ECOLE NORMALE SUPERIEURE

An Approach of Pattern Recognition through Infinite Dimensional Group Action

Alain TROUVÉ

LMENS - 95 - 9

Département de Mathématiques et Informatique

CNRS URA 762

An Approach of Pattern Recognition through Infinite Dimensional Group Action

Alain TROUVÉ

LMENS - 95 - 9

May 1995

Laboratoire de Mathématiques de l'Ecole Normale Supérieure
45 rue d'Ulm 75230 PARIS Cedex 05
Laboratoire d'Analyse Numérique CNRS URA 189
Université Paris 6

Tel : (33)(1) 44 32 00 00
Adresse électronique : .. @dmi.ens.fr

AN APPROACH OF PATTERN RECOGNITION THROUGH INFINITE DIMENSIONAL GROUP ACTION

ALAIN TROUVÉ

May 4, 1995

ABSTRACT. Non rigid deformations of patterns can be interpreted as the action of an infinite dimensional group $\mathcal{A}(n)$ on a given set $\mathcal{P}$ of patterns. Following Lie group ideas, a small deformation can be well described by an element y of the tangent space at identity $T_e\mathcal{A}(n)$. Given a metric n on $T_e\mathcal{A}(n)$, which brings the cost of a small deformation, we show that we can define on $\mathcal{A}(n)$ a left invariant distance $d_{\mathcal{A}(n)}$ which gives the distance between two arbitrary large deformations. This allows to reformulate in a unified framework many pattern recognition tasks. Finally, we propose a sub-optimal algorithm to solve three important classes of pattern recognition problems through a gradient algorithm on $\mathcal{A}(n)$ whose convergence is rigorously established.

CONTENTS

1. Introduction **2**

2. The abstract construction of $A(n)$ **7**

2.1. Control lemmas on A

2.2. Definition of $A(n)$

3. A weak differentiable structure on $A(n)$ **16**

3.1. Tangent spaces

3.2. Exponential

3.3. Differentiable applications

3.4. Important examples

The author would like to thank professor Robert Azencott for many helpful discussions. His beautiful ideas on the application of the Lie group theory in pattern recognition are one of the main sources of the present paper

4. Hilbert sub-groups **20**
4.1. Integration of vector fields on $A(n)$
4.2. Proof of the theorem 4.6
4.3. Convergence of the Cauchy approximates
4.4. Application to gradient algorithms
5. Application to pattern recognition **26**
5.1. Sub-optimal solutions
5.2. Examples
5.3. Choices of n
References **31**

1. Introduction

During the last decade, the use of deformable models in pattern recognition and in pattern matching have became more and more usual. However, the theoretical framework of this approach is not yet fixed and much of the mathematical work is still to be done. At a formal level, the problem can be formulated as follows. Assume that you have a set of "patterns" $\mathcal{P}$ and a set of "actions" or "deformations" or also "transformations" $\mathcal{A}$ such that for each a in $\mathcal{A}$ and each $f \in \mathcal{P}$ we can define the deformation of f by a as a new element of $\mathcal{P}$ denoted af. A natural assumption at this level is that $\mathcal{A}$ has a group structure so that we can define the product of two elements a and $a' \in \mathcal{A}$ denoted aa' and the inverse denoted a^{-1} of an element $a \in \mathcal{A}$. We should also assume quite naturally that $\mathcal{A}$ is acting on $\mathcal{P}$ i.e. $a(a'f) = (aa')f$ and $ef = f$ where e is the identity element of $\mathcal{A}$. Such a situation is quite common, and as a first example, we can choose for $\mathcal{P}$ the subsets of $\mathbb{R}^2$ (binary shapes) and for $\mathcal{A}$ the group of the isometries of $\mathbb{R}^2$ or the group of the affine transformations. Actually, one needs often a larger group allowing non rigid local transformations, to cover for instance the huge variability of biological shape [4].

An important issue is to define an appropriate distance $d_{\mathcal{A}}$ on $\mathcal{A}$ inducing a distance $d_{\mathcal{P}}$ between the elements of $\mathcal{P}$ by

$$d_{\mathcal{P}}(f_1, f_2) = \inf\{\ d_{\mathcal{A}}(e, a) \mid af_1 = f_2\ \}. \tag{1}$$

This approach is quite natural since the distance between two patterns should be a measure of the amount of "deformation" needed to go from f_1 to f_2. However, the symmetry property of the distance $d_{\mathcal{P}}$ $(d_{\mathcal{P}}(f_1, f_2) = d_{\mathcal{P}}(f_2, f_1))$ is fulfilled if $d_{\mathcal{A}}(e, a) = d_{\mathcal{A}}(e, a^{-1})$ so that a property of left invariance $d_{\mathcal{A}}(aa_1, aa_2) = d_{\mathcal{A}}(a_1, a_2))$ or right invariance $(d_{\mathcal{A}}(a_1 a, a_2 a) = d_{\mathcal{A}}(a_1, a_2))$ of the distance $d_{\mathcal{A}}$ is natural and attractive. Moreover, if $d_{\mathcal{A}}$ is left invariant, then $(a, b) \rightarrow d_{\mathcal{A}}(a^{-1}, b^{-1})$ is right invariant so that we can focus on left invariant distances. If $\mathcal{A}$ is a finite dimensional Lie group of transformations, then choosing a metric on the Lie algebra $T_e\mathcal{A}$ and extending this metric on $\mathcal{A}$ by left multiplication, we can define the associated geodesic distance which is left invariant. In the framework of deformable model, this choice of a metric on $T_e\mathcal{A}$ corresponds to our a priori on the cost of the small deformations and should be designed according to our precise application. This approach, applied on infinite dimensional groups of transformations, has been suggested by R. Azencott in [2] for shapes recognition.

However, for infinite dimensional group of transformations, two important difficulties arise. The first one is to define a Lie group structure on a infinite dimensional group of transformations and the second one is to equip this Lie group with a left invariant metric for which we can define the associated geodesic distance [6, 10]. There is in fact a third difficulty which is that we should keep in mind that we want at the very end to have an appropriate numerical scheme to solve various pattern recognition problems in this framework.

In this paper, we will assume that $\mathcal{P}$ is the set of the measurable functions from a finite dimensional Riemannian compact manifold M without boundary to a finite dimensional manifold X. This definition matches numerous examples. For instance, the closed curves in $\mathbb{R}^2$ correspond to the case $M = \mathbb{R}/\mathbb{Z}$ (1 dimensional torus) and $X = \mathbb{R}^2$ (see [12, 11]). The case of the periodic grey-level images correspond to $M = \mathbb{R}^2/\mathbb{Z}^2$ (2 dimensional torus) and $X = \mathbb{R}$. In both last case, X is a vector space. However, more general situations arise if we work with images with bounded grey-level values ($X = \mathbb{R}^+$ or $X = [a, b]$) or with images where $f(m)$ is an unitary vector in $\mathbb{R}^3$ i.e. $X = S^2$ representing for instance the direction in m of a some physical field. The fact that M is without boundary is a bit restrictive. However, this restriction, only technical, could certainly be relaxed and simplifies the results we will prove in this paper.

As group of deformations $\mathcal{A}$, a natural choice is to consider the group $\mathrm{Hom}(M)$ of the homeomorphism on M with the action $\phi f = f \circ \phi$ for $f \in \mathcal{P}$ and $\phi \in \mathrm{Hom}(M)$. However, this choice may be too restrictive in some situations since the transformation does not affect the range of f. For instance, in the case of closed curves in $\mathbb{R}^2$, $f \circ \phi$ is only a change of variable which does not affect the geometric shape of the curve. Hence, in order to modify the range of f, assume that there exists a finite dimensional Lie group G acting on X by the action $(g, x) \rightarrow gx$. In the case of closed curves, one can choose $G = \mathbb{R}^2$ with the action $(g, x) \rightarrow g + x$. In the case of positive images $X = \mathbb{R}^+$, we can choose as G the multiplicative group $\mathbb{R}^+_*$ where gx denotes here the multiplication in $\mathbb{R}$. The case $X = S^2$ of images of unitary vector in $\mathbb{R}^3$ is more interesting since we can choose for G various groups of matrices acting on S^2. Now, if $C(M, G)$ is the set of the continuous mapping from M to G on which we have the group product given by the pointwise multiplication in G (i.e. $hh'(m) = h(m)h'(m)$), $C(M, G)$ acts on $\mathcal{P}$ through the action $(h, f) \rightarrow hf$ where hf is defined by

$$hf(m) = h(m)f(m) \ ; \ m \in M.$$

Now, putting together both previous actions, we will assume that $\mathcal{A} = \text{Hom}(M) \times C(M, G)$, and for any $a = (\phi, h) \in \mathcal{A}$ and $f \in \mathcal{P}$, the action of a on f will be defined by

$$af = h(f \circ \phi),$$

where on $\mathcal{A}$ we have the group product

$$aa' = (\phi' \circ \phi, h(h' \circ \phi)),$$

with $a = (\phi, h)$ and $a' = (\phi', h')$. At this point, we see that $\mathcal{A}$ is not a Lie group, even infinite dimensional. We could have chosen for $\mathcal{A}$ the space $\text{Diff}^\infty(M) \times C^\infty(M, G)$ where $\text{Diff}^\infty(M)$ is the set of smooth diffeomorphisms on M and $C^\infty(M, G)$ the set of smooth mappings from M to G so that we could have seen $\mathcal{A}$ as an ILB (inverse limit of Banach space) Lie group as defined in [8] with Lie algebra $\mathfrak{X}(M) \times C^\infty(M, \mathfrak{G})$ where $\mathfrak{X}(M)$ is the set of the smooth vector fields on M and $C^\infty(M, \mathfrak{G})$ the set of the smooth mappings from M to the Lie algebra $\mathfrak{G}$ of G. However, we do not want to restrict ourself to smooth transformations and we prefer to start with $\mathcal{A} = \text{Hom}(M) \times C(M, G)$, which is in a sense the largest possible group of transformations we can need, and to look for sub-groups of $\mathcal{A}$. On the other side, we will consider the space $\mathfrak{X}(M) \times C^\infty(M, \mathfrak{G})$ as the formal Lie algebra of $\mathcal{A}$ denoted $\widetilde{T}_e\mathcal{A}$. This space could seem to be too small to be the tangent space of $\mathcal{A}$ at identity, but our point of view is that it is, in a sense, the largest tangent space which will be included in all the tangent spaces at identity of the sub-groups we will consider. Now, start from a norm n on $\mathfrak{X}(M) \times C^\infty(M, \mathfrak{G})$. The principle of our construction will be, for any sufficiently smooth path $t \to Y_t$ from $[0,1]$ to $\mathfrak{X}(M) \times C^\infty(M\mathfrak{G})$, to define the integrated path $t \to a_t$ in $\mathcal{A}$ as the solution of the equation formally expressed as

$$\frac{da}{dt} = a_t Y_t \tag{2}$$

where ay denotes the left translation of $y \in \mathfrak{X}(M) \times C^\infty(M, \mathfrak{G})$ by the formal differential of the left multiplication by a. Now, considering paths with finite length according to n i.e. $\int_0^1 n(Y_s)ds < \infty$, we will define the sub-group of transformations $\mathcal{A}(n)$ as the ending points of all the integrated paths of finite length. Hence, formally,

$$\mathcal{A}(n) = \inf\{ \ a_1 \mid \int_0^1 n(a_s^{-1}\frac{da}{ds})ds < \infty, \ a_0 = e \ \}. \tag{3}$$

One should say here that the previous point of view is nothing but a generalisation to infinite dimensional Lie algebra of the construction proposed by R. Palais in [9] for the construction of finite dimensional Lie group of transformations from a finite dimensional Lie sub-algebra $\mathfrak{h}$ of $\mathfrak{X}(M) \times C^\infty(M, \mathfrak{G})$. Here, the condition $a_t^{-1}\frac{da}{dt} \in \mathfrak{h}$ is replaced by the condition of finite length with respect to n. In our case, the sub-group is parametrized by the choice of the norm n and the final sub-group is infinite dimensional.

