, K_x \rangle_{L^2} = \sum_{n=1}^{\infty} \rho_n \varphi_n(x) \langle \varphi_m, \varphi_n \rangle_{L^2} = \rho_m \varphi_m(x)$$

so that

$$\sum_{n=1}^{\infty} \rho_n^{-1} \langle \varphi_n, K_x \rangle_{L^2}^2 = \sum_{n=1}^{\infty} \rho_n \varphi_n(x)^2 = K(x, x) < \infty$$

so that $K_x \in V$ and a similar computation shows that $\langle K_x, K_y \rangle_V = K(x, y)$ so that K is self reproducible.

Finally, if $v \in V$,

$$\langle v, K_x \rangle_V = \sum_{n=1}^{\infty} \rho_n^{-1} \langle v, \varphi_n \rangle_{L^2} \langle \varphi_n, K_x \rangle_{L^2} = \sum_{n=1}^{\infty} \langle v, \varphi_n \rangle_{L^2} \varphi_n(x) = v(x)$$

so that K_x corresponds to the Riesz representation of the evaluation functional on V . Thus, finding a continuous, positive definite, square integrable kernel is enough to make the whole theory work.

9.2.2. Convolution kernels. Mercer's integrability condition cannot be satisfied when $\Omega = \mathbb{R}^k$ and $\chi(x, y) = \gamma(x - y)$ is translation invariant. However, in this case, there is another way for building the set V which is valid in a wide number of situations.

Consider the case when $\gamma(u) = \int_{\mathbb{R}^k} e^{-i^t uv} F(v) dv$ for an integrable positive and even function F on \mathbb{R}^k (we shall assume that F is non-vanishing). We let $G(v) = \sqrt{F(v)}$, assume that G is still integrable and set

$$\tilde{\gamma}(u) = \int_{\mathbb{R}^k} e^{-i^t uv} G(v) dv$$

Letting $\tilde{\chi}(x, y) = \tilde{\gamma}(x - y)$, we define

$$V = \left\{ v \in L^2(\mathbb{R}^k, \mathbb{R}) : \exists u \in L^2(\mathbb{R}^k, \mathbb{R}), v = \int_{\mathbb{R}^k} \tilde{\chi}(x, y) u(y) dy \right\}$$

It can be shown (from elementary properties of the Fourier transform), that

- (1) If $\int_{\mathbb{R}^k} \tilde{\chi}(x, y) u(y) dy = 0$ for all x , then $u = 0$. This implies that the relation

$$u \mapsto \tilde{\chi}.u = \int_{\mathbb{R}^k} \tilde{\chi}(x, y) u(y) dy$$

is one-to-one and allows to define the inner product on V by

$$\langle v, v' \rangle_V = \langle u, u' \rangle_{L^2}, \text{ when } v = \tilde{\chi}.u, v' = \tilde{\chi}.u'$$

which provides V with a structure of Hilbert space.

(2) We have

$$\int_{\mathbb{R}^k} (\tilde{\chi} \cdot u)^2 dx \leq \|F\|_\infty \|u\|_{L^2}^2$$

so that $V \subset L^2(\Omega, \mathbb{R}^k)$ whenever F is bounded.

(3)

$$\chi(x, y) = \int_{\mathbb{R}^k} \tilde{\chi}(x, z) \tilde{\chi}(z, y) dy$$

which implies that $\chi_x : y \mapsto \chi(y, x)$ belongs to V .

(4) We have, if $v = \tilde{\chi} \cdot u$

$$\langle \chi_x, v \rangle_V = \langle \tilde{\chi}_x, u \rangle_{L^2} = v(x)$$

so that χ is the reproducing kernel on V .

9.3. Building V from operators. There is another, direct method, for defining Hilbert spaces V of smooth functions. In this framework, an inner product, defined from the action of an operator, is defined on a subset of $L^2(\Omega; \mathbb{R})$ (within which everything can be defined easily, ie. in a classical sense when speaking of derivatives), but which is not complete for the induced norm, then extended to a larger subspace of $L^2(\Omega, \mathbb{R})$, which will be a Hilbert space. This is called Friedrich's extension of an operator. Since it is not restricted to subspaces of $L^2(\Omega, \mathbb{R})$, we work, in the following with an arbitrary Hilbert space H .

To start, we need a subspace D , included in H and dense in this space, and an operator (ie. a linear functional), $L : D \rightarrow H$. Our typical application will be with $D = C_K^\infty(\Omega, \mathbb{R})$ (the set of C^∞ functions with compact support in Ω) and $H = L^2(\Omega, \mathbb{R})$. In such a case, L may be chosen as a differential operator of any degree, since derivatives of C^∞ function with compact support obviously belong to L^2 . However, L will be assumed to satisfy an additional monotonicity constraint, which is

ASSUMPTION 4. *The operator L is assumed to be symmetrical and strongly monotone on D , which means that there exists a constant $c > 0$ such that, for all $u \in D$,*

$$(11) \quad \langle u, Lu \rangle_H \geq c \langle u, u \rangle_H$$

and for all $u, v \in D$

$$(12) \quad \langle u, Lv \rangle_H = \langle Lu, v \rangle_H$$

An example of strongly monotonic operator on $C_K^\infty(\Omega, \mathbb{R})$ is given by $Lu = -\Delta u + \lambda u$ where Δ is the Laplacian: $\Delta u = \sum_{i=1}^k \frac{\partial^2 u}{\partial x_i^2}$. Indeed, in this case, and when u has compact support, an integration by part yields

$$\int_{\Omega} \Delta u(x) u(x) dx = \sum_{i=1}^k \int_{\Omega} \left(\frac{\partial u}{\partial x_i} \right)^2 dx.$$

The operator L induces an inner product on D , defined by

$$\langle u, v \rangle_L = \langle u, Lv \rangle_H$$

Assumption 4 ensures the symmetry of this product and its positive definiteness. But D is not complete for $\|\cdot\|_L$, and we need to enlarge it (and simultaneously extend L) to obtain a Hilbert space.

THEOREM 14 (Freidrich's extension). *There exists a subspace $V \subset H$ such that D is a dense subspace of V and L can be extended to an operator $\hat{L} : V \rightarrow V^*$ such that:*

- V is embedded in H
- If $u, v \in D$, $\langle u, v \rangle_V = \langle Lu, v \rangle_H = \langle \hat{L}u, v \rangle$

The fact that \hat{L} is an extension of L comes modulo the identification $H = H^*$. Indeed, we have $V \subset H = H^* \subset V^*$ (by the ‘‘duality paradox’’), so that, L , defined on $D \subset V$ can be seen as an operator with values in H^* .

DEFINITION 2. *The operator \hat{L} defined in theorem 14 is called the energetic extension of L . Its restriction to the space*

$$D_L = \left\{ u \in V : \hat{L}u \in H^* = H \right\}$$

is called the Freidrich's extension of L .

The Freidrich extension of L will still be denoted L in the following. (exercise: V_L by continuity of linear forms)

We will not prove this theorem here, but the interested reader may refer to [22]. The Freidrich extension has other interesting properties:

THEOREM 15. *$L : V_L \rightarrow H$ is bijective and self-adjoint ($\langle Lu, v \rangle_H = \langle u, Lv \rangle_H$ for all $u, v \in V_L$).*

Its inverse, $L^{-1} : H \rightarrow H$ is continuous and self-adjoint.

If the embedding $V \subset H$ is compact, then $L^{-1} : H \rightarrow H$ is compact.

In the following, we will mainly be interested in embeddings stronger than the L^2 embedding implied by the monotony assumption. It is important that such embeddings are conserved by the extension provided they are true in the initial space D . This is stated in the next proposition

PROPOSITION 7. *Let D, V and H be as in theorem 14, and B be a Banach space such that $D \subset B \subset H$ and B is canonically embedded in H (there exists $c_1 > 0$ such that $\|u\|_B \geq c_1 \|u\|_H$). Assume that there exists a constant c_2 such that, for all $u \in D$, $\sqrt{\langle Lu, u \rangle_H} \geq c_2 \|u\|_B$. Then $V \subset B$ and $\|u\|_V \geq c_2 \|u\|_B$ for all $u \in V$.*

In particular, if B is compactly embedded in H , then $L^{-1} : H \rightarrow H$ is compact.

PROOF. Let $u \in V$. Since D is dense in V , there exists a sequence $u_n \in D$ such that $\|u_n - u\|_V \rightarrow 0$. Thus u_n is a Cauchy sequence in v , and by our assumption, it is also a Cauchy sequence in B , so that there exists $u' \in B$ such that $\|u_n - u'\|_B$ tends to 0. But since V and B are both embedded in H , we have $\|u_n - u\|_H \rightarrow 0$ and $\|u_n - u'\|_H \rightarrow 0$ which implies that $u = u'$. Thus u belongs to B , and since $\|u_n\|_V$ and $\|u_n\|_B$ respectively converge to $\|u\|_V$ and $\|u\|_B$, passing to the limit in the inequality $\|u_n\|_V \geq c_2 \|u_n\|_B$ ends the proof of proposition 7. \square

We now investigate the relation between this theory and the previous analyses using kernels. From now on, we assume $D = C_K^\infty(\Omega)$ and $H = L^2(\Omega, \mathbb{R})$. The Hilbert space V will hold the same role in both cases, but we need the continuity of the evaluation of functions in V , which here leads to assuming that, there exists a constant such that, for all $u \in D$,

$$(13) \quad \langle Lu, u \rangle_{L^2} \geq c \|u\|_\infty^2$$

since, as we have seen, this implies that V will be continuously embedded in $C^0(\Omega, \mathbb{R})$. Thus, the kernel K is well defined on V , such that $\langle K_x, v \rangle_V = v(x) = (\delta_x, v)$. The linear form δ_x belongs to V^* , and by theorem 14, which implies that $\langle K_x, v \rangle_V = (\hat{L}K_x, v)$, we have $\hat{L}K_x = \delta_x$, or

$$K_x = \hat{L}^{-1}\delta_x$$

This exhibits a direct relationship between the self-reproducing kernel on V and the energetic extension of L .

There is another consequence, which stems from the fact that $C^0(\Omega, \mathbb{R})$ is compactly embedded in $L^2(\Omega, \mathbb{R})$, so that theorem 15 implies that L^{-1} is a compact, self-adjoint operator. Such operators have the important property to admit an orthonormal sequence of eigenvectors: more precisely, there exists a decreasing sequence, (ρ_n) of positive numbers, which is either finite or tends to 0, and an orthonormal sequence φ_n in $L^2(\Omega, \mathbb{R})$, such that, for $u \in L^2(\Omega, \mathbb{R})$

$$L^{-1}u = \sum_{n=1}^{\infty} \rho_n \langle u, \varphi_n \rangle_{L^2} \varphi_n$$

Which directly characterizes V_L as the set

$$V_L = \left\{ u \in L^2(\Omega, \mathbb{R}) : \sum_{n=1}^{\infty} \left(\frac{\langle u, \varphi_n \rangle_{L^2}}{\rho_n} \right)^2 < \infty \right\}$$

and for $u \in V_L$, we have

$$Lu = \sum_{n=1}^{\infty} \rho_n^{-1} \langle u, \varphi_n \rangle_{L^2} \varphi_n$$

so that, for $u, v \in V_L$

$$\langle u, v \rangle_V = \langle Lu, v \rangle_{L^2} = \sum_{n=1}^{\infty} \rho_n^{-1} \langle u, \varphi_n \rangle_{L^2} \langle v, \varphi_n \rangle_{L^2}$$

This indicates that V should be given by

$$V = \left\{ u \in L^2(\Omega, \mathbb{R}) : \sum_{n=1}^{\infty} \rho_n^{-1} \langle u, \varphi_n \rangle_{L^2}^2 < \infty \right\}$$

This is indeed the case, V_L is dense in this set: is $u \in V$, then $u_N = \sum_{n=1}^N \langle u, \varphi_n \rangle \varphi_n$ belongs to V_L and $\|u_N - u\|_V \rightarrow 0$. We summarize what we have just obtained in the

THEOREM 16. *Assume that $D = C_K^\infty(\Omega)$ and $H = L^2(\Omega, \mathbb{R})$, and $L : D \rightarrow H$ is symmetric and satisfies*

$$\langle Lu, u \rangle_{L^2} \geq c \|u\|_\infty^2$$

for some constant $c > 0$. Then the energetic space of L , V , is continuously embedded in $C^0(\Omega, \mathbb{R})$ and its self-reproducing kernel is $K_x = L^{-1}\delta_x$. Moreover, there exists an orthonormal basis, (φ_n) in $L^2(\Omega, \mathbb{R})$ and a decreasing sequence of positive numbers, (ρ_n) , which tends to 0 such that

$$V = \left\{ u \in L^2(\Omega, \mathbb{R}) : \sum_{n=1}^{\infty} \rho_n^{-1} \langle u, \varphi_n \rangle_{L^2}^2 < \infty \right\}$$

Moreover,

$$Lu = \sum_{n=1}^{\infty} \rho_n^{-1} \langle u, \varphi_n \rangle_{L^2} \varphi_n$$

whenever

$$\sum_{n=1}^{\infty} \left(\frac{\langle u, \varphi_n \rangle_{L^2}}{\rho_n} \right)^2 < \infty$$

10. Weak convergence in a Hilbert space

Let us start with the definition:

DEFINITION 3. *When V is a Banach space, a sequence (v_n) in V is said to converge to some $v \in V$ if and only if, for all continuous linear form $\alpha : V \rightarrow \mathbb{R}$, one has $\alpha(v_n) \rightarrow \alpha(v)$ when n tends to infinity.*

The following simple proposition will be useful later:

PROPOSITION 8. *Assume that V and W are Banach spaces and that W is (continuously) embedded in V . Then, if a sequence (w_n) in W weakly converges (in W) to some $w \in W$, then w_n also weakly converges to w in V .*

This just says that if $\alpha(w_n) \rightarrow \alpha(w)$ for all continuous linear functional on W , then the convergence holds for all continuous linear functional on V , which is in fact obvious because the restriction to W of any continuous linear functional on V is a continuous linear functional on W .

In the case of a Hilbert space, the Riesz representation theorem immediately provides the proposition:

PROPOSITION 9. *Let V be a Hilbert space. A sequence v_n in V weakly converges to an element $v \in V$ if and only if, for all $w \in V$,*

$$\lim_{n \rightarrow \infty} \langle w, v_n \rangle_V = \langle w, v \rangle_V$$

Moreover, in v_n weakly converges to v , then $\|v\| \leq \liminf \|v_n\|$.

The last statement comes from the inequality: $\langle v_n, v \rangle_V \leq \|v_n\|_V \|v\|_V$ which provides at the limit

$$\|v\|_V^2 \leq \|v\|_V \liminf_{n \rightarrow \infty} \|v_n\|_V$$

If $\|v\|_V = 0$, there is nothing to prove, and in the other case, one can divide both terms of the inequality by $\|v\|_V$ to obtain the result.

Finally, the following result is essential for us ([22])

THEOREM 17. *If V is a Hilbert space and (v_n) is a bounded sequence in V (there exists a constant C such that $\|v_n\| \leq C$ for all n), then one can extract a subsequence from v_n which weakly converges to some $v \in V$.*

Ordinary differential equations and groups of diffeomorphisms

1. Introduction

In this chapter, we review a few results for the theory of differential equations, which will provide a very efficient way to generate deformations.

It may be time to specify what is meant by deformations in this course. The definition is quite natural: we fix an open subset Ω in \mathbb{R}^k . A deformation is a function φ which assigns to each point $x \in \Omega$ a displaced position $y = \varphi(x) \in \Omega$. There are two undesired behaviors that we would like to forbid: they are

- The deformation should not create holes: every point $y \in \Omega$ should be the image of some point $x \in \Omega$, ie. φ should be onto.
- Folds are also prohibited: two distinct points x and x' in Ω should not be targeting to the same point $y \in \Omega$, ie. φ must be one to one.

Thus deformations must be bijections of Ω . In addition, we require minimal smoothness for φ . The next definition introduces standard vocabulary:

DEFINITION 4. *A homeomorphism of Ω is a continuous bijection $\varphi : \Omega \rightarrow \Omega$ such that its inverse, φ^{-1} is continuous.*

A diffeomorphism of Ω is a continuously differentiable homeomorphism $\varphi : \Omega \rightarrow \Omega$ such that φ^{-1} is continuously differentiable.

One can show in fact that a continuously differentiable homeomorphism is a diffeomorphism. From now on, most of the deformations we shall consider will be diffeomorphisms of some open set $\Omega \subset \mathbb{R}^k$.

If there is information (like grey-level, color, shape boundary) which is carried at position x , it will be moved according to the deformation. We shall say that deformations act on structures carried by Ω . Consider as an example an “image”, ie. a function $I : \Omega \rightarrow \mathbb{R}$ and a diffeomorphism φ of Ω . The deformation will create a new image I' on Ω by letting $I'(y)$ be the value of I at the position x which was targeting to y through the deformation, ie $I'(y) = I(\varphi^{-1}(y))$ or $I' = I \circ \varphi^{-1}$. This describes the action of diffeomorphisms on functions.

However, we shall not be directly concerned with the direct problem of computing the action of diffeomorphisms, but with the inverse problem of estimating the best diffeomorphisms from the output of its action. For example, the image matching problem consists in finding an algorithm which, given two functions I and I' on Ω , is able to recover a plausible diffeomorphism φ such that $I' = I \circ \varphi^{-1}$.

We therefore must face the problem of building diffeomorphisms. This is not an easy matter, because of the nonlinear character of the problem: linear combinations of diffeomorphisms are not necessarily diffeomorphisms.

There is a direct, but limited way of building diffeomorphisms, by small perturbations of the identity. It is provided by the next proposition. We assume that Ω is bounded.

PROPOSITION 10. *Let $u \in C^1(\Omega, \mathbb{R}^k)$, and assume that, for some $\delta > 0$, one has that $u(x) = 0$ for any $x \in \Omega$ such that there exists $y \notin \Omega$ with $|x - y| < \delta$. Then, for small enough ε , $\varphi : x \mapsto x + \varepsilon u(x)$ is a diffeomorphism of Ω .*

PROOF. Indeed, φ is obviously continuously differentiable, and takes values in Ω as soon as $\varepsilon < \delta \|u\|_\infty$.

Since u is C^1 , there exists a constant C such that $|u(x) - u(x')| \leq C|x - x'|$. If $\varphi(x) = \varphi(x')$ we have

$$|x - x'| = \varepsilon |u(x) - u(x')| \leq C\varepsilon |x - x'|$$

which implies $x = x'$ as soon as it is assumed that $\varepsilon < 1/C$, and φ is one to one in this case.

Showing that φ is onto is a little bit harder. The first remark to be made is that if $B(y, \delta)$, the ball centered at y with radius δ is not included in Ω , then, by assumption, $\varphi(y) = y$ so that $y \in \varphi(\Omega)$. Thus assume that $B(y, \delta) \subset \Omega$, and consider the function ψ , defined on $B(0, \delta)$ by $\psi(\eta) = -\varepsilon u(y + \eta)$. If $\varepsilon < \delta \|u\|_\infty$, we have $\psi(\eta) \in B(0, \delta)$, and we have the inequality

$$|\psi(\eta) - \psi(\eta')| \leq \varepsilon C |\eta - \eta'|$$

If $\varepsilon C < 1$, ψ is a contraction, and the fixed point theorem (stated below) implies that there exists $\eta \in B(0, \delta)$ such that $\psi(\eta) = \eta$. But in this case,

$$\varphi(y + \eta) = y + \eta + \varepsilon u(y + \eta) = y + \eta - \psi(\eta) = y$$

so that $y \in \varphi(\Omega)$ and φ is onto. \square

We here recall the standard fixed point theorem, which has been used in the previous proof and will be used later:

THEOREM 18 (Banach fixed point theorem). *Let B be a Banach space, $U \subset B$ and φ be a contraction of U , ie. a map $\varphi : U \rightarrow U$ such that there exists a constant $c \in [0, 1[$ such that, for all $x, y \in U$,*

$$\|\varphi(x) - \varphi(y)\|_B \leq c \|x - y\|_B.$$

Then, φ has a unique fixed point in U , ie. there exists a unique $x_0 \in U$ such that $\varphi(x_0) = x_0$.

