Statistical analysis of shapes

Mads and Anders

November 7, 2018

1 Concepts of Riemannian geometry

1.1 Notation

In the following M denotes a smooth manifold and TM is the tangent bundle of M and $\mathcal{T}(M)$ denotes the space of all vector fields on M. For $I \subset \mathbb{R}$, $\gamma : I \to M$ is a curve in M, i.e. a smooth map.

1.2 Connections

To consider the geodesic distance between two points in a manifold, geodesics need to be defined in a coordinate-invariant way such that the distance is independent of the coordinate charts. One property of geodesics in a Euclidean space, straight lines, is that they have acceleration 0. In order to make sense of acceleration of a curve in a manifold, we need to be able to compute "differences" between tangent spaces along the curve. *Connections* are exactly a way of making computations between tangent spaces possible - they allow us to differentiate vector fields along curves.

Since our use of connections is to define geodesics, we define connections in the tangent bundle of a manifold (instead of defining them generally on smooth sections of vector bundles) follwing chapter 4 of RiemannLee.

Definition 1. A connection in TM is a map

$$\nabla : \mathcal{T}(M) \times \mathcal{T}(M) \to \mathcal{T}(M),$$

written $(X, Y) \mapsto \nabla_X Y$ satisfying (for f, $g \in C^\infty(M)$ and a, $b \in \mathbb{R}$);

$$\begin{aligned}
&a)\ \nabla_{fX_1+gX_2} Y = f\nabla_{X_1} Y + g\nabla_{X_2} Y && \text{(linearity over } C^\infty(M) \text{ in } X) \\
&b)\ \nabla_X(aY_1 + bY_2) = a\nabla_X Y_1 + b\nabla_X Y_2 && \text{(linearity over } \mathbb{R} \text{ in } Y) \\
&c)\ \nabla_X(fY) = f\nabla_X Y + (Xf)Y && \text{(product rule)}
\end{aligned}$$

In accordance with connections allowing "differences" between tangent spaces, $\nabla_X Y$ is called the *covariant derivative of Y in the direction of X*. Note that the product rule for connections is identical to the product rule of derivations. To use connections to derivate along curves, we need the definition of a *vector field along a curve*, which is a smooth map $V : I \to TM$ such that $V(t) \in T_{\gamma(t}M$ for all $t \in I$. The prime example of a vector field along a curve is its velocity, $\dot{\gamma}(t) \in T_{\gamma(t}M$, which acts on functions, $f \in C^\infty(M)$, by

$$\dot{\gamma}(t)f = \frac{\mathrm{d}}{\mathrm{dt}}(f \circ \gamma)(t).$$

We denote by $\mathcal{T}(\gamma)$ all vector fields along γ. To define geodesics all we now need is to define what is means to take the covariant derivative of $V \in \mathcal{T}(\gamma)$ along γ. This covariant derivative is noted $D_t V$ and is has the following properties.

Lemma 2. *Let ∇ be a linear connection on M. For each $\gamma : I \to M$, ∇ determines a unique operator*

$$D_t : \mathcal{T}(\gamma) \to \mathcal{T}(\gamma),$$

satisfying (for f, $g \in C^\infty(I)$ and a, $b \in \mathbb{R}$);

a) $D_t(aV + bW) = aD_tV + bD_tW$ *(linearity over $\mathbb{R}$*

b) $D_t(fV) = \dot{f}V + fD_tV$ *(product rule)*

c) If V is extendible, then for any extension $\tilde{V}$ of V, $D_tV(t) = \nabla_{\dot{\gamma}(t)}\tilde{V}$.

Proof. Proof of Lemma 4.9 in RiemannLee □

V is said to be extendible if it can be constructed by any vector field on M, $\tilde{V}$ by letting $V(t) := \tilde{V}_{\gamma(t)}$. This is not always the case; if V is the velocity of an intersecting curve γ with different covariant derivative at the intersection times. The covariant derivative of the velocity of a curve is now used to define a geodesic.

Definition 3. Let ∇ be a linear connection on M and γ a curve in M. The acceleration of γ is $D_t\dot{\gamma}(t)$. If this vector field is zero, $D_t\dot{\gamma}(t) \equiv 0$, then $\gamma(t)$ is a geodesic with respect to ∇

It follows from Theorem 4.10 in RiemannLee that for any manifold, M, with a linear connection, for any $p \in M$ and $V \in T_pM$ and $t_0 \in \mathbb{R}$ there exists an unextendable geodesic, $\gamma_V : I \to M$, with $\gamma(0) = p$ and $\dot{\gamma}(0) = V$. The geodseic is called the (maximal) geodesic with initial value p and initial velocity V.

In this construction of geodesics the only necessary structure of M is that it should be a smooth manifold. When M is also equipped with a Riemannian metric, making M a Riemannian manifold, the choice of connection (determining the geodesics) should in some way respect the metric. Geodesics resulting from this specific choice of connection are called *Riemannian geodesics*.