Then, we will define on $\mathcal{A}(n)$ the associated geodesic distance

$$d_{\mathcal{A}(n)}(e, a) = \inf\{ \int_0^1 n(a_s^{-1}\frac{da}{ds})ds \mid a_0 = e, \ a_1 = a \ \}. \tag{4}$$

In part 2, we will precise the conditions on which we could give a rigorous meaning to (2) and we will deduce conditions on the norm n such that paths of finite length for n in the tangent space at identity can be integrated in $\mathcal{A}$. We will then give a precise definition of $\mathcal{A}(n)$ and of the distance of $d_{\mathcal{A}(n)}$ and we will prove in theorem 2.15 that $(\mathcal{A}(n), d_{\mathcal{A}(n)})$ is a complete metric space.

In part 3, we will show that even if $\mathcal{A}(n)$ has not a Lie group structure nor a differentiable structure in the usual sense, $\mathcal{A}(n)$ can be equipped with a weak differentiable structure modeled over a Banach space, but strong enough to define the exponential mapping and a useful notion of differentiable real valued functions.

In part 4, we restrict ourself to the important case when n is an hilbertian norm. In this case, the sub-group $\mathcal{A}(n)$ has a weak differentiable modeled over an Hilbert space. We show then that given a differentiable real valued function E, one can define the gradient ∇E and under some additional conditions on E, for every initial deformation a_0, the solution of the formal gradient evolution equation

$$\frac{da}{dt} = -\nabla_{a_t} E. \tag{5}$$

In part 5, we apply this result to three important problems of pattern recognition:

Template fitting: Assume that we single out a pattern $f \in \mathcal{P}$ called the template pattern. Let $L : X \to \mathbb{R}^+$ be a non negative function called the penalty function. The problem of template fitting for the penalty function L is to find $a \in \mathcal{A}(n)$ minimizing

$$\int_M L(af)d\mu + d_{\mathcal{A}(n)}(e,a)^2 \tag{6}$$

where e denotes the identity element in $\mathcal{A}$, $d_{\mathcal{A}(n)}$ the left invariant distance on $\mathcal{A}(n)$ and μ is the normalized Riemannian measure on M. If $\hat{a}$ is a solution of the problem (we do not discuss for the moment the existence of such a solution), then $\hat{f} = \hat{a}f$ will be called the fit of f according to L.

Patterns classification: Assume here that we single out $f_1, \cdots, f_p$, p patterns in $\mathcal{P}$ as template patterns. Now consider a new pattern $\tilde{f} \in \mathcal{P}$. We define the similarity of $\tilde{f}$ with f_i by

$$S(\tilde{f}, f_i) \quad = \quad \inf_{a \in \mathcal{A}(n)} \int_M L(\tilde{f}(m), (af)(m))d\mu + d_{\mathcal{A}(n)}(e,a)^2. \tag{7}$$

The value of $L(x, x')$ is usually a kind of distance between x and x' which controls the similarity between points of X. Now, considering the values of $S(\tilde{f}, f_i)$ for all $i \in \{1, \cdots, p\}$, we can classify $\tilde{f}$ into one of the classes defined by the f_i's.

Pattern matching: Keeping the notation introduced for the classification problem, we denote $\hat{a}_i$ the element of $\mathcal{A}(n)$ (if it exists) achieving the minimal value of $S(\tilde{f}, f_i)$ for i minimizing the values of the $S(\tilde{f}, f_j)$'s. Indeed, if $\hat{\phi}_i$ is the homeomorphism corresponding to $\hat{a}_i$, $\hat{\phi}_i$ is a mapping from the points $(m, \tilde{f}(m))$ of the new pattern to the points $(\hat{\phi}_i(m), f \circ \hat{\phi}_i(m))$ of the template f_i.

We will propose in this last part a sub-optimal solution to these three problems based on a gradient algorithm in $\mathcal{A}(n)$ for the function $E(a) = \int_M L(af)d\mu$ in the case of template fitting and $E(a) = \int_M L(\tilde{f}, af)d\mu$ in the case of pattern classification and pattern matching. This sub-optimal solution can be achieved numerically in various situations as it will be shown in a forthcoming paper.

2. The abstract construction of $\mathcal{A}(n)$

Let us recall briefly the framework and the notations. Let M and X be two finite dimensional manifolds. We assume that M is compact, connected and Riemannian. We denote $\mathcal{P}$ the set of all the measurable functions from M to X. The set $\mathcal{P}$ will be called the space of patterns. Now, let G be a finite dimensional connected Lie group and $\mathfrak{G}$ be its Lie algebra. We assume that G acts on X and the action of $g \in G$ on $x \in X$ will be denoted gx. Now, let $C(M,G)$ be the set of all the continuous functions from M to G and $\mathrm{Hom}(M)$ be the set of all the homeomorphisms on M. One can define on $C(M,G)$ a group product $(h,h')\rightarrow hh'$ for all h and h' in $C(M,G)$ where $h''=hh'$ is defined by $h''(m)=h(m)h'(m)$ and gg' denotes the product on G. Moreover, $C(M,G)$ acts on $\mathcal{P}$ through the following action

$$(hf)(m) \;=\; h(m)f(m) \quad ;\; m\in M,\; h\in C(M,G),\; f\in\mathcal{P}. \tag{8}$$

Consider the set $\mathcal{A}=\mathrm{Hom}(M)\times C(M,G)$ which will be called the set of the actions. For each element $a\in\mathcal{A}$ we will denote ϕ its component on $\mathrm{Hom}(M)$ and h its component on $C(M,G)$. We can define on $\mathcal{A}$ the product $(a,a')\rightarrow aa'$ for all $a=(\phi,h)$ and $a'=(\phi',h')$ by

$$aa' \;=\; (\phi'\circ\phi, h(h'\circ\phi)) \tag{9}$$

where $\circ$ denotes the composition of functions. One verifies easily that for this product, $\mathcal{A}$ is a group acting on $\mathcal{P}$ through the following action

$$af \;=\; h(f\circ\phi), \tag{10}$$

where $a=(\phi,h)$.

Let $\langle\ ,\ \rangle_m^M$ denotes the metric at point $m\in M$ and ∇^M denotes the Riemannian connection on M. Let 1_G denotes the identity element in G and let $\langle\ ,\ \rangle_{1_G}^G$ be a scalar product on $\mathfrak{G}$ (people unfamiliar with Riemannian geometry and Lie group theory could usefully refer to [3] and [5]). We extend this scalar product on each tangent space T_gG through left multiplication so that G becomes a Riemannian manifold. We denote ∇^G the Riemannian connection on G.

Definition 2.1. (1) For all $u \in \mathfrak{X}(M)$, we define

$$|u|_\infty = \sup\{ \ (\langle u(m),\, u(m)\rangle_m^M)^{1/2} \mid m \in M \ \},$$

$$|\nabla u|_\infty = \sup\{ \ |\nabla_v^M u|_\infty \mid v \in \mathfrak{X}(M),\ |v|_\infty = 1 \ \}.$$

(2) For all $z \in C^\infty(M, \mathfrak{G})$, we define

$$|z|_\infty = \sup\{ \ (\langle z(m),\, z(m)\rangle_{1_G}^G)^{1/2} \mid m \in M \ \},$$

$$|\nabla z|_\infty = \sup\{ \ |\nabla_v^{\mathfrak{G}} z|_\infty \mid v \in \mathfrak{X}(M),\ |v|_\infty = 1 \ \},$$

where $z' = \nabla_v^{\mathfrak{G}} z$ is defined by $z'(m) = d_m z(v(m))$ and $d_m z$ is the differential of z at $m \in M$.

Throughout this work, for all $Y \in C^\infty([0,1] \times M, TM \times \mathfrak{G})$, we will denotes U its component on TM and Z its component on $\mathfrak{G}$ so that $U \in C^\infty([0,1] \times M, TM)$ and $Z \in C^\infty([0,1] \times M, \mathfrak{G})$. As usual, for $t \in [0,1]$, Y_t denotes the function in $C^\infty(M, TM \times \mathfrak{G})$ defined by $Y_t(m) = Y(t,m)$ for any $m \in M$. Moreover, π denotes the canonical projection of $TM \times \mathfrak{G}$ on M.

Definition 2.2. Let $\mathcal{T}_c^\infty$ be defined by

$$\mathcal{T}_c^\infty = \{ \ Y \in C^\infty([0,1] \times M, TM \times \mathfrak{G}) \mid \pi \circ Y_t = \mathrm{Id}_M \ \forall t \in [0,1] \text{ and } Y_t \text{ is vanishing outside a compact set of }]0,1[\ \}. \quad (11)$$

To avoid ambiguities, let us say that Y_t is vanishing if $Y_t(m) = (0_m, 0_{\mathfrak{G}})$ for all $m \in M$ where 0_m (resp. $0_{\mathfrak{G}}$) denotes the null vector in $T_m M$ (resp. in $\mathfrak{G}$).

Let e be the identity element of $\mathcal{A}$. This element is defined by $e(m) = (m, 1_G)$ for all $m \in M$. As said before in the introduction, we consider formally the space $\mathfrak{X}(M) \times C^\infty(M, \mathfrak{G})$ as the tangent space at e of $\mathcal{A}$ so that this space will be denoted $\widetilde{T}_e\mathcal{A}$. The tilded notation recalls that $\widetilde{T}_e\mathcal{A}$ is not the tangent space as usually defined on Lie groups. Throughout this work, y will usually denote an element of $\widetilde{T}_e\mathcal{A}$ and u (resp. z) its component on $\mathfrak{X}(M)$, (resp. its component on $C^\infty(M, \mathfrak{G})$). In order to define the formal tangent space $\widetilde{T}_a\mathcal{A}$ for every $a \in \mathcal{A}$, we consider the left multiplication L_a on $\mathcal{A}$ defined by $L_a(a') = aa'$. We can easily compute formally its differential $\widetilde{d}_e(L_a)$

at e given for any $y = (u, z) \in \tilde{T}_e\mathcal{A}$ by

$$\tilde{d}_e(L_a)(y)(m) = \left(u \circ \phi(m), d_{1_G}(L_{h(m)})((z \circ \phi)(m))\right), \tag{12}$$

where $a = (\phi, h)$ and $d_{1_G}(L_g)$ denotes the usual differential at 1_G of the left multiplication by $g \in G$ on G. In order the simplify the formulation, we will use the notation ay to denote $\tilde{d}_e(L_a)(y)$ and $h(z \circ \phi)$ to denote $m \rightarrow d_{1_G}(L_{h(m)})((z \circ \phi)(m))$ so that we get the new definition

$$ay = (u \circ \phi, h(z \circ \phi)). \tag{13}$$

Hence, we define

$$\tilde{T}_a\mathcal{A} = \{\ ay \mid y \in \tilde{T}_e\mathcal{A}\ \}. \tag{14}$$

Let $C([0,1] \times M, M \times G)$ (resp. $C^\infty([0,1] \times M, M \times G)$) be the set of the continuous (resp. smooth) functions from $[0,1] \times M$ to $M \times G$. For any $A \in C([0,1] \times M, M \times G)$, we denote Φ its component on M and H its component on G so that $\Phi \in C([0,1] \times M, M)$ and $H \in C([0,1] \times M, G)$. From the classical theory of O.D.E. on smooth manifolds, we deduce that for all $Y \in \mathcal{T}_c^\infty$, there exists $A \in C^\infty([0,1] \times M, M \times G)$ such that $A_0 = e$ and

$$\frac{\partial A}{\partial t} = A_t Y_t, \tag{15}$$

that is $\Phi_0 = \mathrm{Id}_M$, $H_0 \equiv 1_G$ and

$$\begin{cases} \frac{\partial \Phi}{\partial t}(t,m) = U(t, \Phi(t,m)) \\ \frac{\partial H}{\partial t}(t,m) = H(t,m)Z(t, \Phi(t,m)), \end{cases} \tag{16}$$

where $Y = (U, Z)$ and $A = (\Phi, H)$. We will now establish control lemmas on A as a function of Y in order to define the solution of (15) on a larger set of Y.