We therefore know how to build small deformations. Of course, we cannot be satisfied with this, since they corresponds to a very limited class of diffeomorphisms. However, we can use them to generate large deformation, because diffeomorphisms can be combined through the composition rule. If φ and φ' are diffeomorphisms, then $\varphi \circ \varphi'$ is a diffeomorphism. In fact, diffeomorphisms of Ω form a group for the composition of functions.

Thus, let $\varepsilon_0 > 0$ and u_1, \dots, u_n, \dots be vector fields on Ω which are such that, for $\varepsilon < \varepsilon_0$, $\text{id} + \varepsilon u_i$ is a diffeomorphism of Ω . Consider $\varphi_n = (\text{id} + \varepsilon u_n) \circ \dots \circ (\text{id} + \varepsilon u_1)$. We have

$$\varphi_{n+1} = (\text{id} + \varepsilon u_{n+1}) \circ \varphi_n = \varphi_n + \varepsilon u_{n+1} \circ \varphi_n$$

which can also be written $(\varphi_{n+1} - \varphi_n)/\varepsilon = u_n \circ \varphi_n$. Fixing $x \in \Omega$ and letting $x_0 = x$, $x_n = \varphi_n(x)$, we have the relation $(x_{n+1} - x_n)/\varepsilon = u_n(x_n)$. This looks like a discretized version of a differential equation, of the form (introducing a continuous time variable t):

$$\frac{dx_t}{dt} = u_t(x_t)$$

This should motivate the rest of this chapter, which will be devoted to ordinary differential equations and how they generate flows of diffeomorphisms.

2. A class of ordinary differential equations

2.1. Definitions. We let $\Omega \subset \mathbb{R}^k$ be open and bounded. We have denoted $C_0^1(\Omega, \mathbb{R}^k)$ the Banach space of continuously differentiable vector fields v on Ω such that v and dv vanish on $\partial\Omega$. Elements $v \in C_0^1(\Omega, \mathbb{R}^k)$ can be considered as defined on \mathbb{R}^k by setting $v(x) = 0$ if $x \notin \Omega$.

We define the set $\mathcal{X}^1(T, \Omega)$ of integrable functions from $[0, T]$ to $C_0^1(\Omega, \mathbb{R}^k)$. An element of $\mathcal{X}^1(T, \Omega)$ is a time-dependent vector field, $(v_t, t \in [0, T])$ such that, for each t , $v_t \in C_0^1(\Omega, \mathbb{R}^k)$ and

$$\|v\|_{\mathcal{X}^1, T} := \int_0^T \|v_t\|_{1, \infty} dt < \infty.$$

For $v \in \mathcal{X}^1(T, \Omega)$, we consider the ordinary differential equation, which will be formally denoted $\frac{dy}{dt} = v_t(y)$. A function $t \mapsto y(t)$ is called a solution of the equation on $[0, \tau]$ ($\tau \leq T$) with initial condition x if

- $t \mapsto y_t$ is continuous on $[0, \tau]$, $y_0 = x$.
- for all $t \leq \tau$,

$$y_t = x + \int_0^t v_s(y_s) ds$$

2.2. Main results.

2.2.1. Existence and uniqueness.

THEOREM 19. *Let $v \in \mathcal{X}^1(T, \Omega)$. For all $x \in \Omega$ and $t \in [0, T]$, there exists a unique solution on $[0, T]$ of the ordinary differential equation $\frac{dy}{dt} = v_t(y)$ with initial condition $y_t = x$. This solution is such that $y_s \in \Omega$ for all $s \in [0, T]$.*

PROOF. The proof is standard, and we provide it for completeness.

We denote $\bar{\Omega}$ the closure of the bounded, open set Ω , so that $\bar{\Omega}$ is compact in \mathbb{R}^k . In the following, set let $v_t(z) = 0$ when $z \notin \Omega$. This does not change the value of $\|v_t\|_{1, \infty}$.

Fix $x \in \bar{\Omega}$, $t \in [0, T]$ and $\delta > 0$. Let $I = I(t, \delta)$ denote the interval $[0, T] \cap [t - \delta, t + \delta]$. If φ is a continuous function from I to \mathbb{R}^k such that $\varphi(t) = x$, we define the function $\Gamma(\varphi) : I \rightarrow \mathbb{R}^k$ by

$$\Gamma(\varphi)(s) = x + \int_t^s v_u(\varphi(u)) du$$

This provide a function which is also continuous and such that $\Gamma(\varphi)(t) = x$. The set of continuous functions from the compact interval I to \mathbb{R}^k , with the supremum norm is a Banach space, and we show that for δ small enough, Γ satisfies

$$\|\Gamma(\varphi) - \Gamma(\varphi')\|_\infty \leq \gamma \|\varphi - \varphi'\|_\infty$$

with $\gamma < 1$. The fixed point theorem implies that there is a unique function φ such that $\Gamma(\varphi) = \varphi$, and this is the definition of a solution of the ordinary differential equation on I .

Since $\Gamma(\varphi)(s) - \Gamma(\varphi')(s) = \int_t^s (v_u(\varphi(u)) - v_u(\varphi'(u)))du$ we have

$$\|\Gamma(\varphi) - \Gamma(\varphi')\|_\infty \leq \|\varphi - \varphi'\|_\infty \int_I \|v_u\|_{1,\infty} du$$

but $\int_I \|v_u\|_{1,\infty} du$ can be made arbitrarily small by reducing δ so that existence and uniqueness on I is proved. Now, we can make the additional remark that δ can be taken independent of t . This is because the function $\alpha : s \mapsto \int_0^s \|v_u\|_{1,\infty} du$ is continuous, hence uniformly continuous on the interval $[0, T]$, so that there exists a constant $\eta > 0$ such that $|s - s'| < \eta$ implies that $|\alpha(s) - \alpha(s')| < 1/2$, and it suffices to take $\delta < \eta/2$.

From this remark, we can conclude that a unique solution of the ordinary differential equation exists over all $[0, T]$, because it is now possible, starting from the interval $I(t, \delta)$ to extend the solution from both sides, by jumps of $\delta/2$ at least, until boundaries are reached.

We now prove that solutions such that $y_t \in \Omega$ for some t belong to Ω at all times. This is because if there exists s such that $y_s = x' \notin \Omega$, then the function $\tilde{y}_u = x'$ for all u is a solution of the equation, since $v_u(x') = 0$ for all u . Uniqueness implies $\tilde{y} = y$ which is impossible. \square

DEFINITION 5. Let $v \in \mathcal{X}^1(T, \Omega)$. We denote by $\varphi_{st}^v(x)$ the solution at time t of the equation $\frac{dy}{dt} = v_t(y)$ with initial condition $y_s = x$.

The function $(t, x) \mapsto \varphi_{st}^v(x)$ is called the flow associated to v starting at s . It is defined on $[0, T] \times \Omega$.

From the definition, we have the property

PROPOSITION 11. If $v \in \mathcal{X}^1(T, \Omega)$ and $s, r, t \in [0, T]$, then

$$\varphi_{st}^v = \varphi_{rt}^v \circ \varphi_{sr}^v$$

In particular, $\varphi_{st}^v \circ \varphi_{ts}^v = \text{id}$ and φ_{st}^v is invertible for all s et t .

PROOF. If $x \in \Omega$, $\varphi_{st}^v(x)$ is the value at time t of the unique solution of $dy/dt = v_t(y)$ which is equal to x at time s . It is equal to $x' = \varphi_{sr}^v(x)$ at time r , and thus also equal to $\varphi_{rt}^v(x')$ which is the statement of the proposition. \square

2.2.2. Properties of the flow. The next theorem shows that flows really are interesting objects when diffeomorphisms are needed.

THEOREM 20. Let $v \in \mathcal{X}^1(T, \Omega)$. The associated flow, φ_{st}^v is for all times a diffeomorphism of Ω .

The proof of this result depends on Gronwall's lemma, that we first state and prove.

THEOREM 21 (Gronwall's lemma). Consider two positive functions α_s, u_s , defined for $s \in I$ where I is an interval in \mathbb{R} . We assume that u is bounded, and that, for some function c , and for all $t \in I$,

$$(14) \quad u_t \leq c + \left| \int_0^t \alpha_s u_s ds \right|.$$

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

$$u_t \leq ce^{|\int_0^t \alpha_s ds|}$$

PROOF. To treat simultaneously the cases $t > 0$ and $t < 0$, we let $\varepsilon = 1$ in the first case and $\varepsilon = -1$ in the second cases. Inequality (14) now writes:

$$u_t \leq c + \varepsilon \int_0^t \alpha_s u_s ds$$

Iterating once this inequality yields

$$u_t \leq c + c\varepsilon \int_0^t \alpha_s ds + \varepsilon^2 \int_0^t \int_0^{s_1} \alpha_{s_1} \alpha_{s_2} u_{s_2} ds_2 ds_1.$$

and it may be checked by induction that, repeating this process,

$$\begin{aligned} u_t \leq c + c\varepsilon \int_0^t \alpha_s ds + \dots + \varepsilon^n \int_{0 \leq s_1 \leq \dots \leq s_n \leq t} \alpha_{s_1} \dots \alpha_{s_n} ds_1 \dots ds_n \\ + \varepsilon^{n+1} \int_{0 \leq s_1 \leq \dots \leq s_n \leq t} \alpha_{s_1} \dots \alpha_{s_{n+1}} u_{s_{n+1}} ds_1 \dots ds_n ds_{n+1}. \end{aligned}$$

Consider the integral

$$I_n = \int_{0 \leq s_1 \leq \dots \leq s_n \leq t} \alpha_{s_1} \dots \alpha_{s_n} ds_1 \dots ds_n$$

Let σ be a permutation of $\{1, \dots, n\}$: making the change of variable $s_i \rightarrow s_{\sigma_i}$ in I_n yields

$$I_n = \int_{0 \leq s_{\sigma_1} \leq \dots \leq s_{\sigma_n} \leq t} \alpha_{s_1} \dots \alpha_{s_n} ds_1 \dots ds_n$$

Obviously,

$$\sum_{\sigma} \mathbf{1}_{0 \leq s_{\sigma_1} \leq \dots \leq s_{\sigma_n} \leq t} = 1$$

whenever $s_i \neq s_j$ is $i \neq j$. The n -tuples s_1, \dots, s_n for which $s_i = s_j$ for some j for a set of dimension $n - 1$ which has no influence on the integral. Thus, summing over σ yields

$$n! I_n = \int_{[0, t]^n} \alpha_{s_1} \dots \alpha_{s_n} ds_1 \dots ds_n = \left(\int_0^t \alpha_s ds \right)^n$$

Therefore, using the fact that u_n is bounded, we have

$$u_t \leq c \sum_{k=0}^n \frac{\varepsilon^k}{k!} \left(\int_0^t \alpha_s ds \right)^k + \frac{\varepsilon^{n+1} \sup(u)}{(n+1)!} \left(\int_0^t \alpha_s ds \right)^{n+1}.$$

and passing to the limit yields the result. \square

We now pass to the proof of theorem 20, in which Gronwall's lemma will be used several times.

PROOF OF THEOREM 20. We first show that φ_{st}^v is a homeomorphism. Take $x, y \in \Omega$. We have

$$\begin{aligned} |\varphi_{st}^v(x) - \varphi_{st}^v(y)| &= \left| x - y + \int_s^t (v_s(\varphi_{sr}^v(x)) - v_s(\varphi_{sr}^v(y))) dr \right| \\ &\leq |x - y| + \int_s^t \|v_r\|_{1, \infty} |\varphi_{sr}^v(x) - \varphi_{sr}^v(y)| ds \end{aligned}$$

We apply Gronwall's lemma with $c = |x - y|$, $\alpha_r = \|v_r\|_{1,\infty}$ and $u_r = |\varphi_{0r}^v(x) - \varphi_{0r}^v(y)|$ which is bounded since φ_{sr}^v is continuous in r . This yields

$$(15) \quad |\varphi_{sr}^v(x) - \varphi_{sr}^v(y)| \leq |x - y| \exp\left(\int_s^t \|v_r\|_{1,\infty} dr\right)$$

which shows that φ_{st}^v is continuous on Ω , and even Lipschitz, with Lipschitz constant smaller than $\exp(\|v\|_{\mathcal{X}^1,t})$. Since $(\varphi_{st}^v)^{-1} = \varphi_{ts}^v$, it is also continuous so that φ_{st}^v is a homeomorphism of Ω .

We now prove that we indeed have a diffeomorphism. To see what we aim at, assume first that this is true and let us formally differentiate, at $\varepsilon = 0$ the equation

$$\frac{\partial \varphi_{st}^v}{\partial t}(x + \varepsilon \delta) = v_t(\varphi_{st}^v(x + \varepsilon \delta))$$

to obtain

$$\frac{\partial}{\partial t} d_x \varphi_{st}^v \cdot \delta = d_{\varphi_{st}^v(x)} v_t d_x \varphi_{st}^v \cdot \delta$$

This indicate we should introduce the linear differential equation

$$\frac{\partial W_t}{\partial t} = d_{\varphi_{st}^v(x)} v_t W_t$$

with initial conditions $W_s = \delta$. The same argument we have used for the original equation shows that this equation also admits a unique solution on $[0, T]$. The previous computation indicates that this solution should provide the differential $d_x \varphi_{st}^v$, and we now prove this result. Define

$$a_\varepsilon(t) = (\varphi_{st}^v(x + \varepsilon \delta) - \varphi_{st}^v(x)) / \varepsilon - W_t.$$

We need to show that $a_\varepsilon(t) \rightarrow 0$ when $\varepsilon \rightarrow 0$. For $\alpha > 0$, define

$$\mu_t(\alpha) = \max\{|d_x v_t - d_y v_t| : x, y \in \Omega, |x - y| \leq \alpha\}$$

The function $x \mapsto d_x v_t$ can be extended to a continuous function on the compact set $\bar{\Omega}$, and therefore is uniformly continuous which is equivalent to the fact that $\mu_t(\alpha) \rightarrow 0$ when $\alpha \rightarrow 0$. We can write

$$\begin{aligned} a_\varepsilon(t) &= \frac{1}{\varepsilon} \int_s^t (v_r(\varphi_{sr}^v(x + \varepsilon \delta)) - v_r(\varphi_{sr}^v(x))) dr - \int_s^t d_{\varphi_{sr}^v(x)} v_t W_t dt \\ &= \int_s^t d_{\varphi_{sr}^v(x)} v_r a_\varepsilon(r) dr \\ &\quad + \frac{1}{\varepsilon} \int_s^t (v_r(\varphi_{sr}^v(x + \varepsilon \delta)) - v_r(\varphi_{sr}^v(x))) \\ &\quad - \varepsilon d_{\varphi_{sr}^v(x)} v_r (\varphi_{sr}^v(x + \varepsilon \delta) - \varphi_{sr}^v(x)) dr \end{aligned}$$

We have, for all $x, y \in \Omega$:

$$|v_t(y) - v_t(x) - d_x v_t(y - x)| \leq \mu_t(|x - y|) |x - y|.$$

This inequality, combined with equation (15), yields

$$|a_\varepsilon(t)| \leq \int_s^t \|v_r\|_{1,\infty} |a_\varepsilon(r)| dr + C(v) |\delta| \int_0^T \mu_r(\varepsilon C(v) |\delta|) dr$$

for a constant $C(v)$ which only depends on v . To conclude the proof using Gronwall's lemma we need the fact that

$$\lim_{\alpha \rightarrow 0} \int_0^T \mu_r(\alpha) dr = 0.$$

This is a consequence of the fact that $\mu_r(\alpha) \rightarrow 0$ for each r when $\alpha \rightarrow 0$ and of the upper bound $\mu_r(\alpha) \leq 2 \|\mathbf{v}\|_{1,\infty}$ which allows to apply Lebesgue dominated convergence theorem. \square

We have incidentally shown the following important fact:

PROPOSITION 12. *Let $v \in \mathcal{X}^1(T, \Omega)$. Then for fixed $x \in \Omega$, $d_x \varphi_{st}^v$ is the solution of the linear differential equation*

$$(16) \quad \frac{dW_t}{dt} = d_{\varphi_{st}^v(x)} v_t W_t$$

with initial condition $W_s = \text{Id}_k$.

2.2.3. *Variation with respect to the vector field.* It will be quite important, in the following, to be also able to characterize the effects that a variation of the time-dependent vector v may have on the induced flow. For this purpose, we fix $v \in \mathcal{X}^1(T, \Omega)$, and $h \in \mathcal{X}^1(T, \Omega)$, and proceed to the computation of $\frac{d}{d\varepsilon} \varphi_{st}^{v+\varepsilon h}$ at $\varepsilon = 0$. The argument is similar to the previous section: first, make a guess by formal differentiation of the original ODE, then proceed to a rigorous estimation argument to show that the result which has been guessed is indeed correct. So, consider the equation

$$\frac{\partial \varphi_{st}^{v+\varepsilon h}}{\partial t} = v_t \circ \varphi_{st}^{v+\varepsilon h} + \varepsilon h_t \circ \varphi_{st}^{v+\varepsilon h}$$

and formally compute its derivative with respect to ε . This yields

$$\frac{\partial}{\partial t} \frac{\partial}{\partial \varepsilon} \varphi_{st}^{v+\varepsilon h} = (h_t) \circ \varphi_t^{v+\varepsilon h} + d_{\varphi_{st}^{v+\varepsilon h}}(v_t + \varepsilon h_t) \frac{d}{d\varepsilon} \varphi_{st}^{v+\varepsilon h}$$

and, letting $\varepsilon = 0$,

$$\frac{\partial}{\partial t} \frac{\partial}{\partial \varepsilon} \varphi_{st}^{v+\varepsilon h} \Big|_{\varepsilon=0} = (h_t) \circ \varphi_t^v + d_{\varphi_{st}^v} v_t \frac{d}{d\varepsilon} \varphi_{st}^{v+\varepsilon h} \Big|_{\varepsilon=0}$$

which naturally leads to introduce the solution of the differential equation

$$(17) \quad \frac{\partial}{\partial t} W_t = h_t \circ \varphi_{st}^v + d_{\varphi_{st}^v} v_t W_t$$

with initial condition $W_s = 0$. We now set

$$a_\varepsilon(t) = (\varphi_{st}^{v+\varepsilon h}(x) - \varphi_{st}^v(x)) / \varepsilon - W_t$$

and express it under the form

$$\begin{aligned} a_\varepsilon(t) &= \int_s^t d_{\varphi_{su}^v} v_u a_\varepsilon(u) du + \int_s^t (h_u(\varphi_{su}^{v+\varepsilon h}(x)) - h_u(\varphi_{su}^v(x))) du \\ &\quad + \frac{1}{\varepsilon} \int_s^t (v_u(\varphi_{su}^{v+\varepsilon h}(x)) - v_u(\varphi_{su}^v(x)) - \varepsilon d_{\varphi_{su}^v(x)} v_u (\varphi_{su}^{v+\varepsilon h}(x) - \varphi_{su}^v(x))) du \end{aligned}$$

The proof can proceed exactly as in theorem 20, provided it has been shown that $|\varphi_{su}^{v+\varepsilon h}(x) - \varphi_{su}^v(x)| = O(\varepsilon)$ which is again a direct consequence of Gronwall's lemma and of the inequality

$$\frac{1}{\varepsilon} |\varphi_{st}^{v+\varepsilon h}(x) - \varphi_{st}^v(x)| \leq \int_s^t \|v_u\|_{\mathcal{X}^1, T} \left(\frac{1}{\varepsilon} |\varphi_{su}^{v+\varepsilon h}(x) - \varphi_{su}^v(x)| \right) du + \int_s^t \|h_u\|_{\infty} du$$

Equation (17) is the same as equation (16), with the additional term $h_t \circ \varphi_{st}^v$. This implies that the solution of (17) may be expressed in function of the solution of (16) by variation of the constant, ie. it may be expressed under the form

$$W_t = d_x \varphi_{st}^v A_t$$

with $A_s = 0$ and A_t may be identified by letting

$$h_t \circ \varphi_{st}^v + d_{\varphi_{st}^v} v_t W_t = \frac{dW_t}{dt} = d_x \varphi_{st}^v \frac{dA_t}{dt} + d_{\varphi_{st}^v} v_t W_t$$

so that

$$\frac{dA_t}{dt} = (d_x \varphi_{st}^v)^{-1} h_t \circ \varphi_{st}^v(x) = d_{\varphi_{st}^v(x)} \varphi_{ts}^v h_t \circ \varphi_{st}^v(x)$$

this implies that

$$A_t = \int_s^t d_{\varphi_{su}^v(x)} \varphi_{us}^v h_u \circ \varphi_{su}^v(x)$$

and

$$W_t = \int_s^t d_x \varphi_{st}^v d_{\varphi_{su}^v(x)} \varphi_{us}^v h_u \circ \varphi_{su}^v(x)$$

which, by the chain rule can be written

$$W_t = \int_s^t d_{\varphi_{su}^v(x)} \varphi_{ut}^v h_u \circ \varphi_{su}^v(x)$$