1.3 Riemannian Geodesics and the Exponential Map

Let M be Riemannian manifold with metric g. To define Riemannian geodesics, we must first choose a specific connection on M with two properties - *compatability w.r.t. g* and *symmetric* (these properties arise when trying to generalize the tangential conenction of a manifold submersed in $\mathbb{R}^n$).

Definition 4. A conneciton on M is *compatible with* g if the product rule

$$\nabla_X(Y,Z) = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$$

holds for all vector fields, X, Y and Z.

By Lemma 5.2 in RiemannLee this condiiton is equivalnt with

$$\frac{d}{dt}\langle V, W\rangle = \langle D_t V, W\rangle + \langle V, D_t W\rangle,$$

for V, W being vector fields along any curve γ.

The defintion of a symmetric connection involves the Lie bracket of two vector fields. If we think of vector fields on M, X, Y, as derivtations acting on $C^\infty(M)$, then the Lie brakcet of X and Y, $[X, Y]$, is the vector field (derivation) which acts on $f \in C^\infty(M)$ by

$$[X,Y](f) := X(Y(f)) - Y(X(f)),$$

where $X(f) \in C^\infty(M)$ is the function which evaluated at p is the derivative of f at p in the direction of $X(p)$.

Definition 5. A connection on M is *symmetric* if

$$\nabla_X Y - \nabla_Y X \equiv [X, Y].$$

(Note that interchanging X and Y makes sense, since $[X, Y] = -[Y, X]$.) A symmetric connection is also called *torsion free*, which corresponds to the vector fields along any curve not being "twisted" when they are parallel transported along the curve. By the Fundamental Theorem of Riemannian Geometry, given any Riemannian manifold (M, g), there exists a unique connection, ∇, on M that is symmetric and compatible with M. This connection is called the Riemannian connection of the Levi-Civita connection, and geodesics with respect to this connection are called Riemannian geodesics. Since this choice of connection is unique, geodesics in a Riemannian manifold will always mean Riemannian geodesics. Such geodesics can be used to define the exponential map, exp : $TM \to M$.

Definition 6. Given a point p in a Riemannian manifold, M, and a vector $v \in T_pM$, the *exponential map* is defined by $\exp_p(v) = \gamma_v(1)$ where γ is the unique geodesic with $\gamma(0) = p$ and $\dot\gamma(0) = v$.

The exponential map pushes a point $p \in M$ a unit distance in the direction of v along the geodesic γ. Since $\dot\gamma(0) = v$, the differential of $\exp_p$ at p is v, and

so $\exp_p$ is a local diffeomorphism by the implicit function theorem. Given two points $p, q \in M$, the inverse of the exponential map, $\exp_p^{-1}(q) = \log_p(q) = \overrightarrow{pq}$, yields the element of $v \in T_pM$ for which $\gamma_v(1) = q$

A Riemannian manifold is said to be *geodesically complete* if every un-exentedable geodesic is defined on all of $\mathbb{R}$, and it follows from the Hopf-Rinow theorem that a manifold is geodesically complete if and only if, there exists a $p \in M$ such that exp is defined on all of T_pM.

If the geodesic γ_v is defined on all of $\mathbb{R}$, one can investigate for which values of t, the extended geodesic, $\gamma_v(t) = \exp_p(tv)$ is still a geodesic from p to $\exp_p(tv)$. If γ_v is a geodesic up to time t_0 and not after t_0, then t_0 is called a *cut point*. The set of all cut points of all geodesics starting at p is called the *cut locus* and is denoted $C(p)$.

1.4 Curvature

When defining statistical propoerties such as mean and variance on Riemannian manifolds, measuring geometric propoerties of the manifold is needed. It is therefore of interest to determine on which manifolds these measurements are identical - that is to determine which manifolds are (locally) isomorphic. One way of determining this is to find a local invariant property of a manifold which is preserved by isometries (such that measurements of length are preserved). This property will be the *curvature*.

Let $M = \mathbb{R}^n$ equipped with the Euclidean metric and consider a vector field Z. Given two other vector fields, X and Y, we can now differentiate Z first along X and then along Y by using the Riemannian connection;

$$\nabla_Y \nabla_X Z,$$

and in the opposite order by $\nabla_X \nabla_Y Z$. If $R^n = R^2$ and X and Y where just vector fields corresponding to local coordinates then, by commutativity of second order derivatives,

$$\nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z = \nabla_Y(\partial_1 Z^k \partial_k) - \nabla_X(\partial_2 Z^k \partial_k) = \partial_2 \partial_1 Z^k \partial_k - \partial_1 \partial_2 Z^k \partial_k = 0.$$