2.1. Control lemmas on A.

Definition 2.3. Let d_M and d_G denote the distances associated with the Riemannian structure on M and G. We define the distance $d_\mathcal{A}$ on $\mathcal{A}$ by

$$d_\mathcal{A}(a, a') = \sup_{m \in M} d_M(\phi(m), \phi'(m)) + \sup_{m \in M} d_G(h(m), h'(m)), \tag{17}$$

where $a = (\phi, h)$ and $a' = (\phi', h')$. Moreover, we define on $C([0,1] \times M, M \times G)$ the distance D defined by

$$D(A, A') = \sup_{s \leq 1,\ m \in M} d_M(\Phi_s(m), \Phi'_s(m)) + \sup_{s \leq 1,\ m \in M} d_G(H_s(m), H'_s(m)), \quad (18)$$

where $A = (\Phi, H)$ and $A' = (\Phi', H')$.

From the completeness of M and G (in the topological sense), we deduce that $C([0,1] \times M, M \times G)$ is a complete metric space for the distance D.

Lemma 2.4. *Let $Y \in \mathcal{T}_c^\infty$ and denote $A = (\Phi, H)$ the solution of (15). For all m, m' in M and all $t \in [0,1]$, we have*

$$d_M(\Phi_t(m), \Phi_t(m')) \leq d_M(m, m') \exp(\int_0^t |\nabla U_s|_\infty ds), \quad (19)$$

$$d_G(H_t(m), H_t(m')) \leq d_M(m, m') \int_0^t |\nabla Z_s|_\infty \exp(\int_0^s |\nabla U_{s'}|_\infty ds') ds. \quad (20)$$

Proof. Let $p \in C^\infty([0,1], M)$ be a smooth path such that $p(0) = m$ and $p(1) = m'$. Let $\widetilde{A} = (\widetilde{\Phi}, \widetilde{H}) \in C^\infty([0,1] \times [0,1], M \times G)$ be defined by

$$\widetilde{A}(t, s) = A(t, p(s)). \quad (21)$$

Using covariant derivatives, we get

$$\frac{D}{dt}\left(\frac{\partial \widetilde{\Phi}}{\partial s}\right) = \frac{D}{ds}\left(\frac{\partial \widetilde{\Phi}}{\partial t}\right) = \frac{D}{ds}\Big(U(t, \widetilde{\Phi}(t, s))\Big) = \nabla_{\frac{\partial \widetilde{\Phi}}{\partial s}} U(t, \widetilde{\Phi}(t, s)). \quad (22)$$

Hence, if $r = \langle \frac{\partial \widetilde{\Phi}}{\partial s}, \frac{\partial \widetilde{\Phi}}{\partial s} \rangle_{\widetilde{\Phi}}^M$, we get

$$\frac{\partial r}{\partial t} = 2 \langle \frac{D}{dt}\left(\frac{\partial \widetilde{\Phi}}{\partial s}\right), \frac{\partial \widetilde{\Phi}}{\partial s} \rangle_{\widetilde{\Phi}}^M \leq 2 |\nabla U_t|_\infty r, \quad (23)$$

so that applying the Gronwall's lemma we get

$$r \leq r_0 \exp(2 \int_0^t |\nabla U_s|_\infty ds). \quad (24)$$

Thus, we have finally

$$|\frac{\partial \widetilde{\Phi}}{\partial s}| \leq |\frac{dp}{ds}| \exp(\int_0^t |\nabla U_u|_\infty du), \quad (25)$$

and

$$d_M(\Phi(t, m), \Phi(t, m')) = d_M(\widetilde{\Phi}(t, 0), \widetilde{\Phi}(t, 1)) \leq e^{\int_0^t |\nabla U_u|_\infty du} \int_0^1 |\frac{dp}{ds}| ds. \quad (26)$$

Considering $\frac{\partial \widetilde{H}}{\partial s}$, we get using covariant derivatives and the left invariance of the metric on G

$$\frac{D}{dt}\left(\frac{\partial \widetilde{H}}{\partial s}\right) = \frac{D}{ds}\left(\frac{\partial \widetilde{H}}{\partial t}\right) = \widetilde{H}\nabla_{\frac{\partial \widetilde{\Phi}}{\partial s}} Z(t, \widetilde{\Phi}(s,t)). \tag{27}$$

Now, let $r_t = \langle \frac{\partial \widetilde{H}}{\partial s}, \frac{\partial \widetilde{H}}{\partial s}\rangle^G$, we get from (27)

$$\frac{\partial r}{\partial t} \leq 2r^{1/2}|\nabla Z_t|_\infty |\frac{\partial \widetilde{\Phi}}{\partial s}|. \tag{28}$$

Since $r_0 = 0$, we deduce that

$$\sqrt{r_t} \leq \int_0^t |\nabla Z_u|_\infty |\frac{\partial \widetilde{\Phi}}{\partial s}|(u,s)du, \tag{29}$$

so that using (25) we get

$$d_G(H(t,m), H(t,m')) \leq \int_0^t |\nabla Z_u|_\infty \exp(\int_0^u |\nabla U_{u'}|_\infty du') \int_0^1 |\frac{dp}{ds}|ds. \tag{30}$$

Since (26) and (30) is true for an arbitrary smooth path on M from m to m', we get (19) and (20). Thus the proof is complete □

Lemma 2.5. *Let $Y = (U, Z)$ and $Y' = (U', Z')$ be in $\mathcal{T}_c^\infty$ and let $A = (\Phi, H)$ (resp. $A' = (\Phi', H')$) be the solution of (15). Then for all $t \in [0,1]$ and all $m \in M$, we have*

$$d_M(\Phi_t(m), \Phi'_t(m)) \leq K(t), \tag{31}$$

$$d_G(H_t(m), H'_t(m)) \leq \int_0^t |(Z'-Z)_u|_\infty + (|\nabla Z_u|_\infty + |\nabla(Z'-Z))_u|_\infty) K(u)du, \tag{32}$$

where

$$K(t) = \int_0^t |(U-U')_u|_\infty e^{\int_u^t |\nabla U_{u'}|_\infty + |\nabla(U-U')_{u'}|_\infty du'} du. \tag{33}$$

Proof. Consider $\widetilde{Y} \in C^\infty([0,1]\times[0,1]\times M, TM \times \mathfrak{G})$ defined by

$$\widetilde{Y}(s,t,m) = Y(t,m) + s(Y'(t,m) - Y(t,m)). \tag{34}$$

As usual, we denote $\widetilde{U}$ its component on TM and $\widetilde{Z}$ its component on $\mathfrak{G}$. There exists $\widetilde{A}$ on $C^\infty([0,1]\times[0,1]\times M, M \times G)$ such that

$$\frac{\partial \widetilde{A}}{\partial t} = \widetilde{A}_t \widetilde{Y}_t. \tag{35}$$

We denote $\widetilde{\Phi}$ its component on M and $\widetilde{H}$ its component on G. Using covariant derivatives, we get

$$\frac{D}{dt}\left(\frac{\partial\widetilde{\Phi}}{\partial s}\right) = \frac{D}{ds}\left(\frac{\partial\widetilde{\Phi}}{\partial t}\right) = (U'-U)_t\circ\widetilde{\Phi}+\left(\nabla_{\frac{\partial\widetilde{\Phi}}{\partial s}}U_t+s\nabla_{\frac{\partial\widetilde{\Phi}}{\partial s}}(U'-U)_t\right)\circ\widetilde{\Phi}. \tag{36}$$

Let $r_t=|\frac{\partial\widetilde{\Phi}_t}{\partial s}|^2$. We deduce from (36) and the equality $\frac{\partial r}{\partial t} = 2\langle\frac{D}{dt}\left(\frac{\partial\widetilde{\Phi}}{\partial s}\right),\frac{\partial\widetilde{\Phi}}{\partial s}\rangle^M_{\widetilde{\Phi}}$ that

$$\frac{\partial r}{\partial t}\leq 2(|\nabla U|_\infty+s(\nabla(U'-U)|_\infty)r+2|U'-U|_\infty\sqrt{r}. \tag{37}$$

Applying the Gronwall's lemma to $\sqrt{r}$, we get finally

$$|\frac{\partial\widetilde{\Phi}}{\partial s}|\leq\int_0^t|U'-U|_\infty e^{\int_u^t|\nabla U|_\infty+s|\nabla(U'-U)|_\infty}du, \tag{38}$$

so that

$$d_M(\Phi(t,m),\Phi'(t,m)) = d_M(\widetilde{\Phi}(0,t,m),\widetilde{\Phi}(1,t,m))\leq K(t). \tag{39}$$

Thus, (31) is proved. We turn now to the proof of (32). We have

$$\begin{aligned}\frac{D}{dt}\left(\frac{\partial\widetilde{H}}{\partial s}\right) &= \frac{D}{ds}\left(\frac{\partial\widetilde{H}}{\partial t}\right) = \frac{D}{ds}\left(\widetilde{H}_t\left(Z_t+s(Z'_t-Z_t)\right)\right)\\ &= \widetilde{H}_t\left((Z'-Z)_t\circ\widetilde{\Phi}_t+\nabla_{\frac{\partial\widetilde{\Phi}}{\partial s}}(Z+s(Z'-Z))_t\circ\widetilde{\Phi}_t\right).\end{aligned} \tag{40}$$

Let $r_t=|\frac{\partial\widetilde{H}_t}{\partial s}|^2$. We deduce from (40) that

$$\frac{\partial r}{\partial t} = 2\left(|(Z'-Z)_t|_\infty+|\nabla(Z+s(Z'-Z))_t|_\infty|\frac{\partial\widetilde{\Phi}}{\partial s}|\right)\sqrt{r}. \tag{41}$$

Hence, using the upper bound (38) of $|\frac{\partial\widetilde{\Phi}}{\partial s}|$, we get

$$|\frac{\partial\widetilde{H}}{\partial s}(s,t,m)|\leq\int_0^t|(Z'-Z)_u|_\infty+\left(|\nabla(Z+s(Z'-Z))_u|_\infty\right)K(u)du. \tag{42}$$

Finally, we get

$$\begin{aligned}d_G(H(t,m),H'(t,m)) &= d_H(\widetilde{H}(0,t,m),\widetilde{H}(1,t,m))\\ &\leq\int_0^t|(Z'-Z)_u|_\infty+\left(|\nabla Z_u|_\infty+|\nabla(Z'-Z))_u|_\infty\right)K(u)du,\end{aligned} \tag{43}$$

so that the proof is complete. □

2.2. Definition of $\mathcal{A}(n)$. In the definition below we characterize the norms n for which we will define $\mathcal{A}(n)$.