We summarize this discussion by the theorem

THEOREM 22. *Let $v, h \in \mathcal{X}^1(T, \Omega)$. Then, for $x \in \Omega$*

$$(18) \quad \frac{d}{d\varepsilon} \varphi_{st}^{v+\varepsilon h}(x)|_{\varepsilon=0} = \int_s^t d_{\varphi_{su}^v(x)} \varphi_{ut}^v h_u \circ \varphi_{su}^v(x) du$$

A trivial consequence of this theorem is the fact that φ_{st}^v depends continuously on v . But one can be more specific: the inequalities

$$\begin{aligned} \left| \varphi_{st}^v(x) - \varphi_{st}^{v'}(x) \right| &\leq \left| \int_s^t \left(v_u(\varphi_{su}^v(x)) - v'_u(\varphi_{su}^{v'}(x)) \right) du \right| \\ &\leq \left| \int_s^t v_u(\varphi_{su}^v(x)) - v'_u(\varphi_{su}^v(x)) du \right| + \int_s^t \left| v'_u(\varphi_{su}^v(x)) - v'_u(\varphi_{su}^{v'}(x)) \right| du \\ &\leq \left| \int_s^t v_u(\varphi_{su}^v(x)) - v'_u(\varphi_{su}^v(x)) du \right| + \int_s^t \|v'_u\|_{1, \infty} \left| \varphi_{su}^v(x) - \varphi_{su}^{v'}(x) \right| du \end{aligned}$$

We shall apply to this inequality a more general version of theorem 21 that we state without proof (see [9])

THEOREM 23. *Consider 3 continuous and positive functions c_s, α_s, u_s defined on $[0, T]$ such that*

$$(19) \quad u_t \leq c_t + \int_0^t \alpha_s u_s ds.$$

then

$$u_t \leq c_t + \int_0^t c_s \alpha_s e^{\int_s^t \alpha_r dr}$$

This theorem can be applied in our case by letting $u_\tau = \left| \varphi_{s+\tau}^v(x) - \varphi_{ss+\tau}^{v'}(x) \right|$, $\alpha_\tau = \|v'_{s+\tau}\|_{1,\infty}$ and

$$c_\tau = \left| \int_s^{s+\tau} v_u(\varphi_{su}^v(x)) - v'_u(\varphi_{su}^v(x)) du \right|$$

yielding the inequality

$$(20) \quad \left| \varphi_{st}^v(x) - \varphi_{st}^{v'}(x) \right| \leq c_{t-s} + \int_s^t c_{u-s} \|v'_u\|_{1,\infty} \exp\left(\int_u^t \|v'_{u'}\|_{1,\infty} du'\right) du$$

Consider the function $v' \mapsto \int_s^t v'_u \circ \varphi_{su}^v du$. It is a linear functional on $\mathcal{X}^1(T, \Omega)$ which is obviously continuous, since

$$\left| \int_s^t v'_u \circ \varphi_{su}^v du \right| \leq \int_s^t \|v'_u\|_\infty du \leq \|v'\|_{\mathcal{X}^1, T}$$

Thus, consider a sequence $v^n \in \mathcal{X}^1(T, \Omega)$ which is bounded in $\mathcal{X}^1(T, \Omega)$ and weakly converges to v . If we let c_τ^n be equal to c_τ with v^n instead of v' , weak continuity implies that $c_\tau^n \rightarrow 0$ when n tends to infinity. We have $c_\tau^n \leq \|v^n\|_{\mathcal{X}^1, T} + \|v\|_{\mathcal{X}^1, T}$, so that it is uniformly bounded and the dominated convergence theorem can be applied to equation (20) to obtain the fact that, for all $s, t \in [0, T]$, for all $x \in \Omega$ $\varphi_{st}^{v_n}(x) \rightarrow \varphi_{st}^v(x)$. But the convergence is in fact uniform in x , thanks to the following lemma

LEMMA 4. *For every $\varepsilon > 0$ there exists a finite set x_1, \dots, x_N in Ω such that, for all n ,*

$$\max_{x \in \Omega} |\varphi_{st}^{v_n}(x) - \varphi_{st}^v(x)| \leq \varepsilon + \max_{i=1, \dots, K} |\varphi_{st}^{v_n}(x_i) - \varphi_{st}^v(x_i)|$$

Applying this lemma and letting n tend to infinity implies that $\limsup \|\varphi_{st}^{v_n} - \varphi_{st}^v\|_\infty \leq \varepsilon$. Since this is true for all ε , we have

$$\lim_{n \rightarrow \infty} \|\varphi_{st}^{v_n} - \varphi_{st}^v\|_\infty = 0.$$

Finally, to prove lemma 4, we use the fact that, since v^n is bounded, and since we have shown that the liphitz constant of $\varphi_{st}^{v_n}$ could be expressed in function of $\|v_n\|_{\mathcal{X}^1, T}$ we have the fact that there exists a constant C such that, for all $n > 0$, and $x, y \in \Omega$

$$|\varphi_{st}^{v_n}(x) - \varphi_{st}^{v_n}(y)| \leq C|x - y|$$

and the same inequality with v instead of v^n . Using the boundedness of Ω , we can find, for every $\varepsilon > 0$, a finite sequence of elements $x_1, \dots, x_N \in \Omega$ such that, for any $x \in \Omega$, there exists i such that $|x - x_i| \leq \varepsilon/(2C)$. This implies that

$$\begin{aligned} |\varphi_{st}^{v_n}(x) - \varphi_{st}^v(x)| &\leq |\varphi_{st}^{v_n}(x) - \varphi_{st}^{v_n}(x_i)| + |\varphi_{st}^{v_n}(x_i) - \varphi_{st}^v(x_i)| + |\varphi_{st}^v(x_i) - \varphi_{st}^v(x)| \\ &\leq 2C|x - x_i| + |\varphi_{st}^{v_n}(x_i) - \varphi_{st}^v(x_i)| \\ &\leq \varepsilon + |\varphi_{st}^{v_n}(x_i) - \varphi_{st}^v(x_i)| \end{aligned}$$

Thus, we have just proved the theorem

THEOREM 24. *If $v \in \mathcal{X}^1(T, \Omega)$ and v_n is a bounded sequence in $\mathcal{X}^1(T, \Omega)$ which weakly converges to v , then, for all $s, t \in [0, T]$*

$$\lim_{n \rightarrow \infty} \|\varphi_{st}^{v_n} - \varphi_{st}^v\|_\infty = 0$$

3. Groups of diffeomorphisms

3.1. Admissible Banach spaces.

DEFINITION 6. *A Banach space $V \subset C_0^1(\Omega, \mathbb{R}^k)$ is admissible if it is (canonically) embedded in $C_0^1(\Omega, \mathbb{R}^k)$, ie. there exists a constant C such that, for all $v \in V$,*

$$(21) \quad \|v\|_V \geq C \|v\|_{1, \infty}$$

If V is admissible, we denote by $\mathcal{X}_V^1(\Omega)$ the set of time-dependent vector fields, $(v_t, t \in [0, 1])$ such that, for each t , $v_t \in V$ and

$$\|v\|_{\mathcal{X}_V^1} := \int_0^1 \|v_t\|_{1, \infty} dt < \infty.$$

If the interval $[0, 1]$ is replaced by $[0, T]$, we will use the notation $\mathcal{X}_V^1(T, \Omega)$ and $\|v\|_{\mathcal{X}_V^1, T}$.

3.2. Induced group of diffeomorphisms.

DEFINITION 7. *If $V \subset C_0^1(\Omega, \mathbb{R}^k)$ is admissible, we denote*

$$G_V = \{\varphi_{01}^v, v \in \mathcal{X}_V^1(\Omega)\}$$

the set of diffeomorphisms provided by the flow associated to elements $v \in \mathcal{X}_V^1(\Omega)$ after time 1.

THEOREM 25. *G_V is a group for the composition of functions.*

PROOF. The identity function belongs to G : it corresponds, for example, to φ_{01}^v when $v = 0$. If $\psi = \varphi_{01}^v$ and $\psi' = \varphi_{01}^{v'}$, with $v, v' \in \mathcal{X}_V^1$, then $\psi' \circ \psi = \varphi_{01}^w$ with $w_t = v_{2t}$ for $t \in [0, 1/2]$ and $w_t = v'_{2t-1}$ for $t \in]1/2, 1]$ (details are left to the reader) and w belongs to $\mathcal{X}_V^1(\Omega)$. Similarly, if $\psi = \varphi_{01}^v$, then $\psi^{-1} = \varphi_{01}^w$ with $w_t = -v_{1-t}$. Indeed, we have

$$\varphi_{0, 1-t}^w(y) = y - \int_0^{1-t} v_{1-s} \circ \varphi_{0s}^w(y) ds = y + \int_1^t v_s \circ \varphi_{0, 1-s}^w ds$$

which implies (by the uniqueness theorem) that $\varphi_{0, 1-t}^w(y) = \varphi_{1t}^v(y)$ and in particular $\varphi_{0, 1-t}^w = \varphi_{10}^v$. This proves that G_V is a group. \square

Thus, by selecting a certain Banach space V , we can tailor the group of diffeomorphisms we will be working with. In particular, elements in G_V inherit the smoothness properties of V . We state the following theorems without proof.

THEOREM 26. *Assume that V is continuously embedded in $C_0^p(\Omega, \mathbb{R}^k)$, $p \geq 1$. Then elements in G_V are p times continuously differentiable. If $v \in \mathcal{X}_V^1(\Omega)$, $q \leq p$, $\delta_1, \dots, \delta_q \in \mathbb{R}^k$, the q th derivative of φ_{0t}^v applied to $\delta_1, \dots, \delta_q$, ie. $d_x^q \varphi_{0t}^v(\delta_1, \dots, \delta_q)$, which is a q -linear form in $\delta_1, \dots, \delta_q$, satisfies the equation*

$$(22) \quad \frac{d}{dt} d_x^q \varphi_{0t}^v(\delta_1, \dots, \delta_q) = d_x^q (v \circ \varphi_{0t}^v)(\delta_1, \dots, \delta_q)$$

with $d_x^q \varphi_{00}^v(\delta_1, \dots, \delta_q) = 0$.

Expanding the q -th derivative of $v \circ \varphi_{0t}^v$ in the right-hand-term of equation (22) (the general formula is rather cumbersome, but one can do it at least for small values of q) yields the fact that $d_x^q \varphi_{0t}^v(\delta_1, \dots, \delta_q)$ is the solution of a linear differential equation with right-hand-term and vanishing initial conditions.

3.3. A distance on G_V . Let V be an admissible Banach space. For ψ and ψ' in G_V , we let

$$(23) \quad d_V(\psi, \psi') = \inf_{v \in \mathcal{X}_V^1(\Omega)} \left\{ \|v\|_{\mathcal{X}_V^1}, \psi' = \psi \circ \varphi_{01}^v \right\}$$

We have the theorem:

THEOREM 27 (Trouvé). *The function d_V is a distance on G_V , and (G_V, d_V) is a complete metric space.*

This says that d_V is symmetrical, satisfies the triangle inequality $d_V(\psi, \psi') \leq d_V(\psi, \psi'') + d_V(\psi'', \psi')$ and $d_V(\psi, \psi') = 0$ if and only if $\psi = \psi'$.

PROOF. Note that the set over which the infimum is computed is not empty: is $\psi, \psi' \in G_V$, then $\psi' \circ \psi^{-1} \in G_V$ (since G_V is a group) and therefore can be written under the form φ_{01}^v for some $v \in \mathcal{X}_V^1(\Omega)$.

Let us start with the symmetry: fix $\varepsilon > 0$ and v such that $\|v\|_{\mathcal{X}_V^1} \leq d(\psi, \psi') + \varepsilon$ and $\psi' = \psi \circ \varphi_{01}^v$. This implies that $\psi = \psi' \circ \varphi_{10}^w$, but we know (from the proof of theorem 25) that $\varphi_{10}^w = \varphi_{01}^w$ with $w_t = -v_{1-t}$. Since $\|w\|_{\mathcal{X}_V^1} = \|v\|_{\mathcal{X}_V^1}$, we have, from the definition of d_V :

$$d_V(\psi', \psi) \leq \|w\|_{\mathcal{X}_V^1} \leq d_V(\psi, \psi') + \varepsilon$$

and since this is true for every ε , we have $d_V(\psi', \psi) \leq d_V(\psi, \psi')$. Interverting the roles of ψ and ψ' yields $d_V(\psi', \psi) = d_V(\psi, \psi')$. For the triangular inequality, let v and v' be such that $\|v\|_{\mathcal{X}_V^1} \leq d(\psi, \psi'') + \varepsilon$, $\|v'\|_{\mathcal{X}_V^1} \leq d(\psi'', \psi') + \varepsilon$, $\psi'' = \psi \circ \varphi_{01}^v$ and $\psi' = \psi'' \circ \varphi_{01}^{v'}$. We thus have $\psi' = \psi \circ \varphi_{01}^v \circ \varphi_{01}^{v'}$ and we know, still from the proof of theorem 25, that $\varphi_{01}^v \circ \varphi_{01}^{v'} = \varphi_{01}^w$ with $w_t = v_{2t}$ for $t \in [0, 1/2]$ and $w_t = v_{2t-1}$ for $t \in]1/2, 1]$. But, in this case, $\|w\|_{\mathcal{X}_V^1} = \|v\|_{\mathcal{X}_V^1} + \|v'\|_{\mathcal{X}_V^1}$ so that

$$d(\psi, \psi') \leq \|w\|_{\mathcal{X}_V^1} \leq d(\psi, \psi'') + d(\psi'', \psi') + 2\varepsilon$$

which implies the triangular inequality, since this is true for every $\varepsilon > 0$.

We obviously have $d(\psi, \psi) = 0$ since $\varphi_{01}^0 = \text{id}$. Assume that $d(\psi, \psi') = 0$. This implies that there exists a sequence v_n such that $\|v_n\|_{\mathcal{X}_V^1} \rightarrow 0$ and $\psi' \circ \psi^{-1} = \varphi_{01}^{v_n}$. The continuity of $v \mapsto \varphi_{st}^v$ implies that $\varphi_{01}^{v_n} \rightarrow \varphi_{01}^0 = \text{id}$ so that $\psi = \psi'$.

Let us check now that we indeed have a complete metric space. Let ψ^n be a Cauchy sequence for d_V , so that, for any $\varepsilon > 0$, there exists n_0 such that, for any $n \geq n_0$, $d_V(\varphi^n, \psi^{n_0}) \leq \varepsilon$. Taking recursively $\varepsilon = 2^{-n}$, it is possible to extract a subsequence ψ^{n_k} of ψ^n such that

$$\sum_{k=0}^{\infty} d_V(\psi^{n_k}, \psi^{n_{k+1}}) < \infty$$

Since a Cauchy sequence converges whenever one of its subsequences does, it is sufficient to show that ψ^{n_k} has a limit.

From the definition of d_V , there exists, for every $k \geq 0$ an element v^k in $\mathcal{X}_V^1(\Omega)$ such that $\psi^{n_{k+1}} = \varphi_{01}^{v^k} \circ \psi^{n_k}$ and

$$\|v^k\|_{\mathcal{X}_V^1} \leq d_V(\psi^{n_k}, \psi^{n_{k+1}}) + 2^{-k-1}$$

Let us define a time dependent vector field v by $v_t = 2v_{2t}^0$ for $t \in [0, 1/2[$, $v_t = 4v_{4t-2}^1$ for $t \in [1/2, 3/4[$, and so on: to define the general term, introduce dyadic time $t_0 = 0$ and $t_{k+1} = t_k + 2^{-k-1}$ and let

$$v_t = 2^{k+1} v_{2^{k+1}(t-t_k)}^k$$

for $t \in [t_k, t_{k+1}[$. Since t_k tends to 1 when $t \rightarrow \infty$, this defines v_t on $[0, 1[$, and we fix $v_1 = 0$. We have

$$\begin{aligned} \|v_t\|_{\mathcal{X}_V^1} &= \sum_{k=0}^{\infty} 2^{k+1} \int_{t_k}^{t_{k+1}} \|v_{2^{k+1}(t-t_k)}^k\|_V dt \\ &= \sum_{k=0}^{\infty} \int_0^1 \|v_t^k\|_V dt \\ &\leq 1 + \sum_{k=0}^{\infty} d_V(\psi^{n_k}, \psi^{n_{k+1}}) \end{aligned}$$

so that $v \in \mathcal{X}_V^1(\Omega)$. Now, consider the associated flow φ_{0t}^v : it is obtained by first integrating $2v_{2t}^0$ between $[0, 1/2[$, which yields $\varphi_{0,1/2}^v = \varphi_{01}^{v^0}$, and more generally

$$\varphi_{0t_{k+1}}^v = \varphi_{01}^{v^k} \circ \dots \circ \varphi_{01}^{v^0},$$

so that

$$\psi^{n_{k+1}} = \varphi_{0t_{k+1}}^v \circ \psi^0$$

Let $\psi^\infty = \varphi_{01}^v \circ \psi^{n_0}$. We have $\psi^\infty = \varphi_{t_k 1}^v \circ \psi^{n_k}$. Since $\varphi_{t_k 1}^v = \varphi_{01}^{w^k}$ with $w_t^k = v_{(t-t_k)/(1-t_k)}/(1-t_k)$, and

$$\|w\|_{\mathcal{X}_V^1} = \int_{t_k}^1 \|v_t\|_V dt$$

we obtain the fact that $d_V(\psi^{n_k}, \psi^\infty) \rightarrow 0$ which finishes the proof of theorem 27. \square

3.4. Properties of the distance. We first introduce the set of square integrable (in time) time dependent vector fields:

DEFINITION 8. *Let V be an admissible Banach space. We define $\mathcal{X}_V^2(\Omega)$ as the set of time dependent vector fields $v = (v_t, t \in [0, 1])$ such that, for each t , $v_t \in V$ and*

$$\int_0^1 \|v_t\|_V^2 dt < \infty$$

And we state without proof the important result:

PROPOSITION 13. *$\mathcal{X}_V^2(\Omega)$ is a Banach space with norm*

$$\|v\|_{\mathcal{X}_V^2} = \left(\int_0^1 \|v_t\|_V^2 dt \right)^{1/2}.$$

If V is itself a Hilbert space, then $\mathcal{X}_V^2(\Omega)$ is Hilbert too with

$$\langle v, w \rangle_{\mathcal{X}_V^2} = \int_0^1 \langle v_t, w_t \rangle_V dt.$$

Because $\left(\int_0^1 \|v_t\|_V dt \right)^2 \leq \int_0^1 \|v_t\|_V^2 dt$, we have $\mathcal{X}_V^2(\Omega) \subset \mathcal{X}_V^1(\Omega)$ and if $v \in \mathcal{X}_V^2(\Omega)$, $\|v\|_{\mathcal{X}_V^1} \leq \|v\|_{\mathcal{X}_V^2}$. The computation of d_V can be reduced to a minimization over \mathcal{X}_V^2 by

THEOREM 28. *If v is admissible and $\psi, \psi' \in G_V$, we have*

$$(24) \quad d_V(\psi, \psi') = \inf_{v \in \mathcal{X}_V^2(\Omega)} \left\{ \|v\|_{\mathcal{X}_V^2}, \psi' = \psi \circ \varphi_{01}^v \right\}$$