But if X and Y are arbitrary vector fields, this identity does not necessarily hold, since

$$\nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z = XYZ^k \partial_k - YXZ^k \partial_k = (XYZ^k - YXZ^k)\partial_k,$$

where $XYZ^k = X\left(Z\left(Z^k\right)\right)$ to ease notation. The action of the Lie bracket of X and Y is recognized in the parenthesis, and thus for $\mathbb{R}^n$ the following identity holds;

$$\nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z = \nabla_{[X,Y]} Z. \tag{1}$$

Since this identity depends on the Levi-Civita connection, it holds for all manifolds isometric to $\mathbb{R}^n$, and manifolds for which 1 holds will be called *flat* manifolds. To define the curvature of a manifold is then to determine how "un-flat" the manifold is, by considering the *curvature transformation*, $R : \mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M) \to \mathcal{T}(M)$, defined by

$$R(X,Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[X,Y]} Z,$$

which is identically zero for flat manifolds. The curvature transformation can then be used to determine the curvature of a vector field by defining the *curvature tensor*;

Definition 7. The curvature on a Riemannian manifold is

$$Rm(X,Y,Z,W) = \langle R(X,Y)Z, W \rangle,$$

with $\langle \cdot, \cdot \rangle$ being the inner product determined by the Riemannian metric.

2 Mean and variance

In order to perform statistics on shapes we must first try to define central statistical concepts on manifolds. In this section we focus on a geodsically complete Riemannian manifold, (M, g), of dimension n, and present ways of defining the mean, variance and covariance of M-valued random variables. Given an underlying probability space, $(\Omega, \mathcal{F}, P)$, a M-valued random variable is a $\mathcal{F}/\mathcal{B}(M)$ measurable map, $X : \Omega \to M$, and we denote by $x = X(\omega)$ a realization of X on M.

In order to perform statistics on M we need to construct a measure on M. This measure is induced by the metric g in the following way. Let $x = (x^1, \ldots, x^n)$ be representation of $x \in M$ in local coordinates, and let $\frac{\partial}{\partial x} = (\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n})$ be the corresponding basis of $T_x M$. The metric g is then expressed in this basis by the matrix $G = [g_{ij}(x)]$ where $g_{ij}(x) = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$. The measure on M is then defined by $dM(x) = \sqrt{|\det G(x)|} dx$. X is said to have density p_X w.r.t. dM if

$$P(X \in \mathcal{A}) = \int_{\mathcal{A}} p_X(y) dM(y),$$

holds for all $\mathcal{A} \in \mathcal{B}(M)$ and if the integral over M is equal to 1. Here p_X is a density in the usual sense. It is a real-valued, positive and integrable function. If π is a chart of the manifold, then $r := \pi(X(\omega))$ defines a random vector with density, ρ_r, w.r.t to the Lebesgue measure given by $\rho_r(y) = p_X(y)\sqrt{|\det G(y)|}$. If $\varphi : M \to \mathbb{R}$ is a $\mathcal{B}(M)/\mathcal{B}(\mathbb{R})$-measurable map, then $\varphi(X)$ deifnes a real-valued random variable for which the expection is

$$\mathbb{E}(\varphi(X)) = \int_M \varphi(y) p_X(y) dM(y).$$

Unfortunately, we cannot define the expectation of M-valued random variables in a similar manner, since the real-valued integral does not generalize to an integral with values on M. Instead we generalize the notion of mean value by first defining the variance of a M-valued random variable and then defining the so-called *Frechet means* as minimizers of the variance (this is just one possible way of generalizing).

Definition 8. Let X be a M-valued random variable with density p_X. Given a point $y \in M$, the *variance* of X is then

$$\sigma_X^2(y) = \mathbb{E}(\text{dist}(y, X)^2) = \int_M \text{dist}(y, z) p_X(z) dM(z).$$

Here the distance between two points $x, y \in M$ is the infimum of the lengths of all paths in M from x to y, with the length of a path $c : [0, 1] \to M$ defined by $L(c) = \int_0^1 g_{c(t,\cdot)}(c_t, c_t)dt$.

Definition 9. Let X be a M-valued random variable with density p_X. If $\sigma_X^2(y)$ is finite for all $y \in M$, we define *Frechet mean points* of X as all points in M minimizing $\sigma_X^2(y)$;

$$\mathbb{E}(X) := \arg\min_{y \in M} \sigma_X^2(y).$$

If a mean point $\bar{x}$ exists, the variance of X is defined by $\sigma^2(X) := \sigma_{\bar{x}}^2(X)$. We can further define the *median points* of X as all minimizers of $\mathbb{E}(\text{dist}(y, X))$.

Note 10
Given a series of measurements $X_1, \ldots, X_n$ seen as realizations of X, we define the empirical mean points of X to be the minimizers of

$$\frac{1}{n} \sum_{i=1}^{n} \text{dist}(y, X_i)^2,$$

and the empirical variance is, as before, defined as the variance of X evaluated at a minimizer. Note that this corresponds to the standard notion of the empirical mean as the minimizer of the sum of squares and the empirical variance as the sum of squared deviations from this mean.