Definition 2.6. Let n be a norm on $\mathfrak{X}(M)\times C^\infty(M,\mathfrak{G})$. We say that n is admissible if

(1) there exists $K>0$ such that for all $y=(u,z)\in\mathfrak{X}(M)\times C^\infty(M,\mathfrak{G})$ we have

$$|u|_\infty+|\nabla u|_\infty+|z|_\infty+|\nabla z|_\infty\le Kn(y), \tag{44}$$

(2) the function $y\to n(y)$ from $\mathfrak{X}(M)\times C^\infty(M,\mathfrak{G})$ to $\mathbb{R}$ is continuous for the C^∞ topology on $\mathfrak{X}(M)\times C^\infty(M,\mathfrak{G})$.

Moreover, for all $Y\in\mathcal{T}_c^\infty$, we define for $t\in[0,1]$

$$N^t(Y) \;=\; \int_0^t n(Y_s)ds. \tag{45}$$

The norm N^1 will be usually denoted N.

The condition on n is not restrictive and will be easily checked in most of the particular cases (see for instance the discussion at the end of the paper). Moreover, N defines a norm on $\mathcal{T}_c^\infty$. From now, we consider a fixed admissible norm n

Proposition 2.7. *Let $\mathbf{A}:\mathcal{T}_c^\infty\to C([0,1]\times M,M\times G)$ be such that for all $Y\in\mathcal{T}_c^\infty$, $\mathbf{A}(Y)$ is the solution of (15). Then,*

(1) *there exists $K>0$ such that for all Y and Y' in $\mathcal{T}_c^\infty$*

$$\sup_{s\le t} d_{\mathcal{A}}(\mathbf{A}(Y)_t,\mathbf{A}(Y')_t)\le KN^t(Y-Y')e^{K(N^t(Y)+N^t(Y'))} \tag{46}$$

(2) *the application $\mathbf{A}$ is Lipschitz, uniformly on bounded set, for the norm N on $\mathcal{T}_c^\infty$ and the distance D on $C([0,1]\times M,M\times G)$.*

Proof. The proof is a straightforward consequence of lemma 2.5 □

Corollary 2.8. *Since $(C([0,1]\times M,M\times G),D)$ is a complete metric space, $\mathbf{A}$ has an unique extension on the completion of $\mathcal{T}_c^\infty$ for the norm N. This completion will be denoted $\mathcal{T}_N$.*

Definition 2.9. From proposition 2.7 and corollary 2.8, one can define an application $\mathbf{a}:\mathcal{T}_N\to C(M,M\times G)$ such that for all $Y\in\mathcal{T}_N$

$$\mathbf{a}(Y)(m) \;=\; \mathbf{A}(Y)(1,m)\ ;\ m\in M. \tag{47}$$

We can now define the set $\mathcal{A}(n)$

Definition 2.10. Let $\mathcal{A}(n)$ be the subset of $\mathcal{A}$ define by

$$\mathcal{A}(n) \quad = \quad \{ \mathbf{a}(Y) \mid Y \in \mathcal{T}_N \}. \tag{48}$$

Notation 2.11. (1) For all Y and Y' in $\mathcal{T}_c^\infty$, we define $Y \star Y' \in \mathcal{T}_c^\infty$ by

$$(Y \star Y')(t,m) \quad = \quad 2(Y(2t,m)\mathbf{1}_{t\leq 1/2} + Y'(2t - \frac{1}{2}, m)\mathbf{1}_{t>1/2}. \tag{49}$$

Since $N(Y \star Y') = N(Y) + N(Y')$, we define $Y \star Y'$ by density on $\mathcal{T}_N \times \mathcal{T}_N$.

(2) For all $Y \in \mathcal{T}_c^\infty$, we define $S(Y) \in \mathcal{T}_c^\infty$ by

$$S(Y)(t,m) \quad = \quad -Y(1-t,m). \tag{50}$$

Again, since $N(S(Y)) = N(Y)$, we define $S(Y)$ by density on $\mathcal{T}_N$.

We obviously note that for $Y \in \mathcal{T}^\infty$, $\mathbf{A}(Y)_t \in \mathcal{A}$. Hence, for $Y \in \mathcal{T}_N$, $\mathbf{A}_t \in \mathcal{A}$ for all $t \in [0,1]$. Moreover we have the lemma

Lemma 2.12. *For all Y and Y' in $\mathcal{T}_N$ we have*

$$\mathbf{a}(Y \star Y') \quad = \quad \mathbf{a}(Y)\mathbf{a}(Y'), \tag{51}$$

$$\mathbf{a}(Y)\mathbf{a}(S(Y)) \quad = \quad e. \tag{52}$$

Proof. Using a density argument, if its sufficient to prove the result for Y and $Y' \in \mathcal{T}_c^\infty$. The proof of the first equality is then straightforward. For the second one, one just have to check by derivation that

$$\mathbf{A}(S(Y))_t\mathbf{A}(Y)_1 \quad = \quad \mathbf{A}(Y)_{1-t}\ ;\ t \in [0,1]. \tag{53}$$

Hence, the proof is complete □

Proposition 2.13. *The set $\mathcal{A}(n)$ is a sub-group of $\mathcal{A}$.*

Proof. From the previous lemma, we get that $\mathcal{A}(n)$ is stable for the product and the inverse. Since $e \in \mathcal{A}(n)$ we get the result. □

Definition 2.14. Let $d_{\mathcal{A}(n)} : \mathcal{A}(n) \times \mathcal{A}(n) \to \mathbb{R}^+$ defined by

$$d_{\mathcal{A}(n)}(e,a) \quad = \quad \inf\{ N(Y) \mid Y \in \mathcal{T}_N,\ \mathbf{a}(Y) = a \}\ ;\ a \in \mathcal{A}(n), \tag{54}$$

$$d_{\mathcal{A}(n)}(a,a') \quad = \quad d_{\mathcal{A}(n)}(e, a^{-1}a'), \tag{55}$$

where a^{-1} is the inverse of a in $\mathcal{A}$.

Theorem 2.15. *The application $d_{\mathcal{A}(n)}$ defines a left invariant distance on $\mathcal{A}(n)$ for which $(\mathcal{A}(n), d_{\mathcal{A}(n)})$ is complete.*

Proof. Let us show that $d_{\mathcal{A}(n)}$ is a distance. From proposition 2.7, we get that

$$d_{\mathcal{A}}(e,a) \leq K d_{\mathcal{A}(n)}(e,a) e^{K d_{\mathcal{A}(n)}(e,a)} \tag{56}$$

so that $d_{\mathcal{A}(n)}(e,a) = 0$ implies that $a = e$. Now, from (53), we deduce that for all a, $a' \in \mathcal{A}(n)$ and all $Y \in \mathcal{T}_N$ such that $a' = a\mathbf{a}(Y)$, we have $a = a'\mathbf{a}(S(Y))$. Since we have $N(Y) = N(S(Y))$ we deduce that $d_{\mathcal{A}(n)}(a,a') = d_{\mathcal{A}(n)}(a',a)$ and $d_{\mathcal{A}(n)}$ is symmetric. Finally, let a, a' and a'' be three points in $\mathcal{A}(n)$ and let Y and Y' be in $\mathcal{T}_N$ such that $a' = a\mathbf{a}(Y)$ and $a'' = a'\mathbf{a}(Y')$. From (52) we get

$$a'' \;=\; a\mathbf{a}(Y)\mathbf{a}(Y') \;=\; \mathbf{a}(Y \star Y'). \tag{57}$$

Since $N(Y \star Y') = N(Y) + N(Y')$, we deduce immediately that $d_{\mathcal{A}(n)}$ satisfies the triangle inequality so that $d_{\mathcal{A}(n)}$ is a distance.

We will prove now that $\mathcal{A}(n)$ is complete. Let us first introduce a family of operators on $\mathcal{T}_N$. Consider the sequence $(t_k)_{k\in\mathbb{N}}$ defined by $t_k = 1 - 2^{-k}$. For all p, $q \in \mathbb{N}$ such that $0 \leq p < q$, we consider the application $M_{p,q} : (\mathcal{T}_c^\infty)^{q-p} \rightarrow \mathcal{T}_c^\infty$

$$M_{p,q}(Y_p, \cdots, Y_{q-1}) \;=\; \sum_{k=p}^{q-1} 2^{k+1} Y_k(2^{k+1}(t - t_k)) \mathbf{1}_{t_k \leq t < t_{k+1}}. \tag{58}$$

One easily verifies that for $p < q < r$ we have the following properties

$$\mathbf{a}(M_{p,q}(Y_p, \cdots, Y_{q-1})) \;=\; \mathbf{a}(Y_p) \cdots \mathbf{a}(Y_{q-1}), \tag{59}$$

$$N(M_{p,r}(Y_p, \cdots, Y_{r-1}) - M_{p,q}(Y_p, \cdots, Y_{q-1})) \;=\; \sum_{k=q}^{r-1} N(Y_k), \tag{60}$$

so that we can extend by density $M_{p,q}$ in an application from $(\mathcal{T}_N)^{q-p}$ to $\mathcal{T}_N$ for which (59) and (60) remain true. Consider now a Cauchy sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathcal{A}(n)$. We can assume that $\sum_{n\in\mathbb{N}} d_{\mathcal{A}(n)}(a_n, a_{n+1}) < +\infty$. Thus, there exists a sequence $(Y_n)_{n\in\mathbb{N}}$ in $\mathcal{T}_N$ such that $\sum_{n\in\mathbb{N}} N(Y_n) < +\infty$ and $a_{n+1} = a_n\mathbf{a}(Y_n)$. Since the sequence $(a_n)_{n\in\mathbb{N}}$ is bounded in $(\mathcal{A}(n), d_{\mathcal{A}(n)})$, we deduce from proposition 2.7 (1) that there exists $K' > 0$ such that $d_{\mathcal{A}}(a_n, a_{n+p}) \leq K' d_{\mathcal{A}(n)}(a_n, a_{n+p})$. Hence $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{A}, d_{\mathcal{A}})$ so that there exists $a_\infty \in \mathcal{A}$ such that $d_{\mathcal{A}}(a_\infty, a_n) \rightarrow 0$. From equality (60), the limit in q for fixed p in $\mathcal{T}_N$ of $M_{p,q}(Y_p, \cdots, Y_{q-1})$ exists. Let $\widetilde{Y}_p$ be this limit. From equality (59) we deduce that $a_\infty = a_p \mathbf{a}(\widetilde{Y}_p)$. This last equality shows that $a_\infty \in \mathcal{A}(n)$ and that $d_{\mathcal{A}(n)}(a_\infty, a_p) \leq \sum_{k=p}^{\infty} N(Y_k)$. The proof is ended □

Proposition 2.16. *There exists $K > 0$ such that for all a and $a' \in \mathcal{A}(n)$, we have*

$$d_{\mathcal{A}}(a, a') \leq K d_{\mathcal{A}(n)}(a, a') \exp(K d_{\mathcal{A}(n)}(a, a')). \tag{61}$$

Proof. From proposition 2.7 we get that

$$d_{\mathcal{A}}(e, a) \leq K d_{\mathcal{A}(n)}(e, a) \exp(K d_{\mathcal{A}(n)}(e, a)). \tag{62}$$

Moreover, since

$$d_{\mathcal{A}}(a, a') = \sup_{m \in M} d_M(\phi(m), \phi'(m)) + \sup_{m \in M} d_G(h(m), h'(m)) \tag{63}$$

and

$$\begin{cases} \sup_{m \in M} d_M(\phi(m), \phi'(m)) = \sup_{m \in M} d_M(m, \phi' \circ \phi^{-1}(m)), \\ \sup_{m \in M} d_G(h(m), h'(m)) = \sup_{m \in M} d_G(1_G, h^{-1} \circ \phi^{-1}(m) h' \circ \phi^{-1}(m)), \end{cases} \tag{64}$$

we get that

$$d_{\mathcal{A}}(a, a') = d_{\mathcal{A}}(e, a^{-1} a'), \tag{65}$$

so that the proof is complete. □

3. A weak differentiable structure on $\mathcal{A}(n)$

3.1. Tangent spaces.

Definition 3.1. For all $a \in \mathcal{A}(n)$

(1) we define n^a the norm on $\widetilde{T}_a \mathcal{A}$ defined by

$$n^a(y) = n(a^{-1} y), \tag{66}$$

(2) we denote $\widetilde{T}_a \mathcal{A}(n)$ the completion of $\widetilde{T}_a \mathcal{A}$ for the norm n^a.