PROOF. Let

$$\delta_V(\psi, \psi') = \inf_{v \in \mathcal{X}_V^2(\Omega)} \left\{ \|v\|_{\mathcal{X}_V^2}, \psi' = \psi \circ \varphi_{01}^v \right\}$$

and d_V be given by (23). Since δ_V is the infimum over a larger set than d_V , and minimizes a quantity which is always smaller, we have $d_V(\psi, \psi') \leq \delta_V(\psi, \psi')$ and we now proceed to the reverse inequality. For this, consider $v \in \mathcal{X}_V^1(\Omega)$ such that $\psi' = \psi \circ \varphi_{01}^v$. It suffices to prove that, for any $\varepsilon > 0$ there exists a $w \in \mathcal{X}_V^2(\Omega)$ such that $\psi' = \psi \circ \varphi_{01}^w$ and $\|w\|_{\mathcal{X}_V^2} \leq \|v\|_{\mathcal{X}_V^1} + \varepsilon$. The important remark for this purpose is that, if α is a differentiable, increasing function from $[0, 1]$ onto $[0, 1]$ (which implies $\alpha_0 = 0$ and $\alpha_1 = 1$), then, letting $\dot{\alpha}_s = \frac{d\alpha_s}{ds}$,

$$\begin{aligned} \varphi_{0\alpha_t}^v(x) &= x + \int_0^{\alpha_t} v_u(\varphi_{0u}(x)) du \\ &= x + \int_0^t \dot{\alpha}_s v_{\alpha_s}(\varphi_{0\alpha_s}(x)) ds \end{aligned}$$

so that the flow generated by $w = \dot{\alpha}_s v_{\alpha_s}$ is $\varphi_{0\alpha_t}^v$ and therefore coincides with φ_{01}^v at $t = 1$. We have

$$\|w\|_{\mathcal{X}_V^1} = \int_0^1 \dot{\alpha}_s \|v_{\alpha_s}\|_V ds = \int_0^1 \|v_t\|_V dt = \|v\|_{\mathcal{X}_V^1}$$

so that this time change does not help for the minimization in (23). However, we have, denoting by β_t the inverse of α_t ,

$$\|w\|_{\mathcal{X}_V^2}^2 = \int_0^1 \dot{\alpha}_s^2 \|v_{\alpha_s}\|_V^2 ds = \int_0^1 \dot{\alpha}_{\beta_t} \|v_t\|_V^2 dt$$

so that this transformation can be used to reduce $\|v\|_{\mathcal{X}_V^2}$. If v_t never vanishes, we can choose $\dot{\alpha}_{\beta_t} = c/\|v_t\|_V$ or equivalently $\dot{\beta}_t = \|v_t\|_V/c$ with c chosen so that

$$\int_0^1 \dot{\beta}_t dt = 1$$

which yields $c = \|v\|_{\mathcal{X}_V^1}$, since this gives

$$\|w\|_{\mathcal{X}_V^2}^2 = c \int_0^1 \|v_t\|_V dt = \|v\|_{\mathcal{X}_V^1}^2$$

which is exactly the kind of transform we are looking for. In the general case, we let, for some $\eta > 0$, $\dot{\alpha}_{\beta_t} = c/(\eta + \|v_t\|_V)$ which yields $\dot{\beta}_t = (\eta + \|v_t\|_V)/c$ and $c = \eta + \|v\|_{\mathcal{X}_V^1}$. This gives

$$\|w\|_{\mathcal{X}_V^2}^2 = c \int_0^1 \frac{\|v_t\|_V^2}{\eta + \|v_t\|_V} dt \leq c \|v\|_{\mathcal{X}_V^1} = (\|v\|_{\mathcal{X}_V^1} + \eta) \|v\|_{\mathcal{X}_V^1}$$

By choosing η small enough, we can always arrange that $\|w\|_{\mathcal{X}_V^2} \leq \|v\|_{\mathcal{X}_V^1} + \varepsilon$ which is what we wanted to prove. \square

An important consequence of this result is the following fact:

COROLLARY 2. *If the infimum in (24) is attained at some $v \in \mathcal{X}_V^2(\Omega)$, then $\|v_t\|_V$ is constant.*

Indeed, let v achieve the minimum in (24): we have

$$d_V(\psi, \psi') = \|v\|_{\mathcal{X}_V^2} \geq \|v\|_{\mathcal{X}_V^1}$$

but $\|v\|_{\mathcal{X}_V^1} \geq d_V(\psi, \psi')$ by definition. Thus, we must have $\|v\|_{\mathcal{X}_V^2} = \|v\|_{\mathcal{X}_V^1}$, which corresponds to the equality case in the Schwartz inequality, which can only be achieved by constant functions (a rigorous statement would be *almost everywhere* constant functions).

Corollary 2 is usefully completed by the following theorem:

THEOREM 29. *If V is Hilbert and admissible, and $\psi, \psi' \in G_V$, there exists $v \in \mathcal{X}_V^2(\Omega)$ such that*

$$d_V(\psi, \psi') = \|v\|_{\mathcal{X}_V^2}$$

and $\psi' = \varphi_{01}^v \circ \psi$.

PROOF. By proposition 13, $\mathcal{X}^2(\Omega)$ is a Hilbert space. Let us fix a minimizing sequence for $d_V(\psi, \psi')$, ie. a sequence $v^n \in \mathcal{X}_V^2(\Omega)$ such that $\|v^n\|_{\mathcal{X}_V^2} \rightarrow d_V(\psi, \psi')$ and $\psi' = \varphi_{01}^{v^n} \circ \psi$. This implies that $(\|v^n\|_{\mathcal{X}_V^2})$ is bounded and by theorem 17, one can extract a subsequence of v^n (that we still denote by v^n) which weakly converges to some $v \in \mathcal{X}_V^2(\Omega)$, such that

$$\|v\|_{\mathcal{X}_V^2} \leq \liminf \|v^n\|_{\mathcal{X}_V^2} = d_V(\psi, \psi')$$

But theorem (24) implies that φ^{v^n} converges to φ^v so that $\psi' = \varphi_{01}^{v^n} \circ \psi$ remains true: this proves theorem 29. \square

Minimizing matching functionals

1. General principles

The methods we shall investigate in this chapter are using variational formulations to find optimal matchings between two different structures. They will all obey a very simple general form. Let Ω be an open subset of \mathbb{R}^d and G a group of diffeomorphisms on Ω . Consider a set \mathcal{I} of structures of interest, on which G is “acting”: for every I in \mathcal{I} and every $\varphi \in G$, the result of the action of φ on I is denoted $\varphi.I$ and is a new element of \mathcal{I} . Natural conditions require that $\text{id}.I = I$ and $\varphi.(\psi.I) = (\varphi \circ \psi).I$ (we shall see examples immediately). Then, given two elements I and I' in \mathcal{I} , the optimal matching φ between I and I' is computed by minimizing a functional $E_{I,I'}(\varphi)$ which measures how adequate φ is as a matching function between I and I' . Most of the time, E will be of the kind of the kind

$$(25) \quad E_{I,I'}(\varphi) = \delta(\varphi) + D(\varphi.I, I')$$

where δ is a function which penalizes unwanted behavior, typically ensuring that φ is smooth and not too far from the identity, and D measures the difference between $\varphi.I$ and I' .

Before reviewing examples and algorithms in this framework, we may list some properties that we would like to see satisfied. An obvious one should be that if $I = I'$, then the preferred φ should be identity. If E is given by (25), this will be true when δ is minimal at the identity and $D(I, I')$ minimal when $I = I'$. Another important property is consistency: can one find a minimizer for $E_{I,I'}$ in G (or maybe in some larger space that G on which $E_{I,I'}$ could still be defined) ? Obviously, the problem would otherwise be meaningless. Beside consistency is feasibility: can one find an algorithm which is able to solve the problem ? Is the algorithm stable enough ?

Symmetry may also be an important requirement: if φ is found to provide an optimal correspondence from I to I' , it should be natural that φ^{-1} has the same property from I' to I .

2. Matching terms and greedy descent procedures

2.1. Eulerian gradients. We consider energies of the form (25). For such energies, the matching term largely depends on the matching problem which is considered. Some of them are reviewed in this section. There are two important issues which will be addressed in each case. The first one is the continuity of the matching function. Assume that I and I' are given: what assumptions must be made on a converging sequence $\varphi_n \rightarrow \varphi$ to ensure that $D(\varphi_n.I, I') \rightarrow D(\varphi.I, I')$? Often, pointwise convergence will be enough ($\varphi_n(x) \rightarrow \varphi(x)$ for all x), but

sometimes, some additional requirements will be necessary (like assumptions on the derivatives of φ_n).

The second important issue concerns the differentiability of the matching term, under the following sense, called hereafter the Eulerian gradient. We first provide some necessary definitions. Let V be a Hilbert space included in $L^2(\Omega, \mathbb{R}^k)$ such that $G_V \subset G$. This implies that U is well defined over G_V so that, if $v \in \mathcal{X}_V^1(\Omega)$ and $\varphi \in G$, the function $t \mapsto U(\varphi_{0t}^v \circ \varphi)$ is also well-defined.

DEFINITION 9. *We say that U has a Eulerian gradient with respect to V if and only if, for all $\varphi \in G$, there exists an element denoted $\bar{\nabla}_\varphi^V U \in V$ such that*

$$(26) \quad \frac{d}{dt} U(\varphi_{0t}^v)|_{t=0} = \left\langle \bar{\nabla}_\varphi^V U, v \right\rangle_V$$

in which we identify v with the element $v_t \equiv v \in \mathcal{X}_V^1(\Omega)$.

The interest of this notion is to define the ‘‘gradient descent’’ process which generates a time dependent element of G by setting

$$(27) \quad \frac{d\varphi_t(x)}{dt} = -\bar{\nabla}_{\varphi_t}^V U \circ \varphi_t(x)$$

As long as $\int_0^t \left\| \bar{\nabla}_{\varphi_s}^V U \right\|_V ds$ is finite, this generates a time-dependent element of G_V .

Under some assumptions on the continuity of $\varphi \mapsto U(\varphi)$ and $\varphi \mapsto \bar{\nabla}_\varphi^V U$, it can be shown that

$$\frac{dU(\varphi_t)}{dt} = -\left\| \bar{\nabla}_{\varphi_t}^V U \right\|_V^2$$

so that $U(\varphi_t)$ decreases with time.

2.2. Point matching. Assume that objects are collections of N points $x_1, \dots, x_N \in \Omega$. Diffeomorphisms act on such objects by:

$$(28) \quad \varphi.(x_1, \dots, x_N) = (\varphi(x_1), \dots, \varphi(x_N))$$

Fix two objects $I = (x_1, \dots, x_N)$ and $I' = (x'_1, \dots, x'_N)$ and consider the matching functional

$$(29) \quad U(\varphi) = \sum_{i=1}^N |x'_i - \varphi(x_i)|^2$$

U is obviously continuous for pointwise convergence. Next, we study the existence of a Eulerian gradient with respect to a Hilbert space V . We assume here that V is embedded in $C^0(\Omega, \mathbb{R}^k)$: this implies that the evaluation functions $v \mapsto {}^t av(y)$ are continuous linear forms on V and that there exists a self-reproducing kernel K such that $K_y^a : x \mapsto K(x, y)a$ belongs to V and $\langle v, K_y^a \rangle = {}^t av(y)$ for all $v \in V$. Consider a diffeomorphism, φ , and a vector field $v \in V$, and compute the differential

$$\frac{d}{dt} U(\varphi_{0t}^v \circ \varphi)$$

at time $t = 0$. We have

$$\frac{d}{dt} U(\varphi_{0t}^v \circ \varphi) = 2 \sum_{i=1}^N {}^t \varphi_{0t}^v(\varphi(x_i)) - x'_i v \circ \varphi_{0t}^v(\varphi(x_i))$$

At $t = 0$, one has $\varphi_{00}^v(y) = y$ and

$${}^t(\varphi(x_i) - x'_i)v(\varphi(x_i)) = \left\langle K_{\varphi(x_i)}^{\varphi(x_i) - x'_i}, v \right\rangle_V$$

and we readily obtain the expression of the gradient

$$\bar{\nabla}_{\varphi}^V U = 2 \sum_{i=1}^N K_{\varphi(x_i)}^{\varphi(x_i) - x'_i}$$

The gradient descent algorithm writes

$$\frac{d\varphi_t(x)}{dt} = -2 \sum_{i=1}^N K_{\varphi_t(x_i)}^{\varphi_t(x_i) - x'_i}(\varphi_t(x)) = -2 \sum_{i=1}^N K(\varphi_t(x), \varphi_t(x_i))(\varphi_t(x_i) - x'_i)$$

Denote $y_t^i = \varphi_t(x_i)$. Applying the gradient descent algorithm at $x = x_j$ yields

$$\frac{dy_t^j}{dt} = -2 \sum_{i=1}^N K(y_t^j, y_t^i)(y_t^i - x'_i)$$

This is a differential system in y_t^1, \dots, y_t^N which must be solved with initial conditions $y_0^j = \varphi_0(x_j)$. Once this is done, the value of $\varphi_t(x)$ for a general x is the solution of the differential equation

$$\frac{dy}{dt} = -2 \sum_{i=1}^N K(y, y_t^i)(y_t^i - x'_i)$$

We now summarize this algorithm, assuming that $\varphi_0 = \text{id}$:

ALGORITHM 2 (greedy spline minimization). *Solve the differential system*

$$\frac{dy_t^j}{dt} = -2 \sum_{i=1}^N K(y_t^j, y_t^i)(y_t^i - x'_i)$$

with initial conditions $y_0^j = x_j$.

Then, let φ_t be the flow associated to the ordinary differential equation

$$\frac{dy}{dt} = -2 \sum_{i=1}^N K(y, y_t^i)(y_t^i - x'_i)$$

2.3. Image matching. Now, consider that objects are functions I defined on Ω with values in \mathbb{R} . Diffeomorphisms act on functions by:

$$(\varphi.I)(x) = I(\varphi^{-1}(x))$$

for $x \in \Omega$. Fixing two such functions I and I' , the simplest matching functional which can be considered is

$$(30) \quad U(\varphi) = \int_{\Omega} |I \circ \varphi^{-1}(x) - I'(x)|^2 dx$$

We assume that I has a derivative. Letting V be like above, we compute

$$(31) \quad \frac{d}{dt} U(\varphi_{0t}^v \circ \varphi) = 2 \int_{\Omega} (I \circ \varphi^{-1} \circ \varphi_{t0}^v(x) - I'(x))^t \nabla_{\varphi_{t0}^v(x)} (I \circ \varphi^{-1}) \frac{d}{dt} \varphi_{t0}^v dx$$

Now, differentiating $\varphi_{0t}^v \circ \varphi_{t0}^v = \text{id}$ with respect to time yields

$$v \circ \varphi_{0t}^v \circ \varphi_{t0}^v + d_{\varphi_{t0}^v} \varphi_{0t}^v \frac{d}{dt} \varphi_{t0}^v = 0$$

so that

$$\frac{d}{dt}\varphi_{t0}^v(x) = -(d_{\varphi_{t0}^v(x)}\varphi_{0t}^v)^{-1}v(x) = -d_x\varphi_{t0}^v v(x).$$

Taking $t = 0$ and introducing this into (31) yields

$$\frac{d}{dt}U(\varphi_{0t}^v \circ \varphi)|_{t=0} = -2 \int_{\Omega} (I \circ \varphi^{-1}(x) - I'(x))^t \nabla_x (I \circ \varphi^{-1})v(x) dx$$

Now,

$${}^t \nabla_x (I \circ \varphi^{-1})v(x) = \left\langle K_x^{\nabla_x (I \circ \varphi^{-1})}, v \right\rangle_V$$

so that

$$\overline{\nabla}_{\varphi}^V U = -2 \int_{\Omega} (I \circ \varphi^{-1}(x) - I'(x)) K_x^{\nabla_x (I \circ \varphi^{-1})} dx$$

This provides the greedy gradient descent algorithm, the original version being given in [7] (see also [21])

ALGORITHM 3 (Greedy image matching). *Start with $\varphi_0 = \text{id}$ and solve the evolution equation*

$$\frac{d}{dt}\varphi_t(y) = 2 \int_{\Omega} (I_t(x) - I'(x)) K(\varphi_t(y), x) \nabla_x I_t dx$$

with $I_t = I \circ \varphi_t^{-1}$.

Other, more complex, comparison criteria for images are based on image histograms, or grey-level distributions. For this, define, for $\lambda \in \mathbb{R}$ the level line of an image I at λ by

$$I_{\lambda} = \{x \in \Omega, I(x) = \lambda\}$$

In order to compare images which have been obtained under different lighting conditions, or by different imaging devices, in which case the grey level values are not reliable anymore, it is necessary to use matching criteria which are invariant to one-to-one transformations of the grey-levels. A group of methods which have been used in this context are based on joint local histograms of the grey levels. Such local histograms are computed from a pair of images, I, I' , and are functions of the kind $H_x(\lambda, \lambda')$ which count the frequency of simultaneous occurrence of grey-levels λ in I and λ' in I' at the same location in a small window around x . One computationally feasible way to define it is

$$H_x(\lambda, \lambda') = \int_{\Omega} f(|I(y) - \lambda|) f(|I'(y) - \lambda'|) g(x, y) dy$$

where f is positive and $\int_{\mathbb{R}} f(t) dt = 1$, and vanishes when t is far from 0, g is such that for all x , $\int_{\Omega} g(x, y) dy = 1$ and $g(x, y)$ vanishes when y is far from x .

For each x , H_x is a bidimensional probability function, and there exists several ways for measuring the degree of dependence between each of its components. The simplest one, which is probably sufficient for most applications, is the correlation ratio, given by (reintroducing I and I' in the notation)

$$C_x(I, I') = 1 - \frac{\int_{\mathbb{R}^2} \lambda \lambda' H_x(\lambda, \lambda') d\lambda d\lambda'}{\sqrt{\int_{\mathbb{R}^2} \lambda^2 H_x(\lambda, \lambda') d\lambda d\lambda' \int_{\mathbb{R}^2} (\lambda')^2 H_x(\lambda, \lambda') d\lambda d\lambda'}}$$

It is then possible to define the matching function by

$$U(\varphi) = \int_{\Omega} C_x(I \circ \varphi^{-1}, I') dx$$

Here again, the computation of Eulerian gradients is possible. Since it is quite lengthy, we do not detail it here, but the important fact to remember is that they lead to an algorithm similar to 3 for what is often called multimodal image matching.

2.4. Limit behavior of greedy algorithms. Greedy algorithms, as we presented them here, have a series of very nice features. First, they operate by integrating forward a differential equation, which leads to fast and reasonably simple computer programs. Second, because of the action of the self-reproducing kernel, they provide smooth solutions at all times, which, combined to the fact that they travel along descent lines for some objective function (U), generally guarantees numerical stability, at least in finite time. Maybe the most important thing is that their output is the flow of some differential equation, hence a diffeomorphism at all times. As we will see later, other algorithms designed to provide diffeomorphisms may be much more computationally intensive.

However, the infinite time behaviour of such algorithms is less obvious. One thing we have remarked is that $U(\varphi_t)$ decreases with time, so that, if U is bounded from below (which is a minimal assumption regarding the feasibility of the minimisation problem) it will have a limit when t tends to infinity. One can therefore expect that its derivative, which is the norm (in V) of the left-hand term of the differential equation, vanishes for infinite t , so that variations on φ_t become infinitesimally small. This certainly results in a freezing of any numerical integration procedure when times goes by. But this does not guarantee convergence, since slowly varying sequences may still diverge at infinity.

In fact (note that this discussion is placed on a highly heuristic level) limit behavior should strongly depend on the context in which the procedure is used. For point matching, because the problem is finite dimensional, some reasonable set of conditions can certainly be found in order that greedy gradient descent converges to some solution which achieves exact matching (thus a null value for the cost function) which would belong to G_V . However, there is an infinite number of solutions to the exact matching problem within G_V , and it is hard to guess which is the one which is selected by this algorithm. These issues will be explored, experimentally, in assignment 2.

The image matching case is more complex, because it is not even guaranteed that an exact matching solution exists within G_V . The asymptotic behavior of the algorithm really is an open problem. However, this algorithm (with an adequate stopping rule) provides generally good and smooth matching results, as illustrated in [7, 21, 5].

3. Regularized matching

3.1. Small deformations. A standard way to obtain a (hopefully) unique and smooth solution of a matching problem is to add a penalty term in the matching functional, denoted $\delta(\varphi)$ in (25). A large variety of methods has been designed, in approximation theory, statistics or signal processing for solving ill-posed problems. A typical penalty function would take the form

$$\delta(\varphi) = \|\varphi - \text{id}\|_H^2$$

for some Hilbert (or Banach) space of functions. Some more complex functions of $\varphi - \text{id}$ may also be devised, related to energies of non-linear elasticity (see, among others, [2, 3, 1, 11, 10, 19, 14]). Such methods may be called “small deformation” methods because they work on the deviation of $u = \varphi - \text{id}$, and controlling the size or smoothness of u alone is not enough to guarantee that φ is a diffeomorphism unless u is small, as we have seen in section 1 of chapter 2. For example, there is, in general, no way of proving the existence of a solution of the minimization problem within some group of diffeomorphisms G , unless some restrictive assumptions are made on the objects to be matched.

