THE GEOMETRY AND CURVATURE OF SHAPE SPACES

The idea that the set of all smooth submanifolds of a fixed ambient finite dimensional differentiable manifold forms a manifold in its own right, albeit one of infinite dimension, goes back to Riemann. We quote his quite amazing Habilitatsionschrift:

> *There are, however, manifolds in which the fixing of position requires not a finite number but either an infinite series or a continuous manifold of determinations of quantity. Such manifolds are constituted for example by the possible shapes of a figure in space, etc.*

The group of diffeomorphisms of a fixed finite dimensional manifold is one such infinite dimensional manifold. The differential geometry of the subgroup of volume preserving diffeomorphisms was studied in the ground breaking paper of Arnold [1] where, in particular, he showed that its geodesics (in the simplest L^2 metric) were the solutions of the Euler equation of incompressible fluid flow. In recent years, the demands of medical imaging and, more generally, of object recognition in computer vision, have stimulated work on the space of simple closed plane curves in $\mathbb{R}^2$ and the space of compact surfaces in $\mathbb{R}^3$ homeomorphic to a sphere. One can endow these spaces with a variety of different Riemannian metrics and work out both the geodesic equation and the curvature tensor in these metrics. Many different phenomena appear giving these spaces very different characteristics in different metrics. My lecture will discuss four examples, each illustrating quite different behavior, based largely on joint work in the last ten years with my collaborators and students Peter Michor, Laurent Younes, Jayant Shah, Eitan Sharon, Matt Feiszli, Mario Micheli and Sergey Kushnarev.

§1. The simplest possible example one might look at is the L^2 metric on the space of simple closed plane curves. To fix notation, let $\mathcal{S}$ be this space, the curves being assumed to be smooth, i.e. C^∞. Let $[C] \in \mathcal{S}$ be the point defined the curve $C \subset \mathbb{R}^2$. The tangent space $T_{[C]}\mathcal{S}$ is naturally isomorphic to the space of normal vector fields to C, $\Gamma(Nor(C))$. If $\vec{n}$ is the unit outward normal and s is arc length along C, we put a metric on this via:

$$||a.\vec{n}||^2 = \int_C a(x)^2 ds(x)$$

What does $\mathcal{S}$ 'look like' in this metric? It is an infinite dimensional version of the string theory view of the real world: it is wrapped up more and more tightly in all its higher dimensions. In fact all its sectional curvatures are *non-negative* and go strongly to infinity in the higher frequency dimensions of the local coordinate a. However, the exponential map from the tangent space $T_{[C]}\mathcal{S}$ to $\mathcal{S}$ is locally well-defined as the geodesic equation is a non-linear hyperbolic equation but conjugate points are dense on every geodesic. The global geometry collapses in the sense that the infimum of lengths of paths joining any two curves $[C_1], [C_2]$ is zero. This

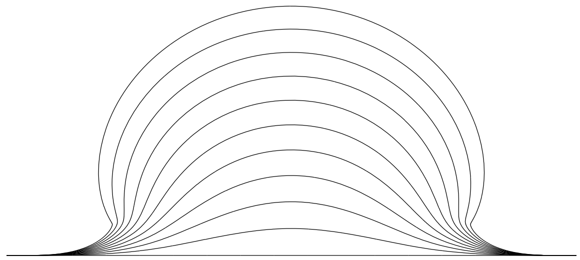

FIGURE 1. A geodesic in the space of plane curves in the L^2 metric. The path starts at the x-axis and moves in the direction of small 'blip'. As the blip enlarges it creates sharper and sharper corners where the curvature goes to infinity so that the geodesic cannot be prolonged.

constellation of properties seems to characterize one possible extreme in the galaxy of infinite dimensional Riemannian manifolds.