For all $a \in \mathcal{A}(n)$, the Banach space $\widetilde{T}_a \mathcal{A}(n)$ will play the role of the tangent space at a in $\mathcal{A}(n)$. However, as mention in the introduction, $\mathcal{A}(n)$ has not a structure of differential manifold modeled over a Banach space so that we will avoid the notation $T_a \mathcal{A}(n)$. Furthermore, since the application $y \rightarrow ay$ from $\widetilde{T}_e \mathcal{A}$ to $\widetilde{T}_a \mathcal{A}$ is an isometry, it can be extended to an application from $\widetilde{T}_e \mathcal{A}(n)$ to $\widetilde{T}_a \mathcal{A}(n)$ so that we have

$$\widetilde{T}_a \mathcal{A}(n) = \{ ay \mid y \in \widetilde{T}_e \mathcal{A}(n) \}. \tag{67}$$

3.2. Exponential.

Definition 3.2. (1) Let $j : \widetilde{T}_e\mathcal{A}\rightarrow\mathcal{T}_N$ be defined by

$$j(y)(t,m) = y(m). \tag{68}$$

Since, $N(j(y)) = n(y)$, j can be uniquely extended from $\widetilde{T}_e\mathcal{A}(n)$ to $\mathcal{T}_N$.

(2) Let $\exp : \widetilde{T}_e\mathcal{A}(n)\rightarrow\mathcal{A}(n)$ be defined by

$$\exp = \mathbf{a}\circ j. \tag{69}$$

(3) For all $a \in \mathcal{A}(n)$, we define $\exp_a : \widetilde{T}_a\mathcal{A}(n)\rightarrow\mathcal{A}(n)$ by

$$\exp_a(y) = a\exp(a^{-1}y). \tag{70}$$

Remark 1. The notation exp comes from the fact that for all $y \in \widetilde{T}_e\mathcal{A}$, for all $m \in M$, the application $t\rightarrow \exp(ty)(m) \in C^\infty(\mathbb{R}, G\times M)$ satisfies

$$\begin{cases} \exp(0y) = e \\ \frac{d}{dt}(\exp(ty)(m)) = (\exp(ty)y)(m), \end{cases} \tag{71}$$

so that $t\rightarrow\exp(ty)$ is a morphism from $\mathbb{R}$ to $\mathcal{A}(n)$. More generally, for all $a \in \mathcal{A}(n)$, all $y \in \widetilde{T}_a\mathcal{A}$ and all $m \in M$, the application $t\rightarrow \exp_a(ty)(m) \in C^\infty(\mathbb{R}, G\times M)$ satisfies

$$\begin{cases} \exp_a(0y) = a \\ \frac{d}{dt}(\exp_a(ty)(m)) = (\exp_a(ty)y)(m), \end{cases} \tag{72}$$

3.3. Differentiable applications. In spite of we have only a weak notion of differentiable structure on $\mathcal{A}(n)$, we will define the differentiability for functions on $\mathcal{A}(n)$

Definition 3.3. Let E be a function from $\mathcal{A}(n)$ to $\mathbb{R}$

(1) We say that E is differentiable at $a \in \mathcal{A}(n)$, if $E\circ\exp_a$ from the Banach space $\widetilde{T}_a\mathcal{A}(n)$ to $\mathbb{R}$ is differentiable at $0 \in \widetilde{T}_a\mathcal{A}(n)$ in the usual sense. We will use the notation d_aE to denote $d_0(E\circ\exp)$.
(2) We say that E is differentiable on $\mathcal{A}(n)$ if E is differentiable at any point $a \in \mathcal{A}(n)$.

3.4. Important examples. We are concerned here by a relevant example of differentiable applications in the context of pattern recognition.

Definition 3.4. (1) Let $R = M \times G$ on which we consider the metric defined for all $r = (r_1, r_2) \in R$ by

$$\langle w,\, w' \rangle_r^R \;=\; \langle w_1,\, w_1' \rangle_{r_1}^M + \langle w_2,\, w_2' \rangle_{r_2}^G \tag{73}$$

where $w = (w_1, w_2)$ and $w' = (w_1', w_2')$ are elements of $T_r R \;=\; T_{r_1} M \times T_{r_2} G$.

(2) For all $g \in C^2(R, \mathbb{R})$, we define

$$|\nabla^R g|_\infty \;=\; \sup\{\ |\nabla^R g(r)| \mid r \in X \times R\ \}, \tag{74}$$

$$|\nabla^R \nabla^R g|_\infty \;=\; \sup\{\ |\nabla^R_w \nabla^R g(r)| \mid (r, w) \in R \times T_r R,\ |w| \leq 1\ \}. \tag{75}$$

Theorem 3.5. *Assume that the action $(g, x) \rightarrow gx$ is C^2 and let f be a C^2 pattern i.e. $f \in C^2(M, X)$.*

(1) *Let $L \in C^2(X, \mathbb{R})$ such that*

$$|\nabla^R l|_\infty + |\nabla^R \nabla^R l|_\infty < +\infty, \tag{76}$$

where $l \in C^2(R, \mathbb{R})$ is defined by $l(r) = L(r_2 f(r_1))$. Then, the function $E : \mathcal{A}(n) \rightarrow \mathbb{R}$ defined by

$$E(a) \;=\; \int_M L(af) d\mu \tag{77}$$

is differentiable on $\mathcal{A}(n)$ and for all $a \in \mathcal{A}(n)$ and all $y \in \tilde{T}_a \mathcal{A}(n)$

$$d_a E(y) \;=\; \int_M \langle \nabla^R l(a),\, y \rangle^R d\mu. \tag{78}$$

(2) *Let $L \in C^2(X \times X, \mathbb{R})$ such that*

$$\sup_{x \in X} (|\nabla^R l_x|_\infty + |\nabla^R \nabla^R l_x|_\infty) < +\infty. \tag{79}$$

where $l \in C^2(X \times R, \mathbb{R})$ is defined by $l(x, r) = L(x, r_2 f(r_1))$ and l_x by $l_x(r) = l(x, r)$. Let $\tilde{f} \in \mathcal{P}$ such that $\tilde{f}(M)$ is relatively compact. Then the function $E : \mathcal{A}(n) \rightarrow \mathbb{R}$ defined by

$$E(a) \;=\; \int_M L(\tilde{f}, af) d\mu \tag{80}$$

is differentiable on $\mathcal{A}(n)$ *and for all* $a \in \mathcal{A}(n)$ *and* $y \in \widetilde{T}_a\mathcal{A}(n)$

$$d_a E(y) = \int_M \langle \nabla^R l(\tilde{f}, a), y \rangle^R d\mu \tag{81}$$

Remark 2. Note that for $y \in \widetilde{T}_a\mathcal{A}$, $d_aE(y)$ can be defined by (78) or (81), but should be extended by density for $y \in \widetilde{T}_a\mathcal{A}(n)$. In part (2), the assumption on the relative compactness of $\tilde{f}(M)$ is just to ensure that $L(\tilde{f}, af)$ is integrable.

Proof. Since part (2) implies obviously part (1), we will proved only the second part.

Let $a \in \mathcal{A}(n)$, $y = (u, z) \in \widetilde{T}_a\mathcal{A}$ and $m \in M$. Now, consider the applications $r = (r_1, r_2) \in C^\infty([0,1], R)$ and $q \in C^2([0,1], \mathbb{R})$ defined by $r(s) = exp_a(sy)(m)$ and $q(s) = l(\tilde{f}(m), r(s))$. We have

$$\frac{dq}{ds} = \langle \nabla^R l, \frac{dr}{ds} \rangle^R, \tag{82}$$

and

$$\frac{d^2q}{ds^2} = \langle \nabla^R_{\frac{dr}{ds}} \nabla^R l, \frac{dr}{ds} \rangle^R + \langle \nabla^R l, \frac{D^R}{ds}\left(\frac{dr}{ds}\right) \rangle^R \tag{83}$$

where $\frac{D^R}{ds}$ is the covariant derivative on R. Since

$$\frac{D^R}{ds}\left(\frac{dr}{ds}\right) = \left(\frac{D^M}{ds}\left(\frac{dr_1}{ds}\right), \frac{D^G}{ds}\left(\frac{dr_2}{ds}\right)\right) \tag{84}$$

and

$$\begin{cases} \dfrac{dr_1}{ds} = u(r_1), \\ \dfrac{dr_2}{ds} = r_2 z(r_1), \end{cases} \tag{85}$$

we get

$$\begin{cases} \dfrac{D^M}{ds}\left(\dfrac{dr_1}{ds}\right) = \nabla^M_u u(r_1), \\ \dfrac{D^G}{ds}\left(\dfrac{dr_2}{ds}\right) = r_2 \nabla^{\mathfrak{G}}_{\frac{dr_1}{ds}} z, \end{cases} \tag{86}$$

so that

$$|\frac{D^R}{ds}(\frac{dr}{ds})| \leq (|\nabla u|_\infty + |\nabla z|_\infty)|u|_\infty \leq n_a(y)^2. \tag{87}$$

Moreover we have $|\frac{dr}{ds}|^2 = |r_1 z(m)|^2 + |u(r_2)|^2 \leq n^a(y)^2$. Thus we deduce from (83) that

$$|\frac{d^2q}{ds}| \leq Mn^a(y)^2, \tag{88}$$

where

$$M = \sup_{x\in X}(|\nabla^R l_x|_\infty + |\nabla^R \nabla^R l_x|_\infty). \tag{89}$$

Hence, by integration by parts we get

$$|q(1) - q(0) - \frac{dq}{ds}(0)| \leq \int_0^1 (1-s)|\frac{d^2q}{ds}|ds \leq Mn^a(y)^2. \tag{90}$$

Since $\frac{dr}{ds}(0) = y(m)$, $\frac{dq}{ds}(0) = \langle \nabla^R l,\, y(m)\rangle^R$ and

$$|\int_M \langle \nabla^R l,\, y\rangle^R d\mu| \leq \sup_{x\in X} |\nabla^R l_x|_\infty n^a(y) \tag{91}$$

we get the result. The proof of the theorem is complete □

4. Hilbert sub-groups

In this part, we are concerned by a interesting particular case of admissible norm n.

Definition 4.1. Let n be an admissible norm. We say that $\mathcal{A}(n)$ is a Hilbert sub-group of $\mathcal{A}$ if there exists a scalar product $\langle\ ,\ \rangle_e$ on $\widetilde{T}_e\mathcal{A}$ such that $n(y) = (\langle y,\, y\rangle_e)^{1/2}$ for all $y \in \widetilde{T}_e\mathcal{A}$.

Finally, defining the scalar product $\langle\ ,\ \rangle_a$ on $\widetilde{T}_a\mathcal{A}$ by

$$\langle y,\, y'\rangle_a = \langle a^{-1}y,\, a^{-1}y'\rangle_e, \tag{92}$$

we get that for all $y \in \widetilde{T}_a\mathcal{A}$, $n^a(y) = (\langle y,\, y\rangle_a)^{1/2}$.