The most apparent question is now whether a mean point exists for X and if it is unique. To give conditions for existence and uniqueness, we follow Riemannian Center of Mass and Mollifier Smoothing and define *Riemannian centers of mass* as local minimizers of $\sigma_X^2(y)$. The Riemannian centers of mass have the added benefit of encoding more information about the distribution of X than the mean points, since the centers of mass represent local maxima of the distribution of X (where the mean points only represent the global maximum). The definition of these centers of mass amounts to finding local extrema of $\sigma_X^2(y)$, in which the following theorem due to Intrinsic Statistics on Riemannian Manifolds plays an important part.

Theorem 11. *Let X be a M-valued random variable with density p_X. If $\sigma_X^2(y) < \infty$ and the image measure of the cut locus is zero, $X(P)(C(y)) = 0$, then $\sigma_X^2(y)$ is differentiable with*

$$(grad\,\sigma^2)(y) = -2\mathbb{E}(\overrightarrow{yx}) = -2\int_{M/C(y)} \overrightarrow{yz} p_X(z) dM(z).$$

The theorem ensures differentiability of $\sigma_X^2(y)$ in points where the cut locus has measure zero, and in this case the extrema of σ^2 are points where (grad $\sigma^2)(y) = 0$. If the cut locus has positive measure, the variance may still attain an extremum. This leads to the following characterization of Riemanninan centers of mass.

Corollary 12. *Let $\mathcal{A}$ be the set of points for which the cut locus has non-zero probability. If $\sigma_X^2(y) < \infty$ for all $y \in M$, then a necessary condition for $\bar{x}$ to be a Riemannian center of mass is $x \in \mathcal{A}$ or $\mathbb{E}(\overrightarrow{\bar{x}x}) = 0$ for $\bar{x} \notin A$.*

Since the Riemannian centers of mass are local minimizers of the variance, the set of mean points is included in the set of centers of mass. If there is only one Riemannian center of mass it therefore follows that it must be the unique mean point. The following corollary then gives uniqueness of centers of mass not in $\mathcal{A}$ for a class of manifolds (Hadamard manifolds).

Corollary 13. *Let M be a simply connected, complete manifold with non-positive Riemannian curvature, and let X be a random variable with values in M and finite variance. Then there exists one and only one Riemannian center of mass characterized by $\mathbb{E}(\overrightarrow{\bar{x}x}) = 0$. If the cut locus has measure zero everywhere, then this point $\bar{x}$ must be a mean point.*

If we want to assert uniqueness for a larger class of manifolds, we have to make assumptions not only on the curvature of the manifold but on the support of the densities. Compact support of the densities is actually not sufficient - the support of the densities have to be contained in a *regular geodesic ball*.

Definition 14. A ball, $B(x, r) = \{y \in M \mid \text{dist}(x, y) < r\}$, is geodesic if $B(x, r) \cap C(x) = \emptyset$ and it is regular if $2r\sqrt{\kappa} < \pi$ where κ is maximum of the Riemannian curvature in $B(x, r)$.

Note that the assumption on the curvature of M is no longer global but local. We have thus replaced the global assumption on the curvature with a local one, but have now restricted ourselves to densities with compact support on locally well-behaved parts of the manifold. Under these assumptions the following results hold (Reference), where the case of $\bar{x} \in A$ is not possible since the balls, on which the densities have support, are geodesic.

Theorem 15. *Let X be a M-valued random variable with density p_X. If the support of p_X is contained in a regular geodesic ball, then there exists one unique Riemannian center of mass on this ball. This center of mass must be the unique mean point of X.*

Theorem 16. *Let X be a M-valued random variable with density p_X. If the support of p_X is contained in a regular geodesic ball, and the ball with twice the radius is also a regular and geodesic, then $\sigma_X^2(y)$ is convex and has a unique critical point, $\bar{x}$, on the ball. This point must be a minimizer and thus the unique mean point of X*

The preceeding corollary and two theorems give us uniqueness and existence statements of mean points for specific cases. One can then wonder if it is possible to relax the global curvature assumption in the corollary or the assumptions on the well-behaviour of the curvature in the domain of the support of the density. As shown in The propeller: a counterexample to a conjectured criterion for the existence of certain harmonic functions some assumptions on the curvature is needed even if there is a unique minimizing geodesic joining any two points.

3 The manifold of curves

A 2-dimensional shape can be thought of as a closed (smooth) curve in $\mathbb{R}^2$. Thus, we want to define a manifold structure on the space of these curves. Doing this mathematically correct is rather technical because we need to consider quotient spaces of infinite dimensional manifolds. In our exposition we shall to a large extend "define our way out of this" by, for example, defining tangent vectors of curves instead of deducing how these look like from the formal definition of the underlying manifold. We shall motivate our definitons geometrically and then deduce some properties from these definitons – properties, which can also be deduced from the formal definitions. At the end of this section we briefly address what we miss with our more informal treatment.