The formulas are quite simple and beautiful. The geodesic equation can be written like this. Suppose $[C_t]$ is a path in $\mathcal{S}$. To describe the second derivative of the path, we can first use orthogonal trajectories to map each C_{t_0} to all nearby C_t's. Then a normal vector field $a(x,t).\vec{n}_{C_t}(x)$, $x \in C$ is defined by a function $a(x,t)$, $x \in C_{t_0}$ too. In particular, the tangents to the path $[C_t]$ are given near t_0 by a function of two variables $a(x,t)$, $x \in C_{t_0}, t \approx t_0$. All geodesic equations express the second derivative along a path as a quadratic function of the first derivatives. In our case, this means that the first derivative of a should be a quadratic function of a and at t_0 this is what it is:

$$\frac{\partial a}{\partial t}(x) = \tfrac{1}{2}\kappa_C(x).a(x)^2, \quad \kappa_C = \text{ curvature of } C.$$

Although it may not look like it, this is a hyperbolic equation: you need only rewrite it using local equations like $y = f(x,t)$ for C_t and the curvature κ contributes an f_{xx} term with positive coefficient. This equation does seem to produce singularities in finite time: see figure 1. Details on this and similar metrics can be found in [2, 3, 4].

The formula for curvature is even more elegant. Recall that sectional curvature is just the Riemann curvature tensor $R(a,b,a,b)$ evaluated on an orthonormal basis of a 2-plane, and that this is a quadratic form on the wedge $a \wedge b$ of its two tangent vector arguments. We will write $R(a,b,a,b)$ as $R(a \wedge b)$. So what could be more natural than:

$$R_{\mathcal{S}}(a \wedge b) = \tfrac{1}{2}\int_C (ab' - ba')^2 ds \geq 0$$

where a and b define two tangent vectors in $T_{[C]}\mathcal{S}$. The formula shows that higher frequencies produce more and more positive curvature. In fact, what happens is that path C_t in $\mathcal{S}$ can be shortened by adding high frequency 'wiggles' to the intermediate curves. This is illustrated in figure 2 below.

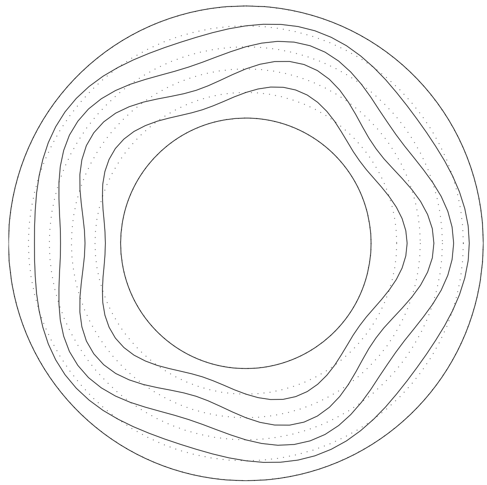

FIGURE 2. The set of all circles with fixed center is a geodesic in the L^2 metric if the radius varies as $t^{2/3}$. However conjugate points are dense on it: Here is a deformation of this geodesic which has a conjugate point when the radius increases by the factor 1.8957.... Beyond that point, it shortens the length of the geodesic.

To summarize: the geodesic equation is a non-linear hyperbolic PDE with well posed initial value problem; the curvature is non-negative, going strongly to infinity at high frequency and with conjugate points dense; and the global metric is identically zero because the infimum of path lengths is zero. This behavior is typical of L^2 metrics.

§2. Positive curvature which is, however, tamer is produced in another elegant situation. This example is due to the work of Laurent Younes [5, 6]. Here we regard the plane as the complex plane. The remarkable idea is to consider the complex square root of the derivative of the curve, i.e. if $t \mapsto f(t) \in \mathbb{C}, t \in \mathbb{R}/2\pi\mathbb{R}$ is the curve, define $g(x) + ih(x) = \sqrt{f'(t)}$. If C is an embedded curve (or more generally any immersed plane curve with odd index), then $g(x + 2\pi) \equiv -g(x), h(x+2\pi) \equiv -h(x)$. The closedness of the curve is expressed by the formula $\int_0^{2\pi} f'(t)dt = 0$ which means that in $L^2([0, 2\pi])$ g and h are orthogonal functions of the same length. We can reverse this process and, starting from such a pair g, h, define a parameterized curve, up to translation, by:

$$x \longmapsto f(x) = \int_0^x (g(x) + i.h(x))^2 dx.$$

The upshot of this ansatz is this: Let $\mathcal{H}$ be the Hilbert space of functions g such that $g(x + 2\pi) \equiv -g(x)$ with norm $||g||^2 = \int_0^{2\pi} g(x)^2 dx$. Let $\mathcal{G}$ be the Grassmannian of 2-planes in $\mathcal{H}$ and let $\mathcal{G}_0$ be the open subset of 2-planes such that there is no x where all functions in the 2-plane vanish. Then using orthonormal bases of these 2-planes as g and h, we find that $\mathcal{G}_0$ is isomorphic to the space of parameterized immersed plane curves of odd index mod translations, rotations and scaling. Not only that but the natural metric on this Grassmannian corresponds to a very natural metric on this space of curves. The tangent space to parameterized curves is given by

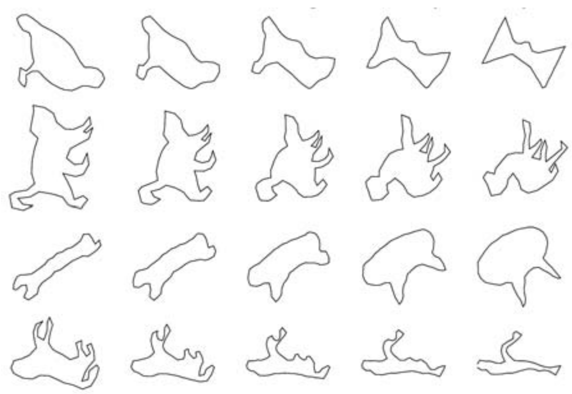

FIGURE 3. Some geodesics in Younes's metric between plane curves representing recognizable shapes. Note how they rotate to make optimal matches, e.g. the tail of the cat with the head of the camel.

all vector fields along the curve, not merely those which are normal, thus, in our case, by a complex valued function along C. The Grassmannian metric turns out to equal the 1-Sobolev norm (with only first derivatives):

$$||a||^2 = \frac{1}{\mathrm{len}(C)} \int_C |a'(x)|^2 ds(x), \quad s = \text{ arc length.}$$

Geodesics and curvature on a Grassmannian are given by quite simple and classical formulas so we also get formulas for these both on this space of *parameterized* curves and on its submersive quotient of *unparameterized* immersed curves both mod translations, rotations and scalings. The geodesic equation is now an integro-differential equation most easily written not in terms of velocity a but in terms of a 'momentum' which is the second derivative $u = -d^2a/ds^2$. Like the Grassmannian itself, these spaces also have entirely non-negative curvature but not so strongly positive that this prevents the Riemannian metric from defining a nice global metric. The space has finite diameter in its global metric and can be completed by adding certain non-immersed curves. Some examples of geodesics in this space are shown in figures 3 and 4. This type of space seems to be the natural infinite dimensional analog of compact symmetric spaces of finite dimension.

§3. A third metric can be put on simple closed plane curves, here modulo translations and scalings, but now with *non-positive* curvature. Interestingly, only one half a derivative is added to the metric in the previous example: it has a Sobolev 3/2 derivative. This is the famous Weil-Peterson metric. It is defined as follows: start with the space of vector fields $v(\theta)$ on the circle and put the WP-norm on it, defined in terms of its Fourier transform by:

$$||v||^2_{WP} = \sum_{n=2}^{\infty} (n^3 - n)|\hat{v}_n|^2.$$

Now SL_2 is a subgroup of the group of diffeomorphisms of the circle with lie algebra consisting of the vector fields $(a + b\cos(\theta) + c\sin(\theta))\frac{\partial}{\partial\theta}$. This is clearly the null