For Hilbert sub-groups $\mathcal{A}(n)$, the tangent spaces are separable Hilbert spaces so that for any differentiable application $E : \mathcal{A}(n)\to\mathbb{R}$, one can define the gradient of E.

Definition 4.2. Let $\mathcal{A}(n)$ be a Hilbert sub-group of $\mathcal{A}$ and let $E : \mathcal{A}(n)\to\mathbb{R}$ be a differentiable application. Then for any $a \in \mathcal{A}(n)$, we denote $\nabla_a E$ the unique element of $\widetilde{T}_a\mathcal{A}(n)$ such that for all $y \in \widetilde{T}_a\mathcal{A}(n)$ we have

$$d_aE(y) = \langle \nabla_a E,\, y\rangle_a. \tag{93}$$

For pattern classification and recognition tasks, we will have to consider non linear evolution equations on $\mathcal{A}(n)$ defined by $\frac{da}{dt} = -\nabla_a E$. More generally, we have to look for integrability conditions of vector fields on $\mathcal{A}(n)$.

4.1. Integration of vector fields on $\mathcal{A}(n)$.

Definition 4.3. Let F be a vector fields on $\mathcal{A}(n)$ i.e. an application from $\mathcal{A}(n)$ to $\widetilde{T}\mathcal{A}(n)$ such that $F(a) \in \widetilde{T}_a\mathcal{A}(n)$ for all $a \in \mathcal{A}(n)$.

(1) We say that F is bounded if there exists $K > 0$ such that

$$\sup_{a\in\mathcal{A}(n)} n^a(F(a)) \leq K. \tag{94}$$

(2) We say that F is strongly Lipschitz if there exists $K > 0$ such that for all a and a' in $\mathcal{A}(n)$ we have

$$n(a^{-1}F(a) - (a')^{-1}F(a')) \leq K d_{\mathcal{A}}(a, a'). \tag{95}$$

Since $d_A(a, a') \leq K d_{\mathcal{A}(n)}(a, a')$, if F is strongly Lipschitz, then F is Lipschitz in the usual sense. We show in the next lemma that there exists a canonical imbedding of $C([0,1], \widetilde{T}_e\mathcal{A}(n))$ in $\mathcal{T}_N$.

Lemma 4.4. *Let $V \in C([0,1], \widetilde{T}_e\mathcal{A}(n))$, then there exists a sequence $(V_n)_{n\in\mathbb{N}}$ in $\mathcal{T}_c^\infty$ such that $\int_0^1 n(V - V_n) \to 0$ when n tends to infinity.*

Proof. Let $(y_p)_{p\in\mathbb{N}}$ be a Hilbert basis of $\widetilde{T}_e\mathcal{A}(n)$. We denote $(\alpha_p)_{p\in\mathbb{N}}$ the family of component functions defined by $\alpha_p(t) = \langle y_p,\, V_t\rangle_e$. Since $V \in C([0,1], \widetilde{T}_e\mathcal{A}(n))$, we deduce that $\alpha_p \in C([0,1], \mathbb{R})$ for all $p \in \mathbb{N}$. Hence, there exists a family $(\alpha_p^k)_{k,p\in\mathbb{N}}$ of elements of $C^\infty([0,1], \mathbb{R})$ with compact supports in $]0,1[$ such that

$$\int_0^1 (\alpha_p^k - \alpha_p)^2 dt \leq 2^{-k}, \tag{96}$$

so that $N(V - \sum_{p\leq n} \alpha_p^n y_p) \to 0$ when n tends to infinity. The proof is complete if we notice that by definition of $\widetilde{T}_e\mathcal{A}$, for each $y_p \in \widetilde{T}_e\mathcal{A}(n)$, there exists a sequence $(y_p^k)_{k\in\mathbb{N}}$ in $\widetilde{T}_e\mathcal{A}$ such that $\alpha_p^k y_p^k \in \mathcal{T}_c^\infty$ and $\lim_{n\to\infty} n(y_p - y_p^k) = 0$. □

We get from the last lemma the following important corollary.

Corollary 4.5. *Assume that $\mathcal{A}(n)$ is a Hilbert sub-group of $\mathcal{A}$ and let F be a bounded and strongly Lipschitz vector field. Then, for all $a \in C([0,1], \mathcal{A}(n))$, $a^{-1}F(a) \in \mathcal{T}_N$.*

Proof. It is sufficient to show that $a^{-1}F(a) \in C([0,1], \widetilde{T}_e\mathcal{A}(n))$. However, from proposition 2.16 and the fact that F is strongly Lipschitz, we get that

$$n(a^{-1}(t)F(a(t)) - a^{-1}(t')F(a(t'))) \leq d_{\mathcal{A}}(a(t), a(t')) \leq K d_{\mathcal{A}(n)}(a(t), a(t')) \exp(K d_{\mathcal{A}(n)}(a(t), a(t'))), \quad (97)$$

so that the proof is complete. □

Theorem 4.6. *Assume that $\mathcal{A}(n)$ is a Hilbert sub-group of $\mathcal{A}$ and let F be a bounded and strongly Lipschitz vector field on $\mathcal{A}(n)$. Let $\hat{F} : \mathcal{A}(n) \rightarrow \widetilde{T}_e\mathcal{A}(n)$ by defined by $\hat{F}(a) = a^{-1}F(a)$. Then there exists $p \in C([0,1], \mathcal{A}(n))$ such that*

$$p = \mathbf{A}(\hat{F} \circ p). \quad (98)$$

The equation (98) is nothing more than an integrated version of the formal evolution equation

$$\frac{\partial p}{\partial t} = F \circ p. \quad (99)$$

Moreover, from corollary 4.5, since $p \in C([0,1], \mathcal{A}(n))$, $\hat{F} \circ p \in \mathcal{T}_N$ and equality (98) is well defined.

4.2. Proof of the theorem 4.6. We will use an iterative scheme. We will built a a sequence of approximates $p_n \in C([0,1], \mathcal{A}(n))$ by induction:

(1) For all $t \in [0,1]$, $p_0(t) = e$,
(2) $p_{n+1} = \mathbf{A}(Y_n)$ where $Y_n = \hat{F}(p_n)$.

The sequence p_n is not defined until we have proved that $Y_n \in \mathcal{T}_N$. It is sufficient to prove that if $p \in C([0,1], \mathcal{A}(n))$ then $\hat{F} \circ p \in C([0,1], \widetilde{T}_e\mathcal{A}(n))$ and that $A(Y) \in C([0,1], \mathcal{A}(n))$ for all $Y = \hat{F} \circ p$. The first part is proved by the corollary 4.5. For the second part, note that if $Y \in C([0,1], \widetilde{T}_e\mathcal{A}(n))$, then for all $0 \leq s \leq t \leq 1$ we have

$$d_{\mathcal{A}(n)}(\mathbf{A}(Y)_s, \mathbf{A}(Y)_t) \leq \int_s^t n(Y_u) du. \quad (100)$$

Since $u \rightarrow n(Y_u)$ is continuous on $[0,1]$, we deduce that $\mathbf{A}(Y) \in C([0,1], \mathcal{A}(n))$. Hence, the sequence $(p_n)_{n \in \mathbb{N}}$ is well defined.

Lemma 4.7. *The sequence $(p_n)_{n \in \mathbb{N}}$ converges in $C([0,1] \times M, M \times G)$.*

Proof. Using proposition 2.7 and the fact that F is bounded, we deduce that there exists $K>0$ such that

$$d_{\mathcal{A}}(p_{n+1}(t),p_n(t)) \leq K\int_0^t n(\hat{F}(p_n)-\hat{F}(p_{n-1}))ds. \tag{101}$$

Hence, if we denote $r_n(t)=\sup_{s\leq t} d_{\mathcal{A}}(p_{n+1}(s),p_n(s))$, we get that there exists K' (independent of n) such that

$$r_n(t)\leq K'\int_0^t r_{n-1}(s)ds, \tag{102}$$

so that

$$r_n(1)\leq \frac{(K')^n}{n!}r_0(1), \tag{103}$$

and the proof is complete. □

From the previous lemma, we get that there exists $a_\infty \in C([0,1],\mathcal{A})$ which is the limit of a_n. To prove that a_∞ is the solution of our problem, we have to show that $a_\infty \in C([0,1],\mathcal{A}(n))$. The key argument is the following.

Lemma 4.8. *The sequence $(Y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{T}_N$.*

Proof. Since F is strongly Lipschitz, we have $N(Y_{n+1}-Y_n)\leq Kr_n(1)$ where r_n is defined in the previous proof. Hence the proof is complete. □

From the previous lemma, we can define Y_∞ as the limit of Y_n in $\mathcal{T}_N$ so that $a_\infty = A(Y_\infty)$. Moreover, using again the fact that F is strongly Lipschitz, we get that there exists $K>0$ such that $N(\hat{F}\circ a_\infty - Y_\infty)\leq KD(a_n,a_\infty)$ so that we obtain that $Y_\infty=\hat{F}\circ a_\infty$. This complete the proof of the theorem.

4.3. Convergence of the Cauchy approximates. Let $0=t_0\leq t_1\leq\cdots\leq t_n=1$ be a subdivision denoted σ of the interval $[0,1]$. We define the Cauchy approximate associated to σ of a_∞ by

- $a^\sigma(0)=e$,
- $a^\sigma(t)=\exp_{a^\sigma(t_k)}((t-t_k)\hat{F}(a^\sigma(t_k))$ for $t_k\leq t<t_{k+1}$.

The path a^σ is obtained through the approximation of $\hat{F}(a^\sigma(t))$ by $\hat{F}(a^\sigma(t_k))$ for $t\in[t_k,t_{k+1}[$. We will show that there exists $K>0$ such that

$$d_{\mathcal{A}}(a^\sigma,a_\infty)\leq K|\sigma|, \tag{104}$$

where $|\sigma|=\sup_{0\leq k<n}(t_{k+1}-t_k)$.

Indeed, let $Y^\sigma = \sum_{k=0}^{n-1} \hat{F}(a^\sigma(t_k))\mathbf{1}_{t_k\le t<t_{k+1}}$. We verifies easily that $a^\sigma = \mathbf{A}(Y^\sigma)$. Hence, since F is bounded, we get from the proposition 2.13 that there exists $K>0$ such that

$$\begin{aligned} d_{\mathcal{A}}(a_\infty(t), a^\sigma(t)) &\le K\int_0^t n(\hat{F}(a_\infty(s)) - Y^\sigma(s))ds \\ &\le K\int_0^t n(\hat{F}(a_\infty(s)) - \hat{F}(a^\sigma(s)))ds + K\int_0^t n(\hat{F}(a^\sigma(s)) - Y^\sigma(s))ds. \quad (105) \end{aligned}$$

Now, since F is strongly Lipschitz, we get first that there exists $K>0$ (independent of σ) such that $n(\hat{F}(a^\sigma(s)) - Y^\sigma(s)) \le K|\sigma|$ and $n(\hat{F}(a_\infty(s)) - \hat{F}(a^\sigma(s))) \le K d_{\mathcal{A}}(a_\infty(s), a^\sigma(s))$. Then, using Gronwall's lemma, we obtain the inequality (104).

4.4. Application to gradient algorithms. We come back the our important examples of differentiable applications introduced in theorem 3.5.