Since we want to focus this course on diffeomorphic matching we shall not detail much of these methods here. However, it is interesting to compute the Eulerian gradients of such functionals, and we shall do this in a simple situation, since the computation rapidly become very heavy.

So, consider the function $\delta(\varphi) = \int_{\Omega} \|d_x \varphi\|^2 dx$ where the matrix norm is

$$\|A\|^2 = \text{trace}({}^t A A) = \sum_{i,j} a_{ij}^2$$

(Hilbert-Schmidt norm). Let V be some Hilbert space of vector fields on Ω , take $v \in V$, and compute

$$\frac{d}{dt} \delta(\varphi_{0t}^v \circ \varphi)|_{t=0}$$

From proposition 12, we have

$$\frac{d}{dt} d_x \varphi_{0t}^v = d_{\varphi_{0t}^v(x)} v d_x \varphi_{0t}^v$$

Taking this equality at time $t = 0$ yields (after applying the chain rule)

$$\frac{d}{dt} \delta(\varphi_{0t}^v \circ \varphi)|_{t=0} = 2 \int_{\Omega} \text{trace}({}^t d_x \varphi d_{\varphi(x)} v d_x \varphi) dx.$$

Let $v^{(\alpha)}$ be the α th component of the \mathbb{R}^k -valued vector field v , $\alpha = 1, \dots, k$, and let e_{α} be the α -th element of the canonical basis of \mathbb{R}^k so that

$$(32) \quad v^{(\alpha)}(x) = {}^t e_{\alpha} v(x) = \langle K_x e_{\alpha}, v \rangle_V = \left\langle K_x^{(\alpha)}, v \right\rangle_V$$

where K is the self-reproducing kernel of V , and $K_x^{(\alpha)}$ is the α -th column of K . Recalling that $K_x(\cdot) = K(\cdot, x)$ is a $k \times k$ matrix, we define $\partial_{\beta} K(\cdot, x)$ as the partial derivative of $K(\cdot, x)$ with respect to x_{β} . A formal differentiation of (32) yields we get

$$\frac{\partial v^{(\alpha)}}{\partial x_{\beta}}(x) = \left\langle \partial_{\beta} K^{(\alpha)}(\cdot, x), v \right\rangle_V$$

(this computation can be made rigorous under sufficient smoothness conditions on K). Now,

$$\begin{aligned} \text{trace}({}^t d_x \varphi d_{\varphi(x)} v d_x \varphi) &= \sum_{i,\alpha,\beta=1}^k \frac{\partial \varphi^{(\alpha)}}{\partial x_i} \frac{\partial \varphi^{(\beta)}}{\partial x_i} \frac{\partial v^{(\alpha)}}{\partial x_{\beta}}(\varphi(x)) \\ &= \sum_{i,\alpha,\beta=1}^k \frac{\partial \varphi^{(\alpha)}}{\partial x_i} \frac{\partial \varphi^{(\beta)}}{\partial x_i} \left\langle \partial_{\beta} K^{(\alpha)}(\cdot, \varphi(x)), v \right\rangle_V \end{aligned}$$

By linearity, this yields the Hilbert gradient of δ on V :

$$(33) \quad \bar{\nabla}_\varphi^V \delta = 2 \int_\Omega \sum_{i,\alpha,\beta=1}^k \frac{\partial \varphi^{(\alpha)}}{\partial x_i} \frac{\partial \varphi^{(\beta)}}{\partial x_i} \partial_\beta K^{(\alpha)}(\cdot, \varphi(x)) dx$$

This yields a new greedy image matching algorithm, which includes a regularization term (a similar algorithm may easily be written for point matching)

ALGORITHM 4. *The following procedure is a Eulerian gradient descent, on V , for the energy*

$$U(\varphi) = \int_\Omega \|d_x \varphi\|^2 dx + \frac{1}{\sigma^2} \int_\Omega |I \circ \varphi^{-1}(x) - I'(x)| dx :$$

Start with an initial $\varphi_0 = \text{id}$ and integrate the differential equation

$$(34) \quad \begin{aligned} \frac{d}{dt} \varphi_t(y) &= -2 \int_\Omega \sum_{i,\alpha,\beta=1}^k \frac{\partial \varphi_t^{(\alpha)}}{\partial x_i} \frac{\partial \varphi_t^{(\beta)}}{\partial x_i} \partial_\beta K^{(\alpha)}(\varphi_t(y), \varphi_t(x)) dx \\ &\quad + \frac{2}{\sigma^2} \int_\Omega (I_t(x) - I'(x)) K(\varphi_t(y), x) \nabla_x I_t dx \end{aligned}$$

with $I_t = I \circ \varphi_t^{-1}$.

This algorithm, which, like the previous greedy procedures we detailed, has the tremendous interest of providing a smooth flow of diffeomorphisms to minimize the matching functional, suffers from the same limitations of its predecessors concerning its limit behavior, which are essentially due to the fact that the variational problem itself is not well-posed: minimizers may not exist, and when they exist they do not necessarily correspond to diffeomorphisms. The methods we present now, because they directly work with flows of diffeomorphisms, will not have this disadvantage, although this will be at the cost of a significant increase of the computational burden.

3.2. Optimizing over flows.

3.2.1. *Existence theorem.* Instead of using a functional norm on the difference between φ and the identity mapping, we here consider, as a regularizing term, the distance d_V which has been defined in section 3.3 of chapter 2. More precisely, we set

$$\delta(\varphi) = d_V(\text{id}, \varphi)^2$$

and henceforth restrict the matching to diffeomorphisms belonging to G_V . In this context, we have the important existence theorem:

THEOREM 30. *Let V be a Hilbert space which is embedded in $C_0^1(\Omega, \mathbb{R}^k)$. Assume that the functional $U : G_V \mapsto \mathbb{R}$ is bounded from below and continuous with respect to pointwise convergence. Then there always exists a minimizer of*

$$(35) \quad E(\varphi) = d_V(\text{id}, \varphi)^2 + F(\varphi)$$

over G_V .

PROOF. E has an infimum u over G_V , since it is bounded from below. The problem is equivalent to finding the infimum, over $\mathcal{X}_V^1(\Omega)$, of

$$\tilde{E}(v) = \int_0^1 \|v_t\|_V^2 dt + U(\varphi_{01}^v)$$

Now, like in theorem 29, one can find, by taking a subsequence of a minimizing sequence, a sequence v^n in $\mathcal{X}_V^1(\Omega)$ which converges weakly to some $v \in \mathcal{X}_V^1(\Omega)$ and $\tilde{U}(v^n)$ tends to u . Since

$$\liminf \int_0^1 \|v_t^n\|_V^2 dt \geq \int_0^1 \|v_t\|_V^2 dt$$

and since weak convergence in $\mathcal{X}_V^1(\Omega)$ implies pointwise convergence of the flow (theorem 24) we also have $F(\varphi_{01}^{v^n}) \rightarrow U(\varphi_{01}^v)$ so that $\tilde{E}(v) = u$ and v is a minimizer. \square

Two examples for which this theorem directly applies are the point matching problem (U given by equation (29)) and the image matching (equation (30)), at least when I and I' are continuous over $\bar{\Omega}$.

3.2.2. Gradient descent algorithms. We now detail the gradient computations for energies of the kind (35), in the particular cases of point matching and image matching. This computation is the starting point of all optimization algorithms, and we refer, for example, to [18] for more details on how to build efficient optimization algorithms when once this issue is solved.

As remarked in the proof of theorem 30, the variational problem which has to be solved is preferably expressed as a problem over $\mathcal{X}_V^1(\Omega)$. The function which is minimized over this space takes the form

$$E(v) = \int_0^1 \|v_t\|_V^2 dt + F(\varphi_{01}^v)$$

What we will be looking for is the expression of the gradient for the Hilbert structure of $\mathcal{X}_V^1(\Omega)$, which is the natural inner-product structure of the problem. It is a function, denoted $\nabla E : v \mapsto \nabla_v^V E \in \mathcal{X}_V^1(\Omega)$, which defined by the formula, valid for v, h in $\mathcal{X}_V^1(\Omega)$

$$\frac{d}{d\varepsilon} E(v + \varepsilon h)|_{\varepsilon=0} = \langle \nabla_v^V E, h \rangle_{\mathcal{X}_V^1} = \int_0^1 \langle (\nabla_v^V)_t, h_t \rangle_V dt$$

Since the set V is fixed in this section, we will drop the exponent from the notation, and simply refer to the gradient $\nabla_v E$. Note that this is different from the Eulerian gradient we have dealt with before, this last definition being closer than the usual definition of the gradient. One important difference is that the gradient we here define is an element of $\mathcal{X}_V^1(\Omega)$, henceforth a time-dependent vector field, whereas the Eulerian gradient simply was an element of V (a vector field on Ω).

Since the first term in E is the squared norm of v in $\mathcal{X}_V^1(\Omega)$, its contribution to the gradient is straightforward, since, at $\varepsilon = 0$,

$$\frac{d}{d\varepsilon} \|v + \varepsilon h\|_{\mathcal{X}_V^1}^2 = 2\langle v, h \rangle_{\mathcal{X}_V^1}$$

inducing a first term given by $2v$ in the gradient. The matching term, $F(\varphi_{01}^v)$ requires more efforts. Instead of writing a general formula, we carry the computation for the two main examples we have considered so far (point and image matching). Once these cases are understood, it is not hard to devise new variants.

We start with image matching, and consider

$$F(\varphi_{01}^v) = \int_{\Omega} (I \circ \varphi_{10}^v - I')^2 dx$$

We assume that I is differentiable, and use theorem 22 which states that

$$\frac{d}{d\varepsilon} \varphi_{10}^{v+\varepsilon h}(x) = - \int_0^1 d_{\varphi_{1t}^v(x)} \varphi_{t0}^v h_t \circ \varphi_{1t}^v(x) dt.$$

This implies that

$$\begin{aligned} \frac{d}{d\varepsilon} F(\varphi_{01}^{v+\varepsilon h})|_{t=0} &= -2 \int_{\Omega} (I \circ \varphi_{10}^v - I')^t \nabla_{\varphi_{01}^v(x)} I \int_0^1 d_{\varphi_{1t}^v(x)} \varphi_{t0}^v h_t \circ \varphi_{1t}^v(x) dt dx \\ &= -2 \int_0^1 \int_{\Omega} {}^t \alpha_t(x) h_t \circ \varphi_{1t}^v(x) dx dt \end{aligned}$$

with $\alpha_t(x) = (I \circ \varphi_{10}^v - I')^t d_{\varphi_{1t}^v(x)} \varphi_{t0}^v \nabla_{\varphi_{01}^v(x)} I$. Introducing the reproducing kernel of V , this may be written as

$$\frac{d}{d\varepsilon} F(\varphi_{01}^{v+\varepsilon h})|_{t=0} = -2 \int_0^1 \int_{\Omega} \langle K_{\varphi_{1t}^v(x)}^{\alpha_t(x)}, h_t \rangle_V dx dt$$

(recall that K_y^a is the vector field: $z \mapsto K(z, y)a$). This implies that

$$(\nabla_v E)_t = 2v_t - 2 \int_{\Omega} K_{\varphi_{1t}^v(x)}^{\alpha_t(x)} dx.$$

We have just obtained the algorithm, where running time now is denoted by τ

ALGORITHM 5 (Flow based gradient descent for image matching). *Start with an initial guess $v^0 \in \mathcal{X}_V^1(\Omega)$ (for example $V = 0$).*

Let v evolve according to the equation

$$\frac{dv_t^\tau(y)}{d\tau} = -\gamma \left(v_t^\tau(y) - \int_{\Omega} K(y, \varphi_{1t}^{v^\tau}(x)) \alpha_t^\tau(x) dx \right)$$

with $\alpha_t^\tau(x) = (I \circ \varphi_{10}^{v^\tau} - I')^t d_{\varphi_{1t}^{v^\tau}(x)} \varphi_{t0}^{v^\tau} \nabla_{\varphi_{01}^{v^\tau}(x)} I$.

Consider now the point matching issue. In this case,

$$F(\varphi_{01}^v) = \sum_{i=1}^N |y_i - \varphi_{01}^v(x_i)|^2$$

so that theorem 22 directly implies that

$$\begin{aligned} \frac{d}{d\varepsilon} F(\varphi_{01}^{v+\varepsilon h})|_{t=0} &= -2 \sum_{i=1}^N {}^t (y_i - \varphi_{01}^v(x_i)) \int_0^1 d_{\varphi_{0t}^v(x_i)} \varphi_{t1}^v h_t \circ \varphi_{0t}^v(x_i) dt \\ &= -2 \sum_{i=1}^N \int_0^1 {}^t \alpha_t^i h_t \circ \varphi_{0t}^v(x_i) dt \end{aligned}$$

with $\alpha_t^i = {}^t d_{\varphi_{0t}^v(x_i)} \varphi_{t1}^v (y_i - \varphi_{01}^v(x_i))$. Thus

$$\frac{d}{d\varepsilon} F(\varphi_{01}^{v+\varepsilon h})|_{t=0} = -2 \sum_{i=1}^N \int_0^1 \langle K_{\varphi_{0t}^v(x_i)}^{\alpha_t^i}, h_t \rangle_V dt$$

which finally yields

$$(36) \quad (\nabla_v E)_t = 2v_t - 2 \sum_{i=1}^N K_{\varphi_{1t}^v(x_i)}^{\alpha_t^i}.$$

It is certainly possible to use this expression to devise an optimization algorithm, but one can be much more efficient in the point matching case. Equation (36) indeed implies that the optimal v (for which $\nabla_v E = 0$) can always be written

$$v_t = \sum_{i=1}^N K_{x_t^i}^{\alpha_t^i}$$

where $x_t^i = \varphi_{0t}^v(x_i)$. In this expression appear two unknown sets of ‘‘trajectories’’, the first one being the landmark trajectories, $x_t^i = \varphi_{0t}^v(x_i)$, the other ones (the $\alpha_t^i \in \mathbb{R}^k$) being called dual trajectories. These trajectories are not independent since we have

$$\frac{dx_t^j}{dt} = v_t(x_t^j) = \sum_{i=1}^N K(x_t^j, x_t^i) \alpha_t^i$$

so that, knowing α_t , we can retrieve x_t by integrating a system of ordinary differential equations, and knowing x_t , we can retrieve α_t by solving a linear system. Moreover, because K is a reproducing kernel, we have

$$\|v_t\|_V^2 = \sum_{i,j=1}^N \alpha_t^j K(x_t^j, x_t^i) \alpha_t^i$$

so that, for the optimal v ,

$$E(v) = \sum_{i,j=1}^N \int_0^1 \alpha_t^j K(x_t^j, x_t^i) \alpha_t^i dt + \sum_{i=1}^N |y_i - x_1^i|^2$$

The initial problem is therefore equivalent to minimizing

$$U(\alpha, x) = \sum_{i,j=1}^N \int_0^1 \alpha_t^j K(x_t^j, x_t^i) \alpha_t^i dt + \sum_{i=1}^N |y_i - x_1^i|^2$$

under the constraints that, for $j = 1, \dots, N$,

$$\frac{dx_t^j}{dt} = \sum_{i=1}^N K(x_t^j, x_t^i) \alpha_t^i$$

By solving these constraints with respect to x or to α , we obtain two different problems. The simplest one is expressed with respect to x . Let $S(x_t) = S(x_t^1, \dots, x_t^N)$ be the block matrix with (i, j) block given by $K(x_t^i, x_t^j)$. The constraints write

$$\frac{dx_t}{dt} = S(x_t) \alpha_t$$

so that $\alpha_t = S(x_t)^{-1} \frac{dx_t}{dt}$ and the minimization reduces to

$$U(x) = \int_0^1 \left(\frac{dx_t}{dt} \right) S(x_t)^{-1} \frac{dx_t}{dt} dt + \|y - x_1\|^2$$

Minimizing this function with respect to x by gradient descent is possible, and has been achieved in [15, 16]. The basic computation is as follows: if $s_{pq}^r = \frac{ds_{pq}}{dx^r}$, we

can write

$$\begin{aligned} \frac{d}{d\varepsilon} U(x + \varepsilon\xi)|_{\varepsilon=0} &= \int_0^1 {}^t \left(\frac{dx_t}{dt} \right) S(x_t)^{-1} \frac{d\xi_t}{dt} dt \\ &\quad - \int_0^1 \sum_{p,q,r} \alpha_t^p \alpha_t^r s_{pq}^r(x) \xi_t^q dt - 2^t (y - x_1) \xi_1 \end{aligned}$$

After a change of variables in the first integral, we obtain a gradient given by

$$-2 \frac{d}{dt} \left(S(x_t)^{-1} \frac{dx_t}{dt} \right) - z_t + \left(S(x_1)^{-1} \frac{dx_t}{dt} \Big|_{t=1} - 2(y - x_1) \right) \delta_1$$

where $z_t^q = \sum_{p,r} \alpha_t^p \alpha_t^r s_{pq}^r(x)$ and δ_1 is the Dirac measure at $t = 1$ (this “gradient” is in fact a generalized function).

This generalized form of the gradient (even if it turns into a computable expression once the problem is discretized) may lead to numerical unstabilities. One way to address this is to compute the gradient in a Hilbert space in which the evaluation function $x(\cdot) \mapsto x_1$ is a continuous linear form. This method has been introduced, in particular, in [12]. Let H be the space of all landmark trajectories $x : t \mapsto x_t = (x_t^1, \dots, x_t^N)$, with fixed starting point x_0 and free end-point x_1 , and square integrable time derivative. This is a space of the form $x_0 + H$ where H is the Hilbert space of time dependent function $t \mapsto \xi_t$, is as a column vector of size Nk :

$$\langle \xi, \xi' \rangle_H = \int_0^1 {}^t \frac{d\xi}{dt} \frac{d\xi'}{dt} dt + {}^t \xi_1 \xi_1'$$

What we need to do, now, in order to obtain the gradient for this inner product, is to express $\frac{d}{d\varepsilon} U(x + \varepsilon\xi)|_{\varepsilon=0}$ under the form $\langle \nabla_x^H U, \xi \rangle_H$. We will make the assumption that

$$\int_0^1 \left| S(x_t)^{-1} \frac{dx_t}{dt} \right|^2 dt < \infty$$

which implies that

$$\int_0^1 {}^t \left(\frac{dx_t}{dt} \right) S(x_t)^{-1} \frac{d\xi_t}{dt} dt \leq \sqrt{\int_0^1 \left| S(x_t)^{-1} \frac{dx_t}{dt} \right|^2 dt \int_0^1 \left| \frac{d\xi_t}{dt} \right|^2 dt}$$

is continuous in ξ . Similarly, the linear form $\xi \mapsto {}^t (y - x_1) \xi_1$ is continuous since

$${}^t (y - x_1) \xi_1 \leq |y - x_1| |x_1|$$

and so is $\xi \mapsto \int_0^1 {}^t z_t \xi_t dt$, since, letting

$$\eta_t = \int_0^t z_s ds,$$

and assuming that it is square integrable over $[0, 1]$, we have

$$\int_0^1 {}^t z_t \xi_t dt = \eta_1 \xi_1 - \int_0^1 \eta_t \frac{d\xi_t}{dt} dt$$

Thus,

$$\xi \mapsto \frac{d}{d\varepsilon} U(x + \varepsilon\xi)|_{\varepsilon=0}$$

is continuous over H , and the Riesz representation theorem implies that $\nabla_x^H U$ exists as an element of H . We now proceed to its computation. Letting

$$\mu_t = 2 \int_0^1 S(x_t)^{-1} \frac{dx_t}{dt}$$

and $a = 2(y - x_1)$, the problem is to find $\zeta \in H$ such that, for all $\xi \in H$,

$$\langle \zeta, \xi \rangle_H = \int_0^1 \frac{d\mu}{dt} \frac{d\xi}{dt} dt + \int_0^1 {}^t z_t \xi_t dt - {}^t a \xi_1$$

which can also be written

$$\int_0^1 \left(\frac{d\zeta}{dt} + \zeta_1 \right) \frac{d\xi}{dt} dt = \int_0^1 \left(\frac{d\mu}{dt} + \eta_1 - \eta_t - a \right) \frac{d\xi}{dt} dt$$

which suggests to select ζ such that $\zeta_0 = 0$ and

$$\frac{d\zeta}{dt} + \zeta_1 = \frac{d\mu}{dt} + \eta_1 - \eta_t - a$$

which implies

$$\zeta_t + t\zeta_1 = \mu_t - \int_0^t \eta_s ds + t(\eta_1 - a).$$

At $t = 1$, this yields

$$2\zeta_1 = \mu_1 - \int_0^1 \eta_s ds + \eta_1 - a$$

and we finally obtain

$$\zeta_t = \mu_t - \int_0^t \eta_s ds + \frac{t}{2} \left(\int_0^1 \eta_s ds - \mu_1 + \eta_1 - a \right) / 2$$

We summarize this in an algorithm, in which τ is again the computation time

ALGORITHM 6 (Gradient descent algorithm for landmark matching). *Start with initial landmark trajectories $x_t(0) = (x_t^1(0), \dots, x_t^N(0))$.*

Solve

$$\frac{dx_t(\tau)}{d\tau} = -\gamma \left(\mu_t(\tau) - \int_0^t \eta_s(\tau) ds + \frac{t}{2} \left(\int_0^1 \eta_s(\tau) ds - \mu_1(\tau) + \eta_1(\tau) - a(\tau) \right) / 2 \right)$$

with $a(\tau) = 2(y - x_1(\tau))$, $\mu_t(\tau) = \int_0^t \alpha_s(\tau) ds$, $\eta_t(\tau) = \int_0^t z_s(\tau) dt$ and

$$\alpha_t(\tau) = S(x_t(\tau))^{-1} \frac{dx_t(\tau)}{dt}$$

$$z_t^q(\tau) = \sum_{p,r} \alpha_t^p(\tau) \alpha_t^r(\tau) s_{pq}^r(x_t(\tau)).$$

There is a third point of view, which we shall not detail here, which consists in considering U as a function of $\alpha_1, \dots, \alpha_N$ instead of a function of x_1, \dots, x_N . The computations are harder in this case, because x cannot be explicitly computed in function of α but only by solving a differential equation (see [4]).