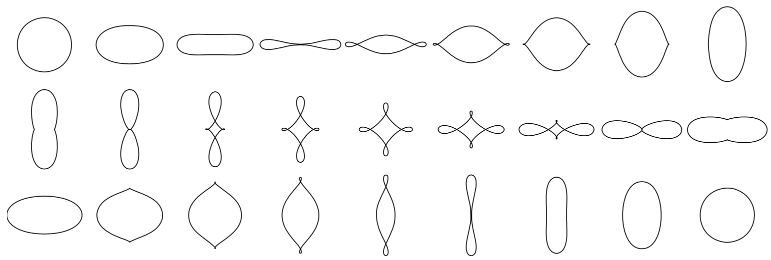

FIGURE 4. If we allow paths to pass through some non-immersed curves, we find many closed geodesics in this metric. This is the simplest example. The path goes from left to right, row by row; a loop flips over in the ellipse-like shapes at the end of the first row and the beginning of the third, hence these are non-immersed.

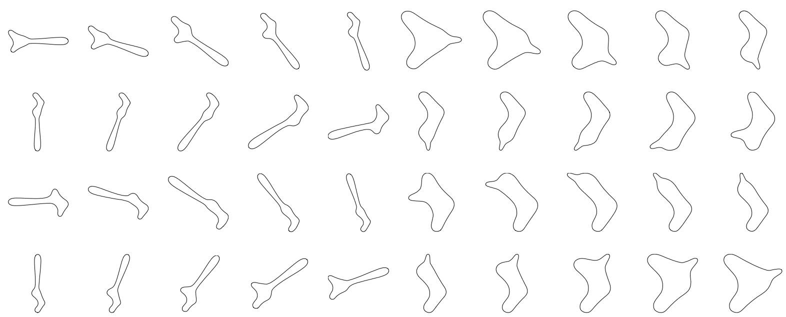

FIGURE 5. A selection of plane curves obtained by composing two diffeomorphisms corresponding to (i) a boomerang-like shape with base point in the middle and (ii) a finger-like shape with base point near one end respectively. In each panel of 15 curves, the data $\vec{t}$ is varied or, equivalently, a variable rotation is added in the middle of the composition.

space of the above WP norm and since – miraculously – the WP-norm is also invariant under the adjoint action of SL_2, this norm extends by right translations to an invariant Riemannian metric on the coset space $SL_2\backslash\mathrm{Diff}(S^1)$. Now the final link: this coset space is isomorphic to the space of simple closed plane curves mod translations and scalings. This comes 'welding': given a diffeomorphism φ, attach two unit disks to each other along their boundaries using the twist φ. The result is a simply connected compact Riemann surface, hence it must be conformal to the sphere. The image of the welded common boundary is our curve. For details, see [7, 9].

One of the remarkable consequences of this construction is that it defines an operation of composition between plane curves. The welding operation also defines a bijection between the group $\mathrm{Diff}(S^1)$ itself and triples $(C, P, \vec{t})$ where C is a simple closed plane curve, P a base point inside C and $\vec{t}$ is a distinguished ray at the base point, all modulo translations and scalings. Thus there is a law of composition of such triples. Some examples are shown in figure 5.

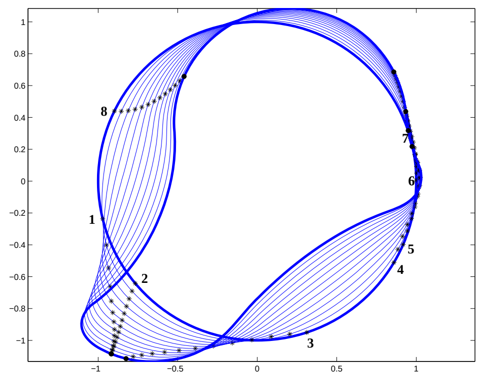

FIGURE 6. A geodesic from the unit circle to a duck like shape using an 8-Teichon. The figure is due to S. Kushnarev [10].