Proposition 4.9. *Assume that $\mathcal{A}(n)$ is a Hilbert sub-group of $\mathcal{A}$ and that the action $(g,x)\to gx$ is C^2. Let f be a C^2 pattern i.e. $f\in C^2(M,X)$.*

(1) *Let $L\in C^2(X,\mathbb{R})$ such that*

$$|\nabla^R l|_\infty + |\nabla^R\nabla^R l|_\infty < +\infty, \quad (106)$$

where $l\in C^2(R,\mathbb{R})$ is defined by $l(r) = L(r_2 f(r_1))$. Let $E:\mathcal{A}(n)\to\mathbb{R}$ defined by

$$E(a) = \int_M L(af)d\mu. \quad (107)$$

Then the gradient field ∇E is bounded and strongly Lipschitz.

(2) *Let $L\in C^2(X\times X,\mathbb{R})$ such that*

$$\sup_{x\in X}(|\nabla^R l_x|_\infty + |\nabla^R\nabla^R l_x|_\infty) < +\infty. \quad (108)$$

where $l\in C^2(X\times R,\mathbb{R})$ is defined by by $l(x,r) = L(x, r_2 f(r_1))$ and l_x by $l_x(r) = l(x,r)$. Let $\tilde{f}\in\mathcal{P}$ such that $\tilde{f}(M)$ is relatively compact. Let $E:\mathcal{A}(n)\to\mathbb{R}$ defined by

$$E(a) = \int_M L(\tilde{f}, af)d\mu \quad (109)$$

Then the gradient field ∇E is bounded and strongly Lipschitz.

Proof. Part (2) obviously implies part (1) so that we will do the proof of the second part. We use the notation of the proof of the theorem 3.5. Let $a \in \mathcal{A}(n)$ and $y \in \widetilde{T}_a\mathcal{A}(n)$. Since

$$|d_aE(y)| \leq |\int_M \langle \nabla^R l(\tilde{f}, a), \, y\rangle^R d\mu| \leq \sup_{x\in X} |\nabla^R l_x|_\infty n^a(y), \tag{110}$$

we deduce that $n^a(\nabla_a E) \leq \sup_{x\in X} |\nabla^R l_x|_\infty$ so that ∇E is bounded. We want to prove now that ∇E is strongly Lipschitz. Let a and a' be in $\mathcal{A}(n)$, let $m \in M$, let $y \in \widetilde{T}_e\mathcal{A}(n)$ and let $p \in C^\infty([0,1], M \times G)$ such that $p(0) = a(m)$ and $p(1) = a'(m)$. Let $q \in C^1([0,1], \mathbb{R})$ be defined by $q(s) = \langle \nabla^R l(\tilde{f}(m), p(s)), \, p(s)y\rangle^R$ where $p(s)y$ denotes $(u(p_1(s)), p_2(s)z(p_1(s)))$ with $p = (p_1, p_2)$ and $y = (u, z)$. One computes

$$\frac{dq}{ds} = \langle \nabla^R_{\frac{dp}{ds}} \nabla^R l(\tilde{f}(m), p(s)), \, p(s)y\rangle^R + \langle \nabla^R l(\tilde{f}(m), p(s)), \, \frac{D^R}{ds}(p(s)y)\rangle^R. \tag{111}$$

Since

$$|\frac{D^R}{ds}(p(s)y)| = |(\nabla^M_{\frac{dp_1}{ds}} u, p_2(s)\nabla^{\mathfrak{G}}_{\frac{dp_1}{ds}} z)| \leq n(y)|\frac{dp}{ds}|, \tag{112}$$

we get

$$|\frac{dq}{ds}| \leq Mn(y)|\frac{dp}{ds}|, \tag{113}$$

so that

$$|\langle \nabla^R l(\tilde{f}(m), a(m)), \, (ay)(m)\rangle^R - \langle \nabla^R l(\tilde{f}(m), a'(m)), \, (a'y)(m)\rangle^R| \leq Mn(y)d_{\mathcal{A}}(a, a'), \tag{114}$$

where

$$M = \sup_{x\in X}(|\nabla^R l_x|_\infty + |\nabla^R\nabla^R l_x|_\infty). \tag{115}$$

After integration under μ, we obtain finally that for all $y \in \widetilde{T}_e\mathcal{A}(n)$ we have

$$|d_aE(ay) - d_{a'}E(a'y)| \leq Mn(y)d_{\mathcal{A}}(a, a') \tag{116}$$

so that

$$n(a^{-1}\nabla_a E - (a')^{-1}\nabla - a'E) \leq Mn(y)d_{\mathcal{A}}(a, a'). \tag{117}$$

Thus, ∇E is strongly Lipschitz and the proof is complete. □

5. Application to pattern recognition

5.1. Sub-optimal solutions. We turn back to the problem of pattern classification and matching as set in the introduction. Let $\mathcal{A}(n)$ be a Hilbert sub-group of $\mathcal{A}$ and assume that the action is C^2. Let $L \in C^2(X \times X, \mathbb{R})$ be non negative. The value of $L(x, x')$ should be interpreted as a distance between x and x'. Let $(f_i)_{1 \leq i \leq p}$ be a family of C^2 patterns in $\mathcal{P}$ called the template patterns. Now, let $\tilde{f} \in \mathcal{P}$ be the observed pattern. For all $i \in \{1, \cdots, p\}$, we define $E_i : \mathcal{A}(n) \rightarrow \mathbb{R}$ by and $W_i : \mathcal{A}(n) \rightarrow \mathbb{R}$ by

$$E_i(a) = \int_M L(\tilde{f}, af_i) d\mu, \tag{118}$$

and

$$W_i(a) = E_i(a) + d_{\mathcal{A}(n)}(e, a)^2. \tag{119}$$

Let $S_i = \inf_{a \in \mathcal{A}(n)} W_i(a)$. We will say that $\tilde{f}$ belongs to the class of f_i if $S_i \leq S_j$ for all $j \neq i$. The problem of classification is then connected to the computation of the value of S_i. This computation cannot be performed directly. Moreover, even if we look for a local minimum of W_i through a gradient algorithm on W_i, one should have to prove that $a \rightarrow d_{\mathcal{A}(n)}(e, a)^2$ is differentiable in the sense of definition 3.3 and that the gradient field is bounded and strongly Lipschitz. This is not likely to be true in general. However, concerning the function E_i, we have proved in the previous part that (cf definition 3.4 and theorem 3.5) if

$$\sup_{x \in X} (|\nabla^R l_i(x, .)|_\infty + |\nabla^R \nabla^R l_i(x, .)|_\infty) < +\infty \tag{120}$$

where $l_i \in C^2(X \times R, \mathbb{R})$ is given by $l_i(x, r) = L(x, r_1 f_i(r_2))$ for all $x \in X$ and $r = (r_1, r_2) \in R$, then E_i is differentiable and the gradient field in bounded and strongly continuous. Hence, there exists $p_i \in C([0, 1], \mathcal{A}(n))$ with is the solution of the formal gradient equation

$$\frac{dp_i}{dt} = -\nabla_{p_i} E_i, \tag{121}$$

i.e.

$$p = \mathbf{A}(Y) \text{ with } Y_t = -p_i(t)^{-1} \nabla_{p_i(t)} E_i. \tag{122}$$

In fact, the gradient equation have a solution in $[0, +\infty[$. We just have to define by induction on $n \in \mathbb{N}$

$$p_i(t + n) = p_i(n) s_n(t) \; ; t \in [0, 1], \tag{123}$$

where s_n is solution of

$$\frac{ds_n}{dt} = -\nabla_{s_n}(E_i \circ L_{p_i(n)}), \tag{124}$$

and $L_{p_i(n)}$ denotes the left multiplication by s_n. Now, let $q_i \in C(\mathbb{R}_+, \mathbb{R}_+)$ be defined by

$$q_i(t) = E_i(p_i(t)) + \left(\int_0^t n(\nabla_{p_i(s)} E_i) ds\right)^2 \tag{125}$$

The function q_i is in fact in $C^1([0, +\infty[, \mathbb{R})$ as shown in the following proposition.

Proposition 5.1. *The function $q_i \in C^1([0, +\infty[, \mathbb{R})$ and*

$$\frac{dq_i}{dt} = -|\nabla_{p_i} E_i|^2_{p_i} + (2\int_0^t |\nabla_{p_i} E_i|_{p_i} ds)|\nabla_{p_i} E_i|_{p_i}. \tag{126}$$

Proof. It is sufficient to prove that $c(t) = E_i(p_i(t))$ is C^1 for $t \in [0,1]$ and that $\frac{de}{dt} = -|\nabla_{p+i} E_i|^2_{p_i}$. Let $(Y^n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{T}_c^\infty$ such that $N(Y^n - Y) \to 0$ where $Y_t = p_i(t)^{-1}\nabla_{p_i(t)} E_i$. Let $p^n = A(Y^n)$ and $c_n(t) = E_i(p^n(t))$. Since $Y^n \in \mathcal{T}_c^\infty$, one easily get that

$$\begin{aligned} c_n(t) - c_n(0) &= -\int_0^t \langle \nabla_{p^n} E_i,\, p^n Y^n \rangle_{p^n} ds \\ &= -\int_0^t \langle \nabla_{p_i} E_i,\, \nabla_{p_i} E_i \rangle_{p_i} ds + \epsilon_n, \end{aligned} \tag{127}$$

where

$$\epsilon_n = \int_0^t \langle \nabla_{p_i} E_i,\, \nabla_{p_i} E_i \rangle_{p_i} ds - \int_0^t \langle \nabla_{p^n} E_i,\, p^n Y^n \rangle_{p^n} ds. \tag{128}$$

However,

$$\begin{aligned} &|\langle \nabla_{p_i} E_i,\, \nabla_{p_i} E_i \rangle_{p_i} - \langle \nabla_{p^n} E_i,\, p^n Y^n \rangle_{p^n}| \\ &\quad \le |\langle \nabla_{p^n} E_i,\, p^n Y^n - p^n p_i^{-1} \nabla_{p_i} E_i \rangle_{p^n} - \langle p^n p_i^{-1} \nabla_{p_i} E_i - \nabla_{p^n} E_i,\, p^n p_i^{-1} \nabla_{p_i} E_i \rangle_{p^n}| \\ &\quad\quad \le |\nabla_{p^n} E_i|_{p^n} |Y^n - p_i^{-1} \nabla_{p_i} E_i|_e + |\nabla_{p_i} E_i|_{p_i} |p_i^{-1} \nabla_{p_i} E_i - (p^n)^{-1} \nabla_{p^n} E_i|_e. \end{aligned} \tag{129}$$

Since ∇E is bounded and strongly Lipschitz, there exists $K > 0$ such that

$$|\epsilon_n| \le KN(Y^n - Y) + KD(p_i, p^n) \tag{130}$$

where D is defined in definition 2.3 so that $\epsilon_n \to 0$ when n tends to infinity. Moreover, $c_n(t) - c_n(0) \to c(t) - c(0)$ so that

$$c(t) - c(0) = -\int_0^t |\nabla_{p_i} E_i|^2_{p_i}, \tag{131}$$

and the proof is complete. □

Proposition 5.2. *There exists $\hat{t}_i \in [0, +\infty[$ so that*

$$q_i(\hat{t}_i) \quad = \quad \inf_{t \geq 0} q_i(t). \tag{132}$$

Proof. We do the proof by contradiction. Assume that there exists an increasing sequence $(t_n)_{n \in \mathbb{N}}$ such that $\lim t_n = +\infty$ and $q_i(t_n) > q_i(t_{n+1})$. Then, for all $n \geq 0$, there exits $t_n^* \in]t_n, t_{n+1}[$ such that $\frac{dq_i}{dt}(t_n^*) < 0$ i.e.

$$|\nabla_{p_i(t_n^*)} E_i| \geq 2 \int_0^{t_n^*} |\nabla_{p_i} E_i| ds. \tag{133}$$