A shape in $\mathbb{R}^2$ can either be thought of as a parametrized object, e.g., as a function $\mathbb{S}^1 \ni \theta \mapsto c(\theta) \in \mathbb{R}^2$; but it can also be thought of as an unparametrized object, e.g., the *image* of such a function $\mathrm{Im}(c) \subset \mathbb{R}^2$ (as illustrated in Figure 1). We impose some smoothness structure on the curves and define the spaces we want to consider formally as

$$\begin{aligned} \mathrm{Imm} &:= \mathrm{Imm}(\mathbb{S}^1, \mathbb{R}^2) := \{...?\}, \\ \mathcal{I} &:= \mathcal{I}(\mathbb{S}^1, \mathbb{R}^2) := \mathrm{Imm}(\mathbb{S}^1, \mathbb{R}^2)/\mathrm{Diff}(\mathbb{S}^1). \end{aligned}$$

Imm is the space of *immersion* of the unit circle into $\mathbb{R}^2$, and $\mathcal{I}$ is then this space modulo reparametrization, i.e., we identify two object $q, p \in \mathrm{Imm}$ in $\mathcal{I}$ if $q = p \circ \varphi$, where $\varphi \in \mathrm{Diff}(\mathbb{S}^1)$ is a diffeomorphism on the unit circle.

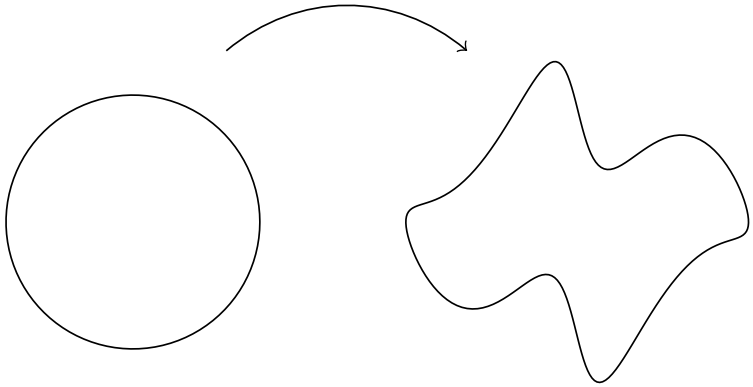

Figure 1: A simple shape. We can think of a shape as the whole mapping, or simply as the subset to the right.

When our interest is upon the *shape*, we are not really interested in the underlying parametrization of this shape, and so we are mostly interested in the space $\mathcal{I}$. However, it is easiest to construct a manifold structure on Imm and then deduce one on $\mathcal{I}$, so we start by considering the parametrized space of immersions.

3.1 The manifold of parametrized curves – $\mathrm{Imm}(\mathbb{S}^1, \mathbb{R}^2)$

The first structure we want to impose on our manifold of curves is tangent spaces and tangent vectors to our elements in the space.

For ordinary finite dimensional manifolds M we have so far worked with the definition of tangents vectors as *derivatives*. This is a rather abstract construction, which, however, turns out to be nice to work with. Fortunalety, we know that this definition corresponds to the more geometrically intuitive definition of tangent vectors at a point $m \in M$ as derivatives of paths going through m. We shall use this a motivation for our definition of tangent vectors to points in Imm.

Consider a point $c \in \mathrm{Imm}$ and a path in Imm defined around 0 that goes through the point c. This path is a map

$$[-\varepsilon, \varepsilon] \ni t \mapsto q(t, \cdot) := (\theta \mapsto q(t, \theta)) \in \mathrm{Imm}, \quad q(0, \theta) = c(\theta), \tag{2}$$

i.e., for each t we get a parametrized curve, and at 0 we get the curve c. We

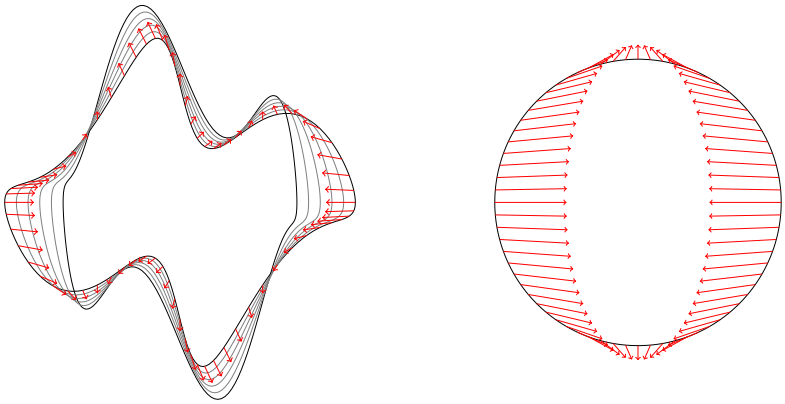

Figure 2: Illustration of tangent vectors in Imm. The left illustration shows a path in the space of curves and how to obtain a tangent vector from this. The right illustration shows how to think of this tangent vector as a vector field on the circle.