Groups, manifolds and invariant metrics

1. Introduction

In the previous sections,

2. Differential manifolds

2.1. Definition. A differential manifold is a set within which points may be described by coordinate systems, which must satisfy some compatibility constraints from which one can develop intrinsic differential operations. We start with the definition:

DEFINITION 10. Let M be a topological Hausdorff space. An N -dimensional local chart on M is a pair (U, Φ) where U is an open subset of M and Φ a homeomorphism between U and some open subset of \mathbb{R}^n .

Two n -dimensional local charts, (U_1, Φ_1) and (U_2, Φ_2) are C^∞ -consistent if the function $\Phi_1 \circ \Phi_2^{-1}$ is a C^∞ -diffeomorphism between $\Phi_2(U_2 \cap U_1)$ and $\Phi_1(U_2 \cap U_1)$.

An n -dimensional atlas on M is a family of pairwise consistent local charts $((U_i, \Phi_i), i \in I)$, such that $M = \bigcup_I U_i$. Two atlases on M are equivalent if their union is also an atlas, ie. if every local chart of the first one is consistent with every local chart of the second one.

A Hausdorff space with an n -dimensional atlas is called an n -dimensional (C^∞) differential manifold n .

If M is a manifold, a local chart on M will always be assumed to be consistent with the atlas on M . If M and N are two manifolds, their product $M \times N$ is also a manifold: if (U, Φ) is a chart on M , (V, Ψ) a chart on N , $(U \times V, (\Phi, \Psi))$ is a chart on $M \times N$.

When a local chart (U, Φ) is given, the coordinates functions x_1, \dots, x_n are defined by $\Phi(m) = (x_1(m), \dots, x_n(m))$ for $m \in U$. Formally, x_i is a function from U to \mathbb{R} . However, when a point m is given, one generally refers to $x_i = x_i(m) \in \mathbb{R}$ as the i th coordinate of m in the chart (U, Φ) .

According to these definition, \mathbb{R}^n is a differential manifold, and so are the open sets in \mathbb{R}^n . An other example is given by the n -dimensional sphere, S^n , defined as the set of points $m \in \mathbb{R}^{n+1}$ with $|m| = 1$. The sphere can be equipped with a $2(n+1)$ -chart atlas (U_i, Φ_i) , $i = 1, \dots, 2(n+1)$, letting

$$U_1 = \{m = (m_1, \dots, m_{n+1}) : m_1 > 0\} \cap S^n$$

and $\Phi_1(m) = (m_2, \dots, m_{n+1}) \in \mathbb{R}^n$, (U_2, Φ_2) being the same, with $m_1 < 0$ instead, and so on with the other coordinates m_2, \dots, m_{n+1} .

We now consider functions on manifolds.

DEFINITION 11. A function $\psi : M \rightarrow \mathbb{R}$ is C^∞ , if, for all local chart (U, Φ) on M , the function $\psi \circ \Phi^{-1} : \Phi(U) \subset \mathbb{R}^n \rightarrow \mathbb{R}$, is C^∞ in the usual sense. The function $\psi \circ \Phi^{-1}$ is called the interpretation of ψ in (U, Φ) .

From the consistency condition, if this property is true for a family of charts which covers M , it is true for all charts. The set of C^∞ functions on M is denoted $C^\infty(M)$. If U is open in M , the set $C^\infty(U)$ contains functions defined on U which can be interpreted as C^∞ functions of the coordinates for all local chart of M which is contained in U . The first example of C^∞ functions are the coordinates: if (U, Φ) is a chart, the i th coordinate $(x_i(m), m \in U)$ belongs to $C^\infty(U)$, since, when interpreted in (U, Φ) , it reduces to $(x_1, \dots, x_n) \mapsto x_i$.

2.2. Vector fields, tangent spaces. We fix, in this section, a differential manifold, denoted M , of dimension n .

DEFINITION 12. A vector field on M is a function $X : C^\infty(M) \rightarrow C^\infty(M)$, such that

$\forall \alpha, \beta \in \mathbb{R}, \forall \varphi, \psi \in C^\infty(M)$:

$$\begin{aligned} X(\alpha.\varphi + \beta.\psi) &= \alpha.X(\varphi) + \beta.X(\psi), \\ X(\varphi\psi) &= X(\varphi)\psi + \varphi X(\psi). \end{aligned}$$

The set of vector fields on M is denoted by $\mathcal{X}(M)$.

DEFINITION 13. If $m \in M$, a tangent vector to M at m is a function $\xi : C^\infty(M) \rightarrow \mathbb{R}$ such that:

$\forall \alpha, \beta \in \mathbb{R}, \forall \varphi, \psi \in C^\infty(M)$:

$$\begin{aligned} \xi(\alpha.\varphi + \beta.\psi) &= \alpha.\xi(\varphi) + \beta.\xi(\psi), \\ \xi(\varphi\psi) &= \xi(\varphi)\psi(m) + \varphi(m)\xi(\psi). \end{aligned}$$

The set of tangent vectors to M at m is denoted $T_m M$.

If $X \in \mathcal{X}(M)$ is a vector field on M , and if $m \in M$, we may define $X_m : C^\infty(M) \rightarrow \mathbb{R}$ by

$$X_m(\varphi) = (X(\varphi))(m)$$

to obtain a tangent vector at m . Conversely, if a collection $(X_m \in T_m M, m \in M)$ is given, we can define $(X(\varphi))(m) = X_m(\varphi)$ for $\varphi \in C^\infty(M)$ and $m \in M$; X will be a vector field on M if and only if, for all $\varphi \in C^\infty(M)$, $m \mapsto X_m(\varphi)$ is C^∞ . Finally, one can show that, for all $\xi \in T_m M$, there exists a vector field X such that $\xi = X_m$ ([13]).

The linear nature of definitions 12 and 13 is clarified in the next proposition

PROPOSITION 14. For all $m \in M$, the tangent space $T_m M$ is an n -dimensional vector space.

PROOF. Let $C = (U, \Phi)$ be a local chart with $m \in U$; denote $x^0 = \Phi(m)$, $x^0 \in \mathbb{R}^n$. If $\varphi \in C^\infty(M)$, then

$$\begin{aligned} \varphi_C : \varphi(U) \subset \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto \varphi \circ \Phi^{-1}(x) \end{aligned}$$

is C^∞ . Define

$$(\partial_m x_i)(\varphi) := \frac{\partial \varphi_C}{\partial x_i} \Big|_{x=x^0}.$$

It is easily checked that $\partial_m x_i$ satisfies the conditions in definition 13, so that $\partial_m x_i \in T_m M$. We show that every $\xi \in T_m M$ may be uniquely written under the form

$$(37) \quad \xi = \sum_{i=1}^n \lambda_i \partial_m x_i.$$

Indeed,

$$\varphi_C(x) = \varphi_C(x^0) + \sum_{i=1}^n (x_i - x_i^0) \psi_i^*(x)$$

with $\psi_i^*(x) = \int_0^1 \frac{\partial \varphi_C}{\partial x_i}(x^0 + t(x - x^0)) dt$. Thus, if $m' \in U$

$$\varphi(m') = \varphi_0 + \sum_{i=1}^n (x_i(m') - x_i(m)) \psi_i(m')$$

with $\psi_i(m') = \psi_i^*(\Phi(m'))$, and $\varphi_0 = \varphi(m)$. If $\xi \in T_m M$ and f is constant, we have $\xi.f = f\xi(1) = f\xi(1^2) = 2f\xi(1)$ so that $\xi.f = 0$. Thus, for all $\xi \in T_m M$,

$$\xi(\varphi) = 0 + \sum_{i=1}^n \psi_i(m) \xi(x_i).$$

But $\psi_i(m) = (\partial_m x_i)(\varphi)$, which yields (37) with $\lambda_i = \xi(x_i)$. \square

This implies that within a local chart, a vector field can always be interpreted under the form

$$X = \sum_{i=1}^n \varphi_i \partial x_i$$

with $\varphi_i \in C^\infty(M)$ and $[\partial x_i]_m = \partial_m x_i$.

There is another standard definition of tangent vectors on M , associated to differentiable curves on M . This starts with the definitions

DEFINITION 14. *Let $t \mapsto \mu_t \in M$ be continuous, $\mu : [0, T] \rightarrow M$. One says that this curve is C^∞ if, for all local chart $C = (U, \Phi)$, the curve $\mu^C : s \mapsto \Phi \circ \mu_s$, defined on $\{t \in [0, T] : \mu_t \in U\}$ is C^∞ .*

Let $m \in M$. One says that two C^∞ curves, μ and ν , starting at m (ie. $\mu_0 = \nu_0 = m$) have the same tangent at m , if and only if, for all chart $C = (U, \Phi)$, the curves μ^C and ν^C have the same derivatives at $t = 0$.

PROPOSITION 15. *The tangential equality in m is an equivalence relation. The tangent space to M at m can be identified to the set of equivalence classes for this relation.*

PROOF. We sketch the argument. If a curve μ is given, with $\mu_0 = m$, define, for $\varphi \in C^\infty(M)$,

$$\xi_\mu \cdot \varphi = \frac{d}{dt} \Big|_{t=0} \varphi \circ \mu$$

It can be checked that $\xi_\mu \in T_m M$, and that $\xi_\mu = \xi_\nu$ if μ and ν have the same tangent at m . Conversely, if $\xi \in T_m M$ there exists a curve μ such that $\xi = \xi_\mu$, and

the equivalence class of μ is uniquely specified by μ . To show this, consider a chart (U, Φ) . We must have

$$\xi_\mu(x_i) = \xi(x_i) = \frac{d}{dt}\Big|_{t=0} x_i \circ \mu(\cdot)$$

which indeed shows that the tangent to $\Phi \circ \mathbf{m}$ is uniquely defined. To define μ , start from a line segment in $\Phi(U)$, passing through $\Phi(m)$, with direction given by $(\xi(x_i), i = 1, \dots, n)$ and apply Φ^{-1} to it to obtain a curve on M . \square

When $X \in \mathcal{X}(M)$ is given, one can consider the differential equation

$$\frac{d\mu_t}{dt} = X_{\mu_t}$$

Such a differential equation always admits a unique solution, with some initial condition $\mu_0 = m$, at least for t small enough. This can be proved quite easily by translating the problem in a local chart and applying the usual theorem on the existence and uniqueness of differential equations on \mathbb{R}^n .

DEFINITION 15. If $\varphi \in C^\infty(M)$, one defines a linear form on $T_m M$ by $\xi \mapsto \xi(\varphi)$. It will be denoted $d_m \varphi$, and called the differential of φ at m .

2.3. Maps between two manifolds.

DEFINITION 16. Let M and M' be two differential manifolds. A map $\Phi : M \rightarrow M'$ has class C^∞ if and only if, for all $\varphi \in C^\infty(M')$, one has $\varphi \circ \Phi \in C^\infty(M)$.

DEFINITION 17. If $m \in M$ and $m' = \Phi(m)$, we define the tangent map of Φ at m ,

$$d_m \Phi : T_m M \rightarrow T_{m'} M'$$

by

$$(d_m \Phi \cdot \xi)(\varphi) = \xi(\varphi \circ \Phi)$$

($\xi \in T_m M$, $\varphi \in C^\infty(M')$).

The tangent map $d_m \Phi$ is also called the differential of Φ at m .

3. Submanifolds

One efficient way for building manifolds is to characterize them as submanifolds of simple manifolds like \mathbb{R}^n . If M is a manifold, a submanifold of M is a set on which it is always possible to express a fixed number of coordinates in function of the others.

DEFINITION 18. Let M and P be two differentiable manifolds of dimension n and p , with $P \subset M$. We say that P is a submanifold of M , if, for all $m_0 \in P$, there exists a local chart (U, Φ) of M such that $m_0 \in U$, with local coordinates (x_1, \dots, x_n) , such that

$$U \cap P = \{m \in M : x_i = 0, i = p + 1, \dots, n\}$$

The next theorem is one of the main tools for defining manifolds:

THEOREM 31. *Let M be a differential manifold, Φ a differentiable map from M to \mathbb{R}^k . Let $a \in \Phi(M)$ and*

$$P = \Phi^{-1}(a) = \{m \in M : \Phi(m) = a\}$$

If there exists an integer q , such that, for all $m \in P$, the linear map $d_m\Phi : T_mM \rightarrow \mathbb{R}^k$ has constant rank q (independent of m), then P is a submanifold of M , with dimension $p = n - q$.

This applies to the sphere S^n , defined by $x_1^2 + \dots + x_{n+1}^2 = 1$, which is a submanifold of \mathbb{R}^{n+1} .

If $P \subset M$ is a submanifold of M defined as in theorem 31, the tangent space to P at m is identified to the kernel of $d_m\Phi$ in T_mM :

$$T_mP = \{\xi \in T_mM, d_m\Phi.\xi = 0\}$$

Another way to define submanifolds is via embeddings, as defined below

DEFINITION 19. *Let M and P be two differentiable manifolds. An embedding of M into P is a C^∞ map $\Phi : M \rightarrow P$, such that*

- i) For all $m \in M$, the tangent map, $d_m\Phi$ is one-to-one, from T_mM to $T_{\Phi(m)}P$.*
- ii) Φ is a homeomorphism between M and $\Phi(M)$ (this last set being considered with the topology induced by P)*

The second condition means the following: Φ is one to one, and, for all open subset U in M , there exists an open subset V in P such that $\Phi(U) = V \cap \Phi(M)$.

We then have:

PROPOSITION 16. *If $\Phi : M \rightarrow P$ is an embedding, then $\Phi(M)$ is a submanifold of P , with same dimension as M .*

4. Lie Groups

4.1. Definitions. A group is a set G with a composition rule $(g, h) \mapsto gh$ which is associative, has an identity element (denoted id_G , or id if there is no risk of confusion) and such that every element in G has an inverse in G . A Lie group is both a group and a differentiable manifold, such that the operations $(g, h) \mapsto g.h$ et $g \mapsto g^{-1}$, respectively from $G \times G$ to G and from G to G are C^∞ .

4.2. Lie algebra of a Lie group. If G is a Lie group, $g \in G$ and $\varphi \in C^\infty(G)$, one defines $\varphi.g \in C^\infty(G)$ by $(\varphi.g)(g') = \varphi(g'.g)$. A vector field on G is right-invariant if, for all $g \in G, \varphi \in C^\infty(G)$, one has $X(\varphi.g) = (X(\varphi)).g$. Denoting by R_g the right translation on G (defined by $R_g(g') = g'.g$), right invariance is equivalent to the identity $X = dR_g.X$. The set of right-invariant vector fields is called the Lie algebra of the group G , and denoted \mathfrak{g} .

Since $(X(\varphi.g))(\text{id}) = ((X(\varphi)).g)(\text{id}) = (X(\varphi))(g)$ whenever $X \in \mathfrak{g}$, an element X of \mathfrak{g} is entirely specified by the values of $X(\varphi)(\text{id})$ for $\varphi \in C^\infty(G)$. This implies that the Lie algebra \mathfrak{g} may be identified to the tangent space to G at id , $T_{\text{id}}G$. If $\xi \in T_{\text{id}}G$, its associated right invariant vector field is

$$X^\xi : g \mapsto X_g = dR_g\xi \in T_gG$$

The operation which provides an algebra structure on \mathfrak{g} is the Lie bracket. Recall that a vector field on a manifold M is a function $X : C^\infty(M) \rightarrow C^\infty(M)$, which satisfies the conditions of definition 12. When X and Y are two vector fields,

it is possible to chain them and compute $(XY)(\varphi) = X(Y(\varphi))$; XY also transforms C^∞ functions into C^∞ functions, but will not satisfy the conditions of definition 12, essentially because it involves second derivatives. However, it is easy to check that the difference $XY - YX$ is a vector field on M , denoted $[X, Y]$, and called the bracket of X and Y . A few important properties of Lie brackets are listed (without proof) in the next proposition:

PROPOSITION 17.

- i. $[X, Y] = -[Y, X]$
- ii. $[[X, Y], Z] = [X, [Y, Z]]$
- iii. $[[X, Y], Z] + [[Z, X], Y] + [[Y, Z], X] = 0$
- iv. If $\Phi \in C^\infty(M, N)$, $d\Phi[X, Y] = [d\Phi X, d\Phi Y]$

The last property is important for Lie groups, since, applied with $\Phi = R_g : G \mapsto G$, and $X, Y \in \mathfrak{g}$, it yields

$$dR_g[X, Y] = [dR_g X, dR_g Y] = [X, Y]$$

so that $[X, Y] \in \mathfrak{g}$. The Lie algebra of G is stable by the Lie bracket operation. Because of the identification of \mathfrak{g} with $T_{\text{id}} G$, the bracket notation can also be used for tangent vectors at the identity, letting $[\xi, \eta] = [X^\xi, X^\eta]_{\text{id}}$.

There is another possible definition of the Lie bracket on \mathfrak{g} . For $g \in G$, one can define the group isomorphism $I_g : h \mapsto ghg^{-1}$. It is differentiable, and the differential of I_g at $h = \text{id}$ is called the adjoint map, denoted $Ad_g : T_{\text{id}} G \rightarrow T_{\text{id}} G$. We therefore have, for $\eta \in T_{\text{id}} G$,

$$Ad_g(\eta) = d_{\text{id}} I_g \eta$$

We may now consider the application $U_\eta : g \mapsto Ad_g(\eta)$ which is defined on G and takes values in $T_{\text{id}} G$. We then have:

$$(38) \quad d_{\text{id}} U_\eta \xi = [\xi, \eta]$$

The notation $ad_\xi \eta = [\xi, \eta]$ is commonly used to represent the Lie bracket.