Assume that $\int_0^\infty |\nabla_{p_i} E_i| ds > 0$. Then considering eventually a subsequence of $(t_n)_{n \in \mathbb{N}}$, we can assume that there exists $\alpha > 0$ such that $2 \int_0^{t_n^*} |\nabla_{p_i} E_i| ds > \alpha$ for all $n \geq 0$. Since ∇E_i is bounded, there exists $K > 0$ (cf proposition 2.16) such that for all $t, t' \leq 0$, $d_{\mathcal{A}}(p_i(t), p_i(t')) \leq K|t - t'| \exp(K(|t - t'|)$. Now, using the fact that ∇E_i is strongly Lipschitz, we get that there exists $\eta > 0$ such that for all $t \in [t_n^* - \eta, t_n^* + \eta]$

$$|\nabla_{p_i(t_n^*)} E_i| \geq \alpha_2. \tag{134}$$

Thus, $\int_0^\infty |\nabla_{p_i} E_i|^2 ds = +\infty$. Since we have proved that $\frac{d}{dt}(E_i \circ p_i) = -|\nabla_{p_i} E_i|^2$ we deduce that $E_i(p_i(t)) \rightarrow -\infty$ which is in contradiction with the fact that E is non negative. Hence $\int_0^\infty |\nabla_{p_i} E_i|^2 ds = 0$. However, we get in this case that $p_i(t) = e$ for all $t \geq 0$ so that q_i is constant. This is again in contradiction with $q_i(t_n) > q_i(t_{n+1})$. The proof is ended. □

From the previous proposition, we define

$$\begin{cases} \hat{S}_i & = & q_i(\hat{t}_i), \\ \hat{p}_i & = & p_i(\hat{t}_i). \end{cases} \tag{135}$$

Since, one has

$$\begin{cases} W_i(e) & = & q_i(0), \\ W_i(p_i(t)) \leq q_i(t), \end{cases} \tag{136}$$

we get that

$$S_i \leq \hat{S}_i. \tag{137}$$

Definition 5.3. We will say that $\hat{i} = \operatorname{argmin} \hat{S}_i$ is the sub-optimal solution of the classification problem and $\hat{p}_{\hat{i}}$ the the sub-optimal solution of the matching problem.

The essential fact here, is that the sub-optimal solution of the classification problem as well as the sub-optimal solution of the matching problem are well defined and can be numerically computed, so that this approach seems very attractive.

5.2. Examples. In this subsection, we will present some applications of the previous scheme to various tasks in images and signals processing.

5.2.1. *Structural restoration of grey level images.* In this framework, we say that a grey level image is a measurable function from $M = \mathbb{R}^2/\mathbb{Z}^2$ to $X = \mathbb{R}$. One could choose for M the unit square $[0,1]^2$ but we prefer the choice of the 2-dimensional torus to have a compact manifold without boundary. This choice also allows to define translated images $f_u(m) = f(m+u)$. We single out a C^2 template in $\mathcal{P}$ denoted f and we consider an observed images $\tilde{f} \in \mathcal{P}$. The problem of structural restoration of images as defined in [1] is described in the following way. We consider a Hilbert space of vector fields on M and we define the solution of the structural restoration problem by

$$\hat{u} \quad = \quad \operatorname{argmin} \int_M (\tilde{f}(m) - f(m+u(m))^2 d\mu + \langle u,\, u\rangle_H. \tag{138}$$

Given $\hat{u}$, we have a complete matching between the points of f and those of $\tilde{f}$ by $m \rightarrow m + u(m)$ (note that $m + u(m)$ should be interpreted as a sum mod 1). This approach has been performed in the case of X-rays images of hands in [1]. However, one of the main drawbacks of this approach is that the matching $m \rightarrow m + u(m)$ is not onto nor injective on M. This problem is particularly visible when large deformations are involved.

In our framework, the problem can be well-posed. It corresponds to the case G reduced to $\{1_G\}$. Then the tangent space $\tilde{T}_e\mathcal{A}$ is isomorphic to $\mathfrak{X}(M)$. Now, if we define the norm n on $\tilde{T}_e\mathcal{A}$ by

$$n(y) \quad = \quad (\langle y,\, y\rangle_H)^{1/2}, \tag{139}$$

and assuming that n is admissible (which is the case for the scalar product considered in [1]), we can defined the Hilbert sub-group $\mathcal{A}(n)$ of $\mathcal{A}$ whose tangent space $\tilde{T}_e\mathcal{A}(n)$ is isomorphic to H. Then, it appears that

$$\int_M (\tilde{f}(m) - f(m+u(m))^2 d\mu + \langle u,\, u\rangle_H \tag{140}$$

is an approximation near $e = \mathrm{Id}_M$ ($u(m) = \phi(m) - m$) to

$$\int_M (\tilde{f}(m) - f(\phi(m))^2 d\mu + d_{\mathcal{A}(n)}(e, \phi)^2. \tag{141}$$

Now, since $(x, x') \to (x - x')^2$ is C^2 and since G is compact, we can define the sub-optimal solution $\hat{\phi}$ of the recognition problem (in this case, we have only one template). Since $\hat{\phi} \in \mathrm{Hom}(M)$, the matching is invertible.

With our framework, we can also allows a simultaneous displacement of the points of M and a variation of the grey levels. It is sufficient to consider $G = \mathbb{R}$ with the action $gx = g + x$. In this case, we have $\mathfrak{G} = \mathbb{R}$ and $\tilde{T}_e\mathcal{A}$ is isomorphic to $C^\infty(M, \mathbb{R}^2 \times \mathbb{R})$. The admissible norm can be given for $y = (u, z) \in \tilde{T}_e\mathcal{A}$ by

$$n(y) \;=\; (\langle u,\, u\rangle_H + \langle z,\, z\rangle_{H'})^{1/2}, \tag{142}$$

where $\langle\ ,\ \rangle_{H'}$ is a scalar product on $C^\infty(M, \mathbb{R})$ such that

$$|z|_\infty + |\nabla z|_\infty \le K \langle z,\, z\rangle_{H'}{}^{1/2}. \tag{143}$$

5.2.2. *Structural restoration of displacement fields.* We consider here that $X = \mathbb{R}^2$, that is the patterns are vectors fields in M. Then one can choose for G either $G = \{1_G\}$ or $\mathbb{R}^2$ according the fact that we want or not to deform the values of $f(m)$ for $m \in M$. A more unusual case is $X = S^1$, that is $f(m)$ is an unit vector, and $G = S^1$ with the action given by the complex product (here we consider S^1 as the set of the complex number with norm 1). Again in this case, $\tilde{T}_e\mathcal{A}$ is isomorphic to $C^\infty(M, \mathbb{R} \times \mathbb{R})$ and we can define the norm n by (142). For the function L, we can choose

$$L(x, x') \;=\; |x - x'|^2, \tag{144}$$

where x and x' are again considered as elements of $\mathbb{C}$ and $|x|$ denotes the usual norm on $\mathbb{C}$. One verifies easily that L is C^∞. Since G is compact, the condition for the differentiability of $E(a) = \int_M L(\tilde{f}, af) d\mu$ is fulfilled and we can define the sub-optimal solution $\hat{p}$ of the matching problem.

5.2.3. *Active contours.* We consider here closed curves living in $\mathbb{R}^p$, so that $M = \mathbb{R}/\mathbb{Z}$ and $X = \mathbb{R}^p$. For G, a natural choice is $\mathbb{R}^p$ with the action $gx = g + x$. Given a admissible norm on $\tilde{T}_e\mathcal{A}$ and a penalty function $L \in C^2(X, \mathbb{R}_+)$, the solution p of the formal gradient equation

$$\frac{dp}{dt} \;=\; -\nabla_p E \tag{145}$$

can be interpreted as a method of deformable contours as introduced in [7]. However, in our framework, the solution of (145) is well defined for all $t \geq 0$.

5.3. Choices of n. We will not go further in our examples, since the framework is sufficiently general to be apply in many situations. We want here to show that the condition of admissibility on n is weak. We will consider the case when $M = \mathbb{R}^p/\mathbb{Z}^p$ and the Lie algebra of G is isomorphic to $\mathbb{R}^q$. Then, $\widetilde{T}_e\mathcal{A}$ is isomorphic to $C^\infty(M, \mathbb{R}^p \times \mathbb{R}^q)$. Let $y = (u, z) \in \widetilde{T}_e\mathcal{A}$ where u is the component on $\mathbb{R}^p$ and z on $\mathbb{R}^q$. Since M is the p-dimensional torus, one can define for all p-uplet $\hat{n} \in \mathbb{N}^p$ the Fourier coefficient $u^k_{\hat{n}}$ (resp. $z^k_{\hat{n}}$) of the k-th component of u (resp. z). Then let $(a_{\hat{n}})_{\hat{n}\in\mathbb{N}^p}$ be a sequence of non negative numbers and define

$$n(y) = \left(\sum_{\hat{n}\in\mathbb{N}^p} a_{\hat{n}}(|u_{\hat{n}}|^2 + |z_{\hat{n}}|^2)\right)^{1/2}. \tag{146}$$

We get from the Sobolev imbeddings that n is an admissible norm if there exist $\beta \geq \alpha \geq p + 3$, $K > 0$ and $K' > 0$ such that for all $\hat{n} \in \mathbb{N}^p$

$$K'(|\hat{n}| + 1)^\beta \geq a_{\hat{n}} \geq K(|\hat{n}| + 1)^\alpha \tag{147}$$

where $|\hat{n}| = \sup |n_i|$. For example, in the case curves in $\mathbb{R}^2$, i.e. $M = \mathbb{R}/\mathbb{Z}$ and $X = \mathbb{R}^2$, we can choose the norm

$$n(y)^2 = \int (\Delta u)^2 + u^2 d\mu + \int |\Delta z|^2 + |z|^2 d\mu \tag{148}$$

where Δ is the Laplacian. The case of norm n defined with the Fourier coefficients is particularly appealing for numerical reasons since one can use Fast Fourier Transform on computer in the implementation. However, the norm should be chosen according to the regularity expected on the elements of $\mathcal{A}(n)$.

References

1. Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable template. preprint.
2. R. Azencott. Random and deterministic deformations applied to shape recognition. 1994. Cortona workshop, 10th-16th April, Italy.
3. W. M. Boothby. *An introduction to differentiable manifolds and riemannian geometry*, volume 63 of *Pure and applied mathematics*. Academic Press, 1975.
4. Y. Chow, U. Grenander, and D. M. Keenan. *HANDS, A pattern Theoetical Study of Biological Shapes*. Springer-Verlag, 1991.
5. R. Godement. *Introduction à la théorie des groupes de Lie*. Publications mathématiques de l'université de Paris VII, 1982.

6. V. Kac, editor. *Infinite Dimensional Groups with Applications*, volume 4 of *Mathematical Sciences Resarch Institute Publications*. Springer-Verlag, 1985.
7. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Actives contour models. *Internat. J. Comput. Vision*, 1:312–331, 1988.
8. H. Omori. *Infinite Dimensional Lie Transformations Groups*, volume 427 of *Lecture Notes in Math.* Springer-Verlag, 1974.
9. R. Palais. *A global formulation of the Lie Theory of Transformations groups*, volume 22. Memoir of the AMS, 1957.
10. R. S. Palais. *Foundations of global non-linear analysis*. W. A. Benjamin, Inc., 1968.
11. M. Piccioni, S. Scarlatti, and A. Trouvé. A variational problem arising from speech recognition. Preprint, 1995.
12. L. Younes. Computable elastic distance between shapes. Preprint, 1995.

LMENS/DIAM, URA 762, Ecole Normale Supérieure, 45, rue d'Ulm, 75230 Paris cedex 05, France

E-mail address: trouve@ens.ens.fr