can also think of this a (smooth) map

$$[-\varepsilon, \varepsilon] \times \mathbb{S}^1 \ni (t, \theta) \mapsto q(t, \theta) \in \mathbb{R}^2.$$

As in the finite dimensional case we can now take the *time derivate* of our path and evaluate this at 0; of course, this is technically not obvious, as our constructed path maps into a function space, but we shall here simply take this derivative to be understood pointwise:

Definition 17. The tangent space $T_c \operatorname{Imm}$ to an element $c \in \operatorname{Imm}$ consists of all function $h \colon \mathbb{S}^1 \to \mathbb{R}^2$ such that

$$h(\theta) = q_t(0, \theta) := \frac{\partial}{\partial t}\bigg|_{t=0} q(t, \theta),$$

where q is some path passing through c at 0 (as defined in (2)).

By this definition, we can essentially think of tangent vectors in Imm as vector fields on the circle. Figure 2 illustrates the idea behind the definition.

We want to make our space of curves into a Riemmanian manifold so the next step is to impose a metric [why do we/others call it a metric – shouldn't it be an inner product?] on the tangent spaces. Because the tangent spaces are function spaces, the most obvious metric to use is some version of the L^2 metric.

Definition 18. The G *metric* or G_c *metric* at the point $c \in \operatorname{Imm}$ is defined as

$$\mathrm{G}_c(h, k) := \left(\int_{\mathbb{S}^1} \langle h(\theta), k(\theta) \rangle \, |c_\theta| \, \mathrm{d}\theta \right)^{\frac{1}{2}},$$

$h, k \in T_c B = C^\infty(\mathbb{S}^1, \mathbb{R}^2)$.

Adding the parameter derivative c_θ makes the metric invariant to reparametrizations which is essential when we want to let this fall down as a metric on $\mathcal{I}$. From this definition it is straightforward to define a notion of *length* of a path in Imm which again allows us to define a *distance* between two point in our space of curves. By defintion, the pointwise time derivative at t of a path $t \mapsto q(t, \cdot)$ in Imm is a tangent vector to the curve $q(t, \cdot)$, so the following definition makes sense.

Definition 19. Let $t \mapsto q(t, \cdot)$ be a path in Imm with $t \in [0, 1]$. The *length* of the path q with respect to the G metric is

$$L(q) := \int_0^1 \mathrm{G}_{q(t,\cdot)}(q_t, q_t)\,\mathrm{d}t = \int_0^1 \left(\int_{\mathbb{S}^1} \|q_t\|^2 |q_\theta|\,\mathrm{d}\theta \right)^{\frac{1}{2}} \mathrm{d}t.$$

The *geodesic distance* [geodesic?] between to curves $b, c \in \mathrm{Imm}$ (with respect to the G metric) is

$$D(b, c) := \inf_{q \in \mathcal{Q}} L(q),$$

where $\mathcal{Q}$ denotes all paths q, such that $q(0, \cdot) = b$ and $q(1, \cdot) = c$.

3.2 The manifold of unparametrized curves – $\mathcal{I}(\mathbb{S}^1, \mathbb{R}^2)$

Using the setup from the previous section, we now define a manifold structure on $\mathcal{I}$. As it is easiest to represent element of this space with element from Imm, we are particularly interested in how to calculate the length of a path in $\mathcal{I}$ directly from a parametrized representative of this path in Imm. For a parametrized curve $c \in \mathrm{Imm}$ we write $\pi(c) := \mathrm{Im}(c) \in \mathcal{I}$ for the projection onto the quotient space, and we refer to $c \in \mathrm{Imm}$ as a *(parametrized) representative* for $\pi(c) \in \mathcal{I}$. Similarly for a path $q = q(t, \cdot) \in \mathrm{Imm}$ we construct the projection $\pi(q) = \pi(q(t, \cdot)) \in \mathcal{I}$ and refer to q as a representative for $\pi(q)$.

First of all we need to know how the tangent vectors of the quotient space $\mathcal{I}$ look like. We cannot directly use the same approach as in the previous subsection, because our definition of tangent vector would then be sensitive to reparametrizations: Consider a time-dependent reparametrization $\varphi(t, \theta)$, or, equivalently, a path $t \mapsto \varphi(t, \cdot)$ in $\mathrm{Diff}(\mathbb{S}^1)$; then the two paths $t \mapsto \pi(q(t, \cdot))$ and $t \mapsto \pi(q(t, \varphi(t, \cdot)))$ are identical in $\mathcal{I}$ but give rise to two different vector fields.