This metric is the closest to the standard metric on $\mathbb{R}^n$ because (a) it is invariant under the transitive action of a group, here Diff(S^1) and (b) it is quite flat in high frequency dimensions because the Ricci curvatures (which are the sum of sectional curvatures $R(a \wedge b_i)$ where $\{b_i\}$ are an orthonormal basis of $a^\perp$ for variable a) are known to be finite. It is also a complete complex Kähler-Hilbert manifold and has unique geodesics between any two points [7, 8]. The metric can also be defined using potential theory which embeds the curve in field lines and thus endows its interior and exterior with a rich additional structure. The geodesic equation is an integro-differential variant of Burger's equation involving the (periodic) Hilbert transform. Among geodesics on this space, there is a special class of soliton-like geodesics, which Daryll Holm named 'teichons'. They are the geodesics generated by vector fields v dual in the WP norm to sums of delta functions, i.e.

$$\langle v, u\rangle_{WP} = \sum_i p_i u(\theta_i), \text{ for all } u$$

for some p_i, θ_i. An example of a teichon is shown in figure 6.

§4. The final example is much more general and deals with the full diffeomorphism group of $\mathbb{R}^n$. Arnold's curvature formula for volume preserving diffeomorphisms was significantly more complicated than anything in the above examples. In his case, there are both positively and negatively curved sections and this also seems to happen for Riemannian manifolds constructed from any higher order Sobolev type metrics on diffeomorphism groups. I would not be surprised if at some point understanding these more complex curvature formulas gives new insight into the unsolved problems of fluid flows.

The situation that my group has studied most intensively is the metric induced on 'landmark space', that is simply the space $\mathcal{L}_{n,N}$ of distinct N-tuples of points in $\mathbb{R}^n$. Fixing a base N-tuple, we get a submersive map from Diff($\mathbb{R}^n$) to $\mathcal{L}_{n,N}$. We

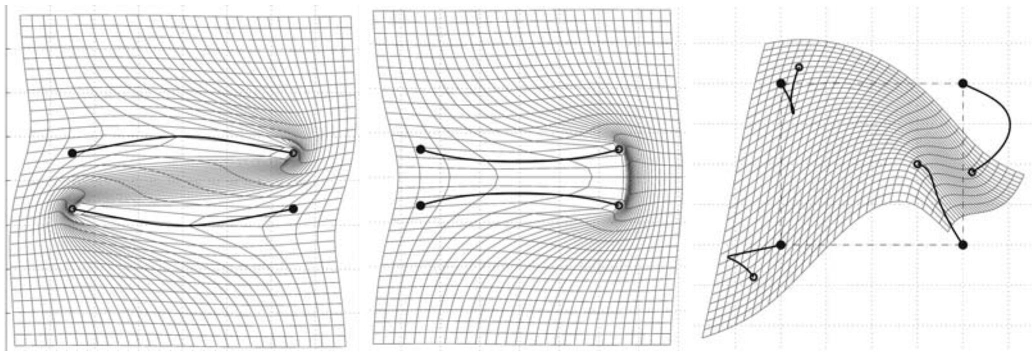

FIGURE 7. Three geodesics on the landmark spaces $\mathcal{L}_{2,2}, \mathcal{L}_{2,2}, \mathcal{L}_{2,4}$ respectively, plus the induce diffeomorphism of the ambient plane. Note how points moving in the same (resp. opposite) direction are drawn together (resp. pushed apart). In the four point case, this causes some complex gyrations. These figures are from the thesis of M. Micheli [11]

may put a Sobolev norm on vector fields X,

$$||X||^2 = \int_{\mathbb{R}^n} \langle X, LX\rangle dx_1 \cdots dx_n$$

where L is a positive definite self-adjoint operator, e.g. $L = (I - \triangle)^s$. This defines a metric on the group of diffeomorphisms provided that L has enough derivatives. In fact, we want the finiteness of the metric to force the diffeomorphisms to be C^1. Then we get an induced Riemannian structure on the quotient space $\mathcal{L}_{n,N}$. It has a simple form. If G is the Green's function associated to L, $\{P^1, \cdots P^N\} \in \mathcal{L}_{n,N}$ and v^a is a vector at P^a, then the metric is:

$$||\{v^1, \cdots, v^N\}||^2 = \sum_{1\le a,b\le N} (G^{-1})_{ab}\langle v^a, v^b\rangle, \quad G_{ab} = G(P^a - P^b).$$

Arguably this is the most natural class of metrics to put on landmark space.