When a vector field $X \in \mathfrak{g}$ is given, the solution of the associated differential equation

$$\frac{d\mu}{dt} = X_{\mu_t}$$

with initial condition $\mu_0 = \text{id}$ always exists, not only for small time, but for arbitrary times. The small time existence comes from the general theory of ODEs, and the existence for arbitrary time comes from the fact that, wherever it is defined, μ_t satisfies the semigroup property $\mu_{t+s} = \mu_t \mu_s$: this implies that if μ_t is defined on some interval $[0, T]$, one can always extend it to $[0, 2T]$ by letting $\mu_t = \mu_{\frac{t}{2}}^2$. The semigroup property can be shown as follows: if $X = X^\xi$, for $\xi \in T_{\text{id}} G$, the ODE can be written

$$\frac{d\mu_t}{dt} = d_{\text{id}} R_{\mu_t} \xi$$

Consider now $\nu : s \mapsto \mu_{t+s}$. It is solution of the same equation with initial condition $\nu_0 = \mu_t$. If $\tilde{\nu}_s = \mu_s \mu_t$, we have

$$\frac{d\tilde{\nu}_s}{ds} = d_{\mu_s} R_{\mu_t} \frac{d\mu_s}{ds} = d_{\mu_s} R_{\mu_t} d_{\text{id}} R_{\mu_t} \xi = d_{\text{id}} (R_{\mu_t} \circ R_{\mu_s}) \xi = d_{\text{id}} R_{\tilde{\nu}_s} \xi$$

Thus, ν_s and $\tilde{\nu}_s$ satisfy the same differential equation, with the same value, μ_t , at $s = 0$, and therefore coincide, which is the semigroup property.

This solution, μ_t , is called the exponential map on G , and is denoted $\exp(tX)$ or $\exp(t\xi)$ if $X = X^\xi$. The semigroup property writes $\exp((t+s)X) = \exp(tX)\exp(sX)$. Equation (38) can be written, using the exponential map

$$\frac{d}{dt} \frac{d}{ds} \exp(t\xi) \exp(s\eta) \exp(-t\xi) = [\xi, \eta].$$

We finally quote the last important property of the exponential map ([13]):

THEOREM 32. *There exists a neighborhood V of 0 in \mathfrak{g} and a neighborhood U of id in G such that \exp is a diffeomorphism between V and U .*

4.3. Example: finite dimensional transformation groups. These transformation groups, and in particular the matrix groups, are fundamental examples of Lie groups. Denote by $\mathcal{M}_n(\mathbb{R})$ the n^2 dimensional space of real $n \times n$ matrices. For $i, j \in \{1, \dots, n\}$, denote ∂x_{ij} for the matrix with (i, j) coefficient equal to 1, and all others equal to 0. Let Id_n denote the identity matrix.

4.3.1. Linear group. $GL_n(\mathbb{R})$ is the group of invertible matrices in $\mathcal{M}_n(\mathbb{R})$. It is open in $\mathcal{M}_n(\mathbb{R})$ and therefore is a submanifold of this space, of same dimension, n^2 . The Lie algebra of $GL_n(\mathbb{R})$ is equal to $\mathcal{M}_n(\mathbb{R})$, and is generated by all $(\partial x_{ij}, i, j = 1, \dots, n)$.

if $\xi \in \mathcal{M}_n(\mathbb{R})$, the associated right-invariant vector field is $X^\xi : g \mapsto \xi g$. The adjoint map is $\eta \mapsto g\eta g^{-1}$, and the Lie bracket is $[\xi, \eta] = \xi\eta - \eta\xi$. Finally, the exponential is the usual matrix exponential:

$$\exp(\xi) = \sum_{k=0}^{\infty} \frac{\xi^k}{k!}$$

4.3.2. Special linear group. $SL_n(\mathbb{R})$ is the subgroup of $GL_n(\mathbb{R})$ containing matrices with determinant 1. The determinant is a C^∞ function. Its derivative at $g \in GL_n(\mathbb{R})$ is a linear map from $\mathcal{M}_n(\mathbb{R})$ to \mathbb{R} , given by $(d_g \det) \cdot \xi = \det(g) \text{trace}(g^{-1}\xi)$. Since it has rank one, theorem 31 implies that $SL_n(\mathbb{R})$ is a submanifold of $GL_n(\mathbb{R})$, of dimension $n^2 - 1$. The Lie algebra of $SL_n(\mathbb{R})$ is defined by $d_{\text{Id}} \det = 0$, and therefore consists of matrices with vanishing trace.

4.3.3. Rotations. $O_n(\mathbb{R})$ is the group of matrices g such that ${}^t g \cdot g = \text{Id}$. $SO_n(\mathbb{R})$ is the subgroup of $O_n(\mathbb{R})$ containing matrices of determinant 1. The map $\Phi : g \mapsto {}^t g \cdot g$ is C^∞ , its differential is

$$d_g \Phi \cdot \xi = {}^t g \cdot \xi + {}^t \xi \cdot g$$

The kernel of $d_g \Phi$ therefore contains matrices $\xi = g\eta, \eta$ skew symmetrical, and has dimension $n(n-1)/2$. Thus, again by theorem 31, $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$ are submanifolds of $\mathcal{M}_n(\mathbb{R})$, of dimension $n(n-1)/2$.

The Lie algebra of $O_n(\mathbb{R})$ contains skew-symmetric matrices.

4.3.4. Similitudes. $Sim_n(\mathbb{R})$ is the group of similitudes, composed with matrices g such that ${}^t g \cdot g = \lambda \text{Id}$, for some $\lambda > 0$ in \mathbb{R} . In fact, one must have $\lambda = \det(g)^{2/n}$, so that $Sim_n(\mathbb{R})$ is the set of invertible matrices for which $\Phi(g) = 0$, with

$$\begin{aligned} \Phi : GL_n(\mathbb{R}) &\rightarrow \mathcal{M}_n(\mathbb{R}) \\ g &\mapsto {}^t g g - \det(g)^{2/n} \text{Id} \end{aligned}$$

One can check that Φ has constant rank and that $Sim_n(\mathbb{R})$ is a submanifold of $\mathcal{M}_n(\mathbb{R})$ of dimension $1 + n(n-1)/2$.

Lie algebra of $Sim_n(\mathbb{R})$, contains matrices of the form $\alpha \text{Id} + \xi$, with ξ skew-symmetrical.

5. Group action

5.1. Definitions. One says that a group G acts (on the left) on a set M , if there exists a map, from $G \times M$ to M which associates to a pair (g, m) the result of the action of g on m , denoted $g.m$, with the properties that $g.(h.m) = (gh).m$ and $\text{id}.m = m$.

The orbit of $m \in M$ under this action is the set $G.m = \{g.m, g \in G\}$. Orbits either coincide or are disjoint, and they form a partition of M . We let $M/G = \{G.m, m \in M\}$.

The action of G is transitive if there exists only one orbit, ie. for all $m, m' \in M$, there exists $g \in G$ such that $g.m = m'$.

The isotropy subgroup of a point $m \in M$ is the collection of elements $g \in G$ such that $g.m = m$. It is denoted G_m , and forms a subgroup of G , (ie. it is stable by the group product and by the group inverse). The isotropy subgroup of M is the intersection of all G_m , and denoted G_M .

When G is a Lie group and M is a manifold, one generally assumes that, in addition, the map $(g, m) \mapsto g.m$ is C^∞ .

5.2. Homogeneous spaces. If H is a subgroup of G , the map $(h, g) \mapsto gh$ defines a right action of H on G . The coset space G/H is the set of orbits, $\{gH, g \in G\}$ for this action. When G is a Lie group and H a closed subgroup of G , G/H is called a homogeneous space. The differential structure of G can be projected onto G/H to provide this set with a structure of differential manifold.

The group G acts on the left on G/H through $g.(g'H) = (gg')H$. This action is transitive and H is the isotropy subgroup of $(\text{id}H)$. Conversely, the following is true

PROPOSITION 18. *Let G be a group acting on the left on a set M . Fix $m \in M$ and let G_m be the isotropy subgroup of m . The map*

$$\begin{aligned} \Phi : \quad G/G_m &\rightarrow M \\ gG_m &\mapsto g.m \end{aligned}$$

is a diffeomorphism which carries the action of G on G/G_m to the action of G on M .

5.3. Group action and distances. Recall that a distance d on a set M is a map: $M^2 \mapsto \mathbb{R}_+$ such that, for all $m, m', m'' \in M$,

$$\text{D1) } d(m, m') = 0 \Leftrightarrow m = m'.$$

$$\text{D2) } d(m, m') = d(m', m)$$

$$\text{D3) } d(m, m'') \leq d(m, m') + d(m', m'').$$

If D1) is not true, and $d(m, m) = 0$ for all m , one says that d is a *pseudo-distance*.

If G is a group acting on M , one says that a distance d on M is G -equivariant if and only if, for all $g \in G$, for all $m, m' \in M$, $d(g.m, g.m') = d(m, m')$. One says that d is G -invariant if and only if it satisfies D2) and D3) and D1) is replaced by D1') $d(m, m') = 0 \Leftrightarrow \exists g \in G, g.m = m'$.

A G -invariant distance is in fact a distance on the coset space M/G , through the identification $d(G.m, G.m') \sim d(m, m')$.

A G -equivariant distance may be projected to induce a G -invariant distance:

PROPOSITION 19. *Let d be a G -equivariant distance on M . The function \tilde{d} , defined by*

$$(39) \quad \tilde{d}(G.m, G.m') = \inf\{d(gm, g'm'), g, g' \in G\}$$

is a (pseudo-)distance on M/G .

Since d is G -equivariant, \tilde{d} is also given by

$$\tilde{d}(G.m, G.m') = \inf\{d(gm, m'), g \in G\}$$

PROOF. Let us show that \tilde{d} is a pseudo distance. Symmetry is obvious from (39). For the triangle inequality (D3), it suffices to show that, for all $g_1, g'_1, g'_2, g''_1 \in G$, there exists $g_2, g''_2 \in G$ such that

$$d(g_2m, g''_2m'') \leq d(g_1m, g'_1m') + d(g'_2m', g''_1m'')$$

But, since $d(g'_2m', g''_1m'') = d(g'_1m', g'_1(g'_2)^{-1}g''_1m'')$, one can take $g_2 = g_1$ and $g''_2 = g'_1(g'_2)^{-1}g''_1$ and apply the triangle inequality for d . \square

This can be applied in a slightly different way. Let G be a group acting on M . We here show how a distance on the product space $M^G = G \times M$ can be projected on M .

The group G acts on the left on M^G , letting, for $k \in G$, $o = (h, m) \in M^G$: $k.o = (hk^{-1}, km)$. For $o = (h, m) \in M^G$, denote $\pi(o) = h.m$. For all $k \in G$, one has $\pi(k.o) = \pi(o)$.

Let d^G be a distance on M^G . We let, for $m, m' \in M$

$$(40) \quad d(m, m') = \inf\{d^G(o, o'), o, o' \in \mathcal{O}, \pi(o) = m, \pi(o') = m'\}$$

We have the proposition

PROPOSITION 20. *If d^G is G -equivariant, then d , as defined in (40) is a pseudo-distance on M*

This result is in fact a corollary of the previous one, provided it is noted that the quotient space M^G/G can be identified to M and that, though this identification, the definitions of d in (40) and in (39) are the same.

6. Riemannian manifolds

6.1. Introduction. In this section, M is a differential manifold of dimension n .

DEFINITION 20. *A Riemannian structure on M is the definition of a C^∞ inner product between vector fields:*

$$(X, Y) \in \mathcal{X}(M) \times \mathcal{X}(M) \mapsto \langle X, Y \rangle \in C^\infty(M)$$

such that $\langle X, Y \rangle_M = \langle Y, X \rangle_M$, $\langle X, X \rangle \geq 0$, and $\langle X, X \rangle = 0$ if and only if $X = 0$, and for all $\varphi, \psi \in C^\infty(M)$

$$\langle \varphi X + \psi X', Y \rangle = \varphi \langle X, Y \rangle + \psi \langle X', Y \rangle$$

The value of $\langle X, Y \rangle$ at $m \in M$ will be denoted $\langle X, Y \rangle_m$, and it can be shown that it only depends on the values X_m and Y_m . It is equivalent to assume that, for all $m \in M$, an inner product denoted $\langle \cdot, \cdot \rangle_m$ is given on $T_m M$, which is such that, if X and Y are vector fields, the function $m \mapsto \langle X_m, Y_m \rangle_m$ is C^∞ . We shall use the notation

$$\|\xi\|_m = \sqrt{\langle \xi, \xi \rangle_m}.$$

In a local chart, $C = (U, \Phi)$, with coordinates (x_1, \dots, x_n) , a tangent vector at $m \in U$ can be written as a linear combination of the $\partial_m x_i$. From elementary linear algebra, there exists a positive definite symmetric matrix S_m , the coefficients of which being C^∞ functions of m , such that, if $\xi = \sum \lambda_i \partial_m x_i$, $\eta = \sum \mu_i \partial_m x_i$, then,

$$\langle \xi, \eta \rangle_m = {}^t \lambda S_m \mu.$$

Such a structure allows, among other things, to measure lengths of displacements on the manifold. If $\mu : [0, T] \rightarrow M$ is continuous, piecewise differentiable, its length is defined by

$$L(\mu) = \int_0^T \left\| \frac{d\mu_t}{dt} \right\|_{\mu(t)} dt.$$

In other terms, one defines an infinitesimal length element from the norms on the tangent spaces to M . Similarly, the energy of M is defined by

$$E(\mu) = \int_0^T \left\| \frac{d\mu_t}{dt} \right\|_{\mu(t)}^2 dt.$$

The extremal curves of the energy are called geodesics (one says that a curve is an extremal of a given variational problem if any first order local perturbation of the curve has only second order effects on the functional). In a chart where $\mu_t = (y_t^1, \dots, y_t^n)$, and $S_y = (s_{ij}(y))$ is the matrix which is associated to the inner product, we have

$$\left\| \frac{d\mu_t}{dt} \right\|_{\mu_t}^2 = \sum_{ij} s_{ij}(y(t)) \frac{dy^i}{dt} \frac{dy^j}{dt}.$$

Making a local variation of the kind $y^i \mapsto y^i + h^i$, extremals are characterized by: for all i

$$2 \int_0^T \sum_{lj} \frac{dh^l}{dt} s_{lj}(y(t)) \frac{dy^j}{dt} + \sum_{i,j,l} \frac{dy^i}{dt} \frac{dy^j}{dt} \frac{\partial s_{ij}}{\partial x_l} h^l = 0,$$

which yields, after an integration by parts

$$-2 \int_0^T \sum_{l,j} h^l \frac{d}{dt} \left[s_{lj}(y(t)) \frac{dy^j}{dt} \right] + \sum_{i,j,l} \frac{dy^i}{dt} \frac{dy^j}{dt} \frac{\partial s_{ij}}{\partial x_l} h^l = 0.$$

This relation being true for every h , we have, for all l ,

$$-2 \sum_j s_{lj}(y) \frac{d^2 y^j}{dt^2} - 2 \sum_{ij} \frac{\partial s_{lj}}{\partial x_i} \frac{dy^i}{dt} \frac{dy^j}{dt} + \sum_{i,j} \frac{dy^i}{dt} \frac{dy^j}{dt} \frac{\partial s_{ij}}{\partial x_l} = 0.$$

Denote s^{ij} for the coefficients of S^{-1} . The previous identities write (with a symmetrized second term)

$$-2 \frac{d^2 y^k}{dt^2} = \sum_{ij} \frac{dy^i}{dt} \frac{dy^j}{dt} \sum_l s^{kl} \left(\frac{\partial s_{lj}}{\partial x_i} + \frac{\partial s_{li}}{\partial x_j} - \frac{\partial s_{ij}}{\partial x_l} \right)$$

Denoting

$$(41) \quad \Gamma_{ij}^k = \frac{1}{2} \left[\sum_l s^{kl} \left(\frac{\partial s_{lj}}{\partial x_i} + \frac{\partial s_{li}}{\partial x_j} + \frac{\partial s_{ij}}{\partial x_l} \right) \right],$$

this is

$$(42) \quad \frac{d^2 y^k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dy^i}{dt} \frac{dy^j}{dt} = 0.$$

The coefficients Γ_{ij}^k only depend on the Riemannian metric. They are called the Christoffel's symbols of the manifold at a given point. Therefore, geodesics (expressed in a local chart) are solutions of a second order differential equation. This implies that they are uniquely specified by their value at time, say $t = 0$, and their derivatives $\frac{d\mu_t}{dt}$ at $t = 0$. In particular, one defines the Riemannian exponential at $m \in M$ in the direction $v \in T_m M$ by

$$\text{Exp}_m(tv) = \mu_t$$

where μ_t is the geodesic with $\mu_0 = m$ and $\frac{d\mu_t}{dt} = v$ at time $t = 0$. Such a geodesic exists, as a solution of a differential equation, at least for small times, so that the exponential is well defined at least for small enough t . If this exponential exists at all x for all times, M is said to be a complete manifold.

6.2. Geodesic distance. When M is a Riemannian manifold, one defines the distance between two points m and m' in M by the length of the shortest path which links them, setting

$$d(m, m') = \inf \{ L(\mu) : \mu : [0, 1] \rightarrow M, \\ \mu \text{ continuous, piecewise differentiable, } \mu_0 = m, \mu_1 = m' \}$$

The following theorem is standard, and may be proved as an exercise or read, for example in [6]:

THEOREM 33. *The function d which is defined above is a distance M . Moreover,*

$$d(m, m') = \inf \{ \sqrt{E(\mu)} : \mu : [0, 1] \rightarrow M, \\ \mu \text{ continuous, piecewise differentiable, } \mu_0 = m, \mu_1 = m' \}$$

6.3. Lie group with a left invariant metric. On Lie groups, Riemannian structures are preferably coupled with invariance constraints. As we have seen in considering groups of diffeomorphisms, the suitable way for “moving” within a group is by iterating small steps through the composition rule. For a curve g_t on the group, the length of a portion between g_t and $g_{t+\varepsilon}$ should measure the increment

$g_{t+\varepsilon}g_t^{-1}$. Fix t and let $u_\varepsilon = g_{t+\varepsilon}g_t^{-1}$: one has $u_0 = 0$ and $g_{t+\varepsilon} = u_\varepsilon g_t$. If there is a Riemannian structure on G , the length of the displacement from g_t to $g_{t+\varepsilon}$ is

$$\int_t^{t+\varepsilon} \left| \frac{dg_s}{ds} \right|_{g_s} ds \simeq \varepsilon \left| \frac{d}{d\varepsilon} g_{t+\varepsilon} \right|_{g_t}$$

where the last derivative is taken at $\varepsilon = 0$. The right invariance constraint says that it should in fact be only measured by the increment on the group and therefore be a function of u_ε . But u_ε is itself a curve on G , between id and $g_{t+\varepsilon}g_t^{-1}$ and its length is therefore essentially given by $\varepsilon \left| \frac{du_\varepsilon}{d\varepsilon} \right|_{\varepsilon=0} \Big|_{\text{id}}$. Thus, under the invariance constraint

$$\left| \frac{d}{d\varepsilon} g_{t+\varepsilon} \right|_{g_t} = \left| \frac{du_\varepsilon}{d\varepsilon} \right|_{\text{id}}$$

Introduce the right translation in the Lie group:

$$\begin{aligned} R_g : G &\rightarrow G \\ h &\mapsto hg \end{aligned}$$

so that $g_{t+\varepsilon} = R_{g_t}(u_\varepsilon)$. We have, by the chain rule

$$\left. \frac{dg_{t+\varepsilon}}{d\varepsilon} \right|_{\varepsilon=0} = d_{\text{id}} R_{g_t} \left. \frac{du_\varepsilon}{d\varepsilon} \right|_{\varepsilon=0}$$

This leads to the following definition

DEFINITION 21. *A Riemannian metric on a Lie group G is said to be right invariant if and only if, for all $u \in \mathfrak{g} = T_{\text{id}}G$, for all $g \in G$*

$$|d_{\text{id}} R_g u|_g = |u|_{\text{id}}$$

Thus, the metric on any $T_g G$ may be obtained from the metric on \mathfrak{g} by right translation.

6.4. Application to groups of diffeomorphisms. So far, we have discussed definitions and results which are valid for finite dimensional manifold. In an infinite dimensional situation, eg. for diffeomorphisms, most of the definitions can be extended, and some of the results, but not all, but it is not our intent here to detail these aspects. But it is quite interesting to analyse, at least formally the result of the general construction above when applied to diffeomorphisms.

Let Ω be a bounded open subset of \mathbb{R}^k , and G a group of diffeomorphisms which reduce to the identity on $\partial\Omega$. Since tangent vectors to this group are derivatives of $\frac{d\varphi_t}{dt}$ where $\varphi_t \in G$, one can expect them to vanish on $\partial\Omega$. In particular, \mathfrak{g} is composed with vector fields on Ω which vanish on $\partial\Omega$.