To make better sense of the tangents vectors of the quotient space, we use the following result.

Proposition 20. *For every path $t \mapsto q(t,\cdot)$ in* Imm *there exists a time-dependent reparametrization $t \mapsto \varphi(t,\cdot) \in \mathrm{Diff}(\mathbb{S}^1)$ such that the path*

$$t \mapsto \tilde{q}(t,\theta) := q(t,\varphi(t,\theta))$$

fulfills $\langle \tilde{q}_t, \tilde{q}_\theta \rangle = 0$ for all $(t,\theta) \in [0,1] \times \mathbb{S}^1$, and such that $\varphi(0,\theta) = \theta$. Furthermore, it holds that

$$\varphi_t = a \circ \varphi = -\frac{\langle q_t \circ \varphi, q_\theta \circ \varphi \rangle}{|q_\theta \circ \varphi|^2}, \quad a := -\frac{\langle q_t, q_\theta \rangle}{|q_\theta|^2}. \tag{3}$$

Proof. todo or ref. □

Remark 21

1. When we write $\tilde{q}$ in the following, we shall we refer to a path obtained from another path q by reparametrizing with φ above.
2. Note that for every vector field $h \in T_c(\mathrm{Imm})$, determined from the path q, we can make a pointwise decomposition of h onto $q_\theta(0,\cdot)$ and $iq_\theta(0,\cdot)$ by using the pointwise orthogonal projection. Explicitly we have that

   $$h = q_t = p_{q_\theta}(q_t) + p_{iq_\theta}(q_t),$$

   where p is taken to be the standard *pointwise* $\mathbb{R}^2$ orthogonal projection, which is given as

   $$p_v(u) = \frac{\langle v, u \rangle}{|v|^2} v, \quad u, v \in \mathbb{R}^2.$$

   More correctly we should thus write

   $$h(\theta) = q_t(0,\theta) = p_{q_\theta(0,\theta)}(q_t(0,\theta)) + p_{iq_\theta(0,\theta)}(q_t(0,\theta)).$$

   From this we see that the time derivative of the reparametrization in the previous Proposition is the coefficient function for the projection onto the parameter derivative of the original path q; this becomes relevant in a moment.

We can use this result to define tangent vectors to elements of $\mathcal{I}$ in a consistent way:

Definition 22. A *tangent vector* h to an element $\pi(c) \in \mathcal{I}$ is defined as a vector fields obtained from some path $(-\varepsilon, \varepsilon) \ni t \mapsto q(t,\cdot) \in \mathrm{Imm}$, with $q(0,\cdot) = c$, by

$$h(\theta) = \frac{\partial}{\partial t}\bigg|_{t=0} \tilde{q}(t,\theta) = \frac{\partial}{\partial t}\bigg|_{t=0} q(t,\varphi(t,\theta)), \tag{4}$$

with $\tilde{q}$ and φ given in accordance with remark 21.

[NB: Does this actually solve the problem about define the ubiquity in defining tangent vectors? Not clear that applying φ to a reparametrization of q will yield the same result? However, it shows that we can always think of are path as moving orthonormally; but maybe we should compine this definition with the proposition below?]

First we note that this gives us the following visualization of the tangents spaces of B.

Proposition 23. *The tangent space to an element $\pi(c) \in \mathcal{I}$ consists of orthonormal vector fields on the circle, i.e.,*

$$T_{\pi(c)}(\mathcal{I}) = \left\{ bic_\theta \mid b \in C^\infty(\mathbb{S}^1, \mathbb{R}) \right\}.$$

Proof. This follows from Definition 22 and the property of the reparametrization φ. □

As the length of a path in $\mathcal{I}$ is our primary concern, we skip straight to this without actually defining the inner product on the tangent spaces. The central idea is to mimic Definition 19 on the reparametrized path $\tilde{q}$; and though we don't bother to go trough a inner product on the tangent spaces, we note that by this construction the time derivative of the path $\tilde{q}$ is a valid tangent vector in $\mathcal{I}$ at every point $\pi(q(t, \cdot))$

Definition 24. The *length* in $\mathcal{I}$ of a path $\mathrm{q} = \pi(q)$ is

$$\mathcal{L}(\mathrm{q}) = \mathcal{L}(\pi(q)) := L(\tilde{q}) = \int_0^1 \left(\int_{\mathbb{S}^1} \|\tilde{q}_t\|^2 |\tilde{q}_\theta| \, \mathrm{d}\theta \right)^{\frac{1}{2}} \mathrm{d}t,$$

with $\tilde{q}(t, \theta) = q(t, \varphi(\theta))$ as in remark 21. The *geodesic distance* between to shapes $\mathrm{b}, \mathrm{c} \in \mathcal{I}$ represented by $c, b \in \mathrm{Imm}$, is

$$\mathcal{D}(\mathrm{b}, \mathrm{c}) = \mathcal{D}(\pi(b), \pi(c)) := \inf_{q \in \mathrm{Q}} \mathcal{L}(q),$$

where Q denotes all paths q in $\mathcal{I}$, such that $\mathrm{q}(0) = \pi(b)$ and $\mathrm{q}(1) = \pi(c)$.