The geodesic equation on landmark space is quite elegant. To any geodesic there is a natural set of momenta u_a for which a geodesic is a solution of:

$$\frac{du_a}{dt} = -\sum_b \nabla G(P^a - P^b)\langle u_a, u_b\rangle$$

$$\frac{dP^a}{dt} = \sum_b G(P^a - P^b)u_b$$

These equations create a world governed by a weird sort of physics in which points moving together attract and points moving in opposite directions repel and occasionally one even gets planetary systems. Since $\mathcal{L}_{n,N}$ is a submersive quotient of the diffeomorphism group, geodesics in $\mathcal{L}$ lift to horizontal geodesics in the group. So these geodesics induce warpings of the ambient Euclidean space. In fact any geodesic in the diffeomorphism group can be approximated by one of these landmark geodesics if we take enough landmark points. Some examples are shown in Figure 6.

The sectional curvature of landmark space has four terms, similarly to Arnold's formula for the curvature of Lie groups [1]. The reason for this seems to be that it reflects several different modes of interactions of the points. There are, in particular, at least two 'causes' of positive curvature. One is seen in the middle panel of Figure 4: when two points must move in a similar direction, it saves energy for them to come close to each other. But if the distances are in a certain range, there will be two geodesics joining the pair at the initial position on the left and the same pair translated to the final position on the right. One geodesic has the points moving nearly independently and nearly parallel and the other has them first coming close, then moving together and finally moving apart again. This non-uniqueness causes positive curvature.

Another cause of positive curvature occurs in dimension three and higher. Suppose two points are to be interchanged, the first moving to the position of the second and the second to that of the first. Since the distance is infinite if they were to move directly towards each other, they must move around each other and there are many planes in which to do this. We show in [12] that when only two points have momenta, these are, in a sense, the only ways positive curvature can arise.

But negative curvature is caused all the time by the turbulence caused by landmark point motion. Take the situation where a single point P is moving with non-zero momentum but that there are many other landmark points around it with zero momentum. These extra points are dragged along, compressed together in front of P. If P moves from A to B, we wind up with a configuration $C(B)$ of the whole set of landmarks. Take $B_1 \neq B_2$. Then what will the geodesic from $C(B_1)$ to $C(B_2)$ look like? You can't just put momentum on P because you need to move the points bunched up near B_1 back apart and create a new bunch near B_2. The only way to do this is unwind the mess you made in one geodesic and recreate the new mess in the second. This is negative curvature: to connect the endpoints of two trips, it is better to go nearly back home. For details, see [12].

Anyway, the somewhat daunting formula for sectional curvature is this. Let $v_1 = \{v_1^a\}$ and $v_2 = \{v_2^a\}$ be two tangent vectors to $T\mathcal{L}$ at some point $\{P^a\}$. Let v_i be extended to a vector field on $\mathbb{R}^n$ by its horizontal lift to the diffeomorphism group. Let $v_1^\flat$ and $v_2^\flat$ be the co-vectors dual to the v's. Then the numerator of sectional curvature is given by:

$$
\begin{aligned}
&R(v_1 \wedge v_2) = R_1 + R_2 + R_3 + R_4 \\
&R_1 = \tfrac{1}{2} \sum_{a \neq b} \left[(v_2^\flat)_a \otimes \delta^{ab} v_1 - (v_1^\flat)_a \otimes \delta^{ab} v_2 \right] \cdot H^{ab} \cdot \left[(v_2^\flat)_b \otimes \delta^{ab} v_1 - (v_1^\flat)_b \otimes \delta^{ab} v_2 \right], \\
&R_2 = \left\langle D_{11}, F_{22} \right\rangle - \left\langle D_{12} + D_{21}, F_{12}) \right\rangle + \left\langle D_{22}, F_{11} \right\rangle, \\
&R_3 = \left\| F_{12} \right\|^2_{T^*\mathcal{L}} - \left\langle F_{11}, F_{22} \right\rangle_{T^*\mathcal{L}}, \\
&R_4 = -\tfrac{3}{4} \| D_{12} - D_{21}) \|^2_{T\mathcal{L}}, \\
&\qquad \text{where } \delta^{ab} v = v^a - v^b, \text{ and } C_{ab}(v) = \left\langle \delta^{ab} v, \nabla G(P_a - P_b) \right\rangle \text{ for any } v \\
&\qquad \text{and } D^a_{ij} = \sum_{b \neq a} C_{ab}(v_i)(v_j^\flat)_b \in T\mathcal{L}, \\
&\qquad \text{and } (F_{ij})_a = \tfrac{1}{2} \left(D_a v_i \cdot (v_j^\flat)_a + D_a v_j \cdot (v_i^\flat)_a \right) \in T^*\mathcal{L}, \ (D_a = \text{deriv. at } P^a) \\
&\qquad \text{and } H^{ab} = I \otimes D^2 G(P^a - P^b)
\end{aligned}
$$

The term R_4 above is the main cause of the turbulence related negative curvature: it is the only term which involves points with no momentum of their own.

It is natural to generalize this formula to get more insight into it. A paper is under preparation analyzing the spaces of submanifolds of any type in any fixed ambient finite dimensional manifold M with respect to a very general Sobolev-type metric on the group of diffeomorphisms of M.

References

[1] V. Arnold, Sur la géométrie diffrentielle des groupes de Lie de dimension infinie et ses applications l'hydrodynamique des fluides parfaits, *Annales de l'institut Fourier*, 16:319-36, 1966.

[2] P. Michor and D. Mumford, Riemannian Geometries on Spaces of Plane Curves, *J. of the Europ Math. Society*, 8:1-48, 2006.

[3] P. Michor and D. Mumford, Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms, *Documenta Mathematica*, 10, 2005.

[4] P. Michor and D. Mumford, An overview of the Riemannian metrics on spaces of curves using the Hamiltonian Approach, *Applied and Computational Harmonic Analysis*, 23:74-113, 2007.

[5] L. Younes, Computable elastic distances between shapes, *SIAM J. Appl. Math.* 58:565-586, 1998.

[6] P. Michor, D. Mumford, J. Shah and L. Younes, A metric on shape space with explicit geodesics, *Rendiconti Lincei – Matematica e Applicazioni*, 19:25-57, 2009.

[7] T. Takhtajan and L.-P. Teo, *Weil-Petersson Metric on Universal Teicmüller Space*, Memoirs of the AMS, **86**, 2008.

[8] F. Gay-Balmaz, *Infinite Dimensional Flows and the Universal Teichmüller Space*, PhD thesis, École Polytechnique de Lausanne, 2009.

[9] D. Mumford and E. Sharon, 2D-Shape Analysis using Conformal Mapping (with Eitan Sharon), *Int. J. of Comp. Vision* 70:55-75, 2006.

[10] S. Kushnarev, Teichon: Soliton-like Geodesics on Universal Teichmüller Space, *J. Exp. Math.*, 18:325-336, 2009.

[11] M. Micheli, *The Differential Geometry of Landmark Shape Manifolds*, PhD thesis, Div. of Appl. Math., Brown Univ., 2008.

[12] M. Micheli, P. Michor and D. Mumford, Sectional Curvature in terms of the Cometric, with Applications to the Riemannian Manifolds of Landmarks, submitted.