Since the product in G is the usual composition of functions, the right translation is $R_\varphi : \psi \mapsto \psi \circ \varphi$. If ψ_t is a curve on G with $\varphi_0 = 0$ and $\frac{d\psi_t}{dt} = v \in \mathfrak{g}$ at time $t = 0$, we have

$$\frac{d}{dt}(R_\varphi(\psi_t)) = \frac{d}{dt} \psi_t \circ \varphi = v \circ \varphi$$

(at time $t = 0$). But, by the chain rule, this is also given by $d_{\text{id}} R_\varphi \frac{d\psi_t}{dt}$ so that

$$d_{\text{id}} R_\varphi v = v \circ \varphi$$

Thus, a Riemannian metric on G is right invariant if and only if, for all $v \in \mathfrak{g}$, for all $\varphi \in G$,

$$|v \circ \varphi|_\varphi = |v|_{\text{id}}$$

Thus, consider a curve $\varphi : t \mapsto \varphi_t$ on G , defined for $t \in [0, 1]$. Its energy is given, according to the previous sections, by

$$E(\varphi) = \int_0^1 \left| \frac{d\varphi_t}{dt} \right|_{\varphi_t}^2 dt = \int_0^1 \left| \frac{d\varphi_t}{dt} \circ \varphi_t^{-1} \right|_{\text{id}}^2 dt$$

by right invariance. If one lets $v_t = \frac{d\varphi_t}{dt} \circ \varphi_t^{-1}$, we get

$$E(\varphi) = \int_0^1 |v_t|_{\text{id}}^2 dt$$

with $\frac{d\varphi_t}{dt} = v_t \circ \varphi_t$. One retrieves the formulas of chapter 2. In fact, the construction on G_V from an admissible Hilbert space V is a way to rigorously formalize the above formal discussion. In particular, we see that this Hilbert space V may be considered as the ‘‘Lie algebra’’ of G_V ... with the exception that G_V is not necessarily a Lie group (right translation is differentiable on G_V by construction, but left translation may fail to be). Another illustration of the difficulties in making V be a Lie algebra comes from the computation of Lie brackets. The Lie brackets are given by the computation (the derivatives being taken at $s, t = 0$):

$$\begin{aligned} [\xi, \eta] &= \frac{d}{dt} \frac{d}{ds} (\text{id} + t\xi) \circ (\text{id} + s\eta) \circ (\text{id} + t\xi)^{-1} \\ &= \frac{d}{dt} (d(\text{id} + t\xi)\eta \circ (\text{id} + t\xi)^{-1}) \\ &= \frac{d}{dt} (\eta \circ (\text{id} + t\xi)^{-1} + td\xi\eta \circ (\text{id} + t\xi)^{-1}) \\ &= -d\eta\xi + d\xi\eta \end{aligned}$$

since $\frac{d}{dt}(\text{id} + t\xi)^{-1} = -\xi$ at $t = 0$. We therefore obtain $[\xi, \eta] = d\xi\eta - d\eta\xi$. For V to be a Lie algebra, we would need the property $\xi, \eta \in V \Rightarrow [\xi, \eta] \in V$. Assume, only for this discussion, that $\Omega = \mathbb{R}$, so that constant vector fields belong to Ω . In this case, we would have, taking $d\eta = 0$, $\xi \in V \Rightarrow d\xi b \in V$ for any $b \in \mathbb{R}$, since elements of V are continuous, this implies that they are C^1 , and by induction C^∞ . This is quite restrictive, and we do not want to make such an assumption, therefore relaxing the Lie group property.

6.5. Manifolds of deformable objects.

6.5.1. *A landmark manifold.* We now apply the paradigm of section 5.3, in the particular cases of landmark comparison and image comparison, starting with the former. We therefore consider the set $M = \Omega^N$, which contains collections of N landmarks. We let $G = G_V$ for some admissible Hilbert space V , and consider the product space $M^G = G \times M$. The action of G on M is

$$g.(x_1, \dots, x_N) = (g(x_1), \dots, g(x_N))$$

and its action on \mathcal{O} is

$$g.(h, x_1, \dots, x_N) = (h \circ g^{-1}, g(x_1), \dots, g(x_N))$$

The projection of M^G on M is

$$\pi(h, x_1, \dots, x_N) = (h(x_1), \dots, h(x_N))$$

We want to apply proposition 20, and build a left equivariant distance on M^G . This can be done by building an invariant metric on M^G : denote $A_g((h, m)) =$

$(hg^{-1}, g.m)$ for the action of g of (h, m) . A left invariant metric should satisfy, for all $g \in G$, for all $(h, m) \in M^G$, for all $\eta \in T_{(h,m)}M^G$

$$|d_{(h,m)}A_g\eta|_{g.(h,m)} = |\eta|_{(h,m)}$$

and taking $g = h$

$$|\eta|_{(h,m)} = |d_{(h,m)}A_h\eta|_{(\text{id}, h.m)}$$

This implies that it suffices to define the metric at (h, m) when $h = \text{id}$ to know how to define it everywhere.

For landmark deformation, with $m = (x_1, \dots, x_N) \in M$, a tangent vector to M^G at (h, m) takes the form $\eta = (\xi, \alpha_1, \dots, \alpha_N)$ where ξ is a vector field on Ω and $\alpha_i \in \mathbb{R}^k$, $i = 1, \dots, N$. We have

$$d_{(h,m)}A_h\eta = (\xi \circ h^{-1}, d_{x_1}h\alpha_1, \dots, d_{x_N}h\alpha_N)$$

Now, define a norm on $T_{(\text{id}, m)}M^G$ under the form

$$|\xi, \alpha_1, \dots, \alpha_N|_{(\text{id}, m)}^2 = \|\xi\|_V^2 + \frac{1}{\sigma^2} \sum_{i=1}^N |\alpha_i|^2$$

V being, as usual, an admissible Hilbert space. We have

$$|\xi, \alpha_1, \dots, \alpha_N|_{(h,m)}^2 = \|\xi \circ h^{-1}\|_V^2 + \frac{1}{\sigma^2} \sum_{i=1}^N |d_{x_i}h\alpha_i|^2$$

This implies that the energy of a path (h_t, m_t) in M^G is

$$\int_0^1 \left\| \frac{dh_t}{dt} \circ h_t^{-1} \right\|_V^2 dt + \frac{1}{\sigma^2} \sum_{i=1}^N \int_0^1 \left| d_{x_t^i} h_t \frac{dx_t^i}{dt} \right|^2 dt$$

Minimizing path lengths on M^G with fixed endpoints provides a left invariant metric in M^G . Optimizing it over endpoints of the kind (id, m) , $(h, h.m')$ will yield a metric in M , according to theorem 20. We define as before $v_t = \frac{dh_t}{dt} \circ h_t^{-1}$. We also define $y_t^i = h_t(x_t^i)$ so that

$$\frac{dy_t^i}{dt} = \frac{dh_t}{dt}(x_t^i) + d_{x_t^i}h_t \frac{dx_t^i}{dt}$$

and the energy of the path (h_t, m_t) can be written

$$\int_0^1 \|v_t\|_V dt + \frac{1}{\sigma^2} \sum_{i=1}^N \int_0^1 \left| \frac{dy_t^i}{dt} - v_t(y_t^i) \right|^2 dt$$

Note that $(h_t, x_t^1, \dots, x_t^N) \mapsto (v_t, y_t^1, \dots, y_t^N)$ makes a valid change of variable for the fixed endpoint minimization problem, provided the boundary condition becomes $(y_1^1, \dots, y_1^N) = g_{01}^v.m'$. But to compute the distance between m and m' , this must be minimized over targets of the kind $(h, h.m')$, we see that, in fact, the induced distance on M can be computed by minimizing

$$(43) \quad U(v, y) = \int_0^1 \|v_t\|_V dt + \frac{1}{\sigma^2} \sum_{i=1}^N \int_0^1 \left| \frac{dy_t^i}{dt} - v_t(y_t^i) \right|^2 dt$$

over all $v \in \mathcal{X}_V^1(\Omega)$ and all landmark trajectories (y_t^1, \dots, y_t^N) which coincide with m at time $t = 0$ and with m' at time $t = 1$.

This Riemannian analysis of landmark deformations provides, in addition to a distance between landmark configurations, a new matching energy, which can be compared to the one which has been studied in section 3.2. In that case, the matching between m and m' was estimated by minimizing

$$(44) \quad U(v) = \int_0^1 \|v_t\|_V^2 dt + \frac{1}{\sigma^2} \sum_{i=1}^N |x'_i - \varphi_{01}^v(x_i)|^2 dt$$

with respect to $v \in V$. Thus, landmark trajectories are additional unknowns in problem (43). On the other hand, (43) does not directly involve the integrated flow φ^v and may therefore be computed more easily.

Like in section 3.2, the problem can be reduced with the help of the self-reproducing kernel of V . Indeed, computing the differential with respect to V of $U(v, y)$ given by (43) yields

$$\begin{aligned} \frac{dU(v + \varepsilon h, y)}{d\varepsilon} \Big|_{\varepsilon=0} &= 2 \int_0^1 \langle v_t, h_t \rangle_V - 2 \sum_{i=1}^N \int_0^1 \left(\frac{dy_t^i}{dt} - v_t(y_t^i) \right) h_t(y_t^i) \\ &= 2 \int_0^1 \langle v_t, h_t \rangle_V - 2 \sum_{i=1}^N \int_0^1 \left\langle K_{y_t^i}^{\frac{dy_t^i}{dt} - v_t(y_t^i)}, h_t \right\rangle_V \end{aligned}$$

which implies that, at the minimum,

$$v_t(\cdot) = \frac{1}{\sigma^2} \sum_{i=1}^N K(\cdot, y_t^i) \left(\frac{dy_t^i}{dt} - v_t(y_t^i) \right)$$

Thus, introducing the auxilliary unknowns $a_t^i = (\frac{dy_t^i}{dt} - v_t(y_t^i))/\sigma^2$, we have $v_t(\cdot) = \sum_{i=1}^N K(\cdot, y_t^i) a_t^i$ and

$$U(v, y) = \sum_{i,j=1}^N \int_0^1 a_t^j K(y_t^j, y_t^i) a_t^i dt + \frac{1}{\sigma^2} \int_0^1 \sum_{i=1}^N \left| \frac{dy_t^i}{dt} - \sum_{j=1}^N K(y_t^i, y_t^j) a_t^j \right|^2 dt$$

The gradient of this functional with respect to the a trajectories and to the landmark trajectories is quite easy to write down and this computation is left to the reader.

6.5.2. An image manifold. We want to reiterate what we have done for the landmarks in the case of functions. In this case, M is a set of functions $m : \Omega \rightarrow \mathbb{R}$ (or \mathbb{R}^d , but we restrict to scalar functions to simplify) and therefore is infinite dimensional. We still want to consider M as a manifold, which we will do informally, and carry the same analysis as in the previous section. In the present case, a tangent vector to a function $m \in M$ simply is a function $\eta : \Omega \rightarrow \mathbb{R}$, since M is a linear space.

The action of diffeomorphisms on functions is defined by

$$g.m = m \circ g^{-1}$$

and the action of $G = G_V$ on M^G therefore is

$$A_g(h, m) = g.(h, m) = (h \circ g^{-1}, m \circ g^{-1})$$

The differential of this action is straightforward, and given by

$$d_{(h,m)} A_g(\xi, \eta) = (\xi \circ g^{-1}, \eta \circ g^{-1})$$

and a left invariant metric on M^G must satisfy

$$\|\xi, \eta\|_{h,m} = \|d_{(h,m)}A_h(\xi, \eta)\|_{(\text{id}, h.m)} = \|(\xi \circ h^{-1}, \eta \circ h^{-1})\|_{(\text{id}, h.m)}$$

Consider the following norm

$$\|\xi, \eta\|_{(\text{id}, m)}^2 = \|\xi\|_V^2 + \frac{1}{\sigma^2} \|\eta\|_{L^2}^2$$

Using translation invariance, the energy of a curve (h_t, m_t) on M^G is given by

$$E(g_t, m_t) = \int_0^1 \left\| \frac{dg_t}{dt} \circ g_t^{-1} \right\|_V^2 dt + \int_0^1 \left\| \frac{dm_t}{dt} \circ g_t^{-1} \right\|_{L^2}^2 dt$$

To compute the distance between a function m and another function m' , this must be minimized with boundary conditions $m_0 = (\text{id}, m)$ and m_1 in the orbit $\{(g^{-1}, m' \circ g^{-1}), g \in G\}$. Introduce the variable $\mu_t = m_t \circ g_t^{-1}$ so that

$$\frac{dm_t}{dt} = \frac{d\mu_t}{dt} \circ g_t + {}^t\nabla_{g_t\mu_t} v_t \circ g_t.$$

with as usual $\frac{dg_t}{dt} = v_t \circ g_t$. One can write

$$E(g_t, m_t) = \int_0^1 \|v_t\|_V^2 dt + \int_0^1 \left\| \frac{d\mu_t}{dt} - {}^t\nabla_{\mu_t} v_t \right\|_{L^2}^2 dt$$

and minimizing this expression with respect to v and μ , with boundary conditions $\mu_0 = m$ and $\mu_1 = m'$ is equivalent to the initial problem.

7. Gradient descent on Riemannian manifolds

7.1. Gradient of a functional.

DEFINITION 22. *Let M be a Riemannian manifold. The gradient of a differentiable function $U : M \rightarrow \mathbb{R}$ is the vector field denoted*

$$\begin{aligned} \nabla^M u : M &\rightarrow TM \\ m &\mapsto \nabla_m^M U \end{aligned}$$

such that, for all vector field X on M , and $m \in M$

$$(45) \quad XU = \langle X, \nabla^M U \rangle$$

An alternative definition is: for all curve $\mu : t \mapsto \mu_t$ on M ,

$$(46) \quad \frac{dU(\mu_t)}{dt} = \left\langle \frac{d\mu_t}{dt}, \nabla_{\mu_t}^M U \right\rangle_{\mu_t}$$

Gradient descent on M is defined by the following proposition:

PROPOSITION 21. *Let U be a differentiable function on M and γ a positive real number. If μ is a solution of the ordinary differential equation*

$$(47) \quad \frac{dU}{dt} = -\gamma {}^t\nabla_{\mu_t}^M U$$

on some interval $[0, t]$, then $t \mapsto U(\mu_t)$ decreases on $[0, T]$.

Indeed, we have, by definition of the gradient

$$\frac{dU(\mu_t)}{dt} = \left\langle \frac{d\mu_t}{dt}, \nabla_{\mu_t}^M U \right\rangle_{\mu_t} = -\gamma |\nabla_{\mu_t}^M U|_{\mu_t}^2 \leq 0$$

These straightforward definition and proposition lead to quite interesting algorithms when applied to practical situations. We here consider the case of Lie groups.

7.2. Gradient descent on Lie groups with a right invariant metric.

In this section, the Riemannian manifold we consider is a Lie group G with a right invariant metric. Consider in this framework a differentiable function μ and the function $f_g : v \mapsto U(\exp(v)g)$ for $g \in G$ and $v \in \mathfrak{g}$. Since f_g is defined on the Euclidean (or Hilbert) vector space \mathfrak{g} , it has a gradient in the usual sense which will be denoted ∇f_g . One has

$$\left. \frac{df(tv)}{dt} \right|_{t=0} = \langle \nabla_0 f_g, v \rangle_{\text{id}} = \langle \nabla_g^G, d_{\text{id}} R_g v \rangle_g = \langle (d_{\text{id}} R_g)^{-1} \nabla_g^G U, v \rangle_{\text{id}}$$

so that $(d_{\text{id}} R_g) \nabla_0 f_g = \nabla_g^G U$. This implies that the gradient descent algorithm for U can be written under the form

$$\frac{dg_t}{dt} = -\gamma (d_{\text{id}} R_{g_t}) \nabla_0 f_{g_t}$$

so that we are able to perform gradient descent on G from the computation of a gradient of a function defined on a Lie algebra (which is generally easier) followed by a right translation.

In the case of diffeomorphisms ($G = G_V$), this gradient descent provides algorithms we already have seen. Indeed, in this case, we have seen that right invariant vector field have the form $X_g = v \circ g$ with $v \in V$, and the exponential $\exp(tv)$ is the solution of the equation $\frac{dg_t}{dt} = v \circ g_t$ with initial condition $g_0 = \text{id}$, which has precisely been called φ_t^v . Thus in this case, $f_g(tv) = U(\varphi_t^v \circ g)$ and the identity $\left. \frac{df(tv)}{dt} \right|_{t=0} = \langle \nabla_0 f_g, v \rangle_{\text{id}}$ is precisely the definition we have given for the Eulerian gradient $\tilde{\nabla}_g^V U$. The gradient descent now writes

$$\frac{dg_t}{dt} = -\gamma \tilde{\nabla}_g^V U \circ g$$

which is exactly the equation we have written in (27). Thus, we have an the greedy algorithms we have considered in chapter 3 as a gradient descent in the group of diffeomorphisms equipped with a right invariant metric.

Bibliography

- [1] Y Amit and P Piccioni. A non-homogeneous markov process for the estimation of gaussian random fields with non-linear observations. *Ann. of Proba.*, 19:1664–1678, 1991.
- [2] R Bajcsy and C Broit. Matching of deformed images. In *The 6th international conference in pattern recognition*, pages 351–353, 1982.
- [3] R Bajcsy and S Kovacic. Multiresolution elastic matching. *Comp. Vision, Graphics, and image proc.*, 46:1–21, 1989.
- [4] F Beg, M. *Variational and Computational methods for Flows of Diffeomorphisms in Image Matching and Growth in Computational Anatomy*. PhD thesis, Johns Hopkins University, 2003.
- [5] F Beg, M, I Miller, M, A Trouvé, and L Younes. Computing large deformation metrics via geodesic flows of diffeomorphisms. Technical report, University Johns Hopkins, 2002. To appear in *International Journal of Computer Vision*.
- [6] M. P. Do Carmo. *Riemannian geometry*. Birkhäuser., 1992.
- [7] E Christensen, G, D Rabbitt, R, and I Miller, M. Deformable templates using large deformation kinematics. *IEEE trans. Image Proc.*, 1996.
- [8] R Courant and D Hilbert. *Methods of Mathematical Physics, I*. J. Wiley and Sons, 1955. Original edition: 1937.
- [9] J. Dieudonné. *Calcul infinitésimal (seconde édition)*. Hermann, 1980. English translation: Jean Dieudonné, Infinitesimal Calculus.
- [10] R Downie, T, L Shepstone, and W Silverman, B. a wavelet approach to deformable templates. In V Mardia, K, A Gill, C, and L Dryden, I, editors, *Image fusion and shape variability techniques*, pages 163–169. Leeds University Press, 1996.
- [11] C Gee, J and D Peralta, P. Continuum models for bayesian image matching. In H Hanson, K and N Silver, R, editors, *Maximum entropy and bayesian methods*. Kluwer academics, 1995.
- [12] J Glaunès, M Vaillant, and I Miller, M. Landmark matching via large deformation diffeomorphisms on the sphere. *Journal of Mathematical Imaging and Vision, MIA 2002 special issue*, (to appear) 2003.
- [13] S. Helgason. *Differential Geometry, Lie groups and Symmetric spaces*. Academic Press, 1978.
- [14] G Hermosillo. *Variational methods for multimodal matching*. PhD thesis, Universit de Nice, 2002.
- [15] S Joshi. *Large deformation diffeomorphisms and Gaussian random fields for statistical characterization of brain sub-manifolds*. PhD thesis, Sever institute of technology, Washington University, 1997.
- [16] S Joshi and M Miller. Landmark matching via large deformation diffeomorphisms. *IEEE transactions in image processing*, 9(8):1357–1370, 2000.
- [17] G Luenberger, D. *Optimization by Vector Space Methods*. J. Wiley and Sons, 1969.
- [18] J Nocedal and J Wright, S. *Numerical Optimization*. Springer, 1999.
- [19] D Rabbitt, R, A Weiss, J, E Christensen, G, and I Miller, M. Mapping of hyperelastic deformable templates using the finite element method. In *Proceeding of San Diego's SPIE conference*, 1995.
- [20] F Riesz and S Nagy, B. *Functional analysis*. Ungar Pub Co, 1955.
- [21] Alain Trouv. Diffeomorphism groups and pattern matching in image analysis. *Int. J. of Comp. Vis.*, 28(3):213–221, 1998.
- [22] E Zeidler. *Applied Functional Analysis. Applications to mathematical physics*. Springer, 1995.