Proposition 25. *$\mathcal{L}$ is well-defined, and for any representative $t \mapsto q(t, \cdot) \in \mathrm{Imm}$ of the path $t \mapsto \tilde{q}(t) \in B$ the length can be calculated as*

$$\mathcal{L}(\tilde{q}) = \int_0^1 \left(\int_{\mathbb{S}^1} \frac{\langle q_t, i q_\theta \rangle^2}{|q_\theta|} \, \mathrm{d}\theta \right)^{\frac{1}{2}} \mathrm{d}t. \tag{5}$$

Proof. For ease of notation, write $q \circ \varphi$ to mean $q(t, \varphi(t, \theta))$ and so on during this proof. First, we shows that (5) implies that $\mathcal{L}$ is well-defined; so assume (5) holds and let $q(t, \cdot)$ and $p(t, \cdot)$ be two different representatives for $\tilde{q}(t)$. This means that we must have a reparametrization $\psi(t, \theta)$ such that

$$p(t, \psi(t, \theta)) = q(t, \theta).$$

Then

$$p_t = q_t \circ \psi + \psi_t (q_t \circ \psi), \quad p_\theta = \psi_\theta (q_\theta \circ \psi),$$

so

$$\begin{aligned}\langle p_t, ip_\theta\rangle &= \langle q_t \circ \psi + \psi_t(q_\theta \circ \psi), \psi_\theta(iq_\theta \circ \psi)\rangle \\ &= \langle q_t \circ \psi, \psi_\theta(iq_\theta \circ \psi)\rangle \\ &= (\langle q_t, iq_\theta\rangle \circ \psi)\psi_\theta,\end{aligned}$$

and thus

$$\int_{\mathbb{S}^1} \frac{\langle p_t, ip_\theta\rangle^2}{|p_\theta|}\,\mathrm{d}\theta = \int_{\mathbb{S}^1} \left(\frac{\langle q_t, iq_\theta\rangle^2}{|q_\theta|}\right) \circ \psi |\psi_\theta|\,\mathrm{d}\theta = \int_{\mathbb{S}^1} \frac{\langle q_t, iq_\theta\rangle^2}{|q_\theta|}\,\mathrm{d}\theta,$$

which shows that the length does not depend on the parametrization of the path.

Next, by construction, the tangent vectors along the path $\tilde{q}$ in B is given as

$$\frac{\partial}{\partial t}(q \circ \varphi) = q_t \circ \varphi + \varphi_t(q_\theta \circ \varphi)$$

Now, as in remark 21, decompose $q_t \circ \varphi$ by projecting pointwise onto $q_t \circ \varphi$ and $iq_t \circ \varphi$. Then we get

$$q_t \circ \varphi = \left(\frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta\right) \circ \varphi + \left(\frac{\langle q_t, q_\theta\rangle}{|q_\theta|^2} q_\theta\right) \circ \varphi,$$

and by Proposition 20 we see that the last term cancels with $\varphi_t(q_\theta \circ \varphi)$, so

$$\frac{\partial}{\partial t}(q \circ \varphi) = \left(\frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta\right) \circ \varphi.$$

For any fixed $t \in [0,1]$, the reparametrization φ is just an ordinary reparametrization of the curve $\theta \mapsto q(t,\theta)$, so by invariance of the metric we have that

$$\begin{aligned}&G^2_{q(t,\varphi(t,\cdot))}\left(\frac{\partial}{\partial t}(q \circ \varphi), \frac{\partial}{\partial t}(q \circ \varphi)\right) \\ &\quad= G^2_{q(t,\varphi(t,\cdot))}\left(\left(\frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta\right) \circ \varphi, \left(\frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta\right) \circ \varphi\right) \\ &\quad= G^2_{q(t,\cdot)}\left(\frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta, \frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta\right) \\ &\quad= \int_{\mathbb{S}^1} \left\|\frac{\langle q_t, iq_\theta\rangle}{|q_\theta|^2} iq_\theta\right\|^2 |q_\theta|\,\mathrm{d}\theta \\ &\quad= \int_{\mathbb{S}^1} \frac{\langle q_t, iq_\theta\rangle^2}{|q_\theta|}\,\mathrm{d}\theta,\end{aligned}$$

from which (5) follows immediately by definition. □

Some concluding text...

3.3 Comparing this to the formal construction as a Frechet manifold

... a lot of references ...

[1]

3.4 The L2 metric vanishes

In the previous section we went through some work to construct a measure of length for a path q in the space $\mathcal{I}$. As the tangent spaces were seen to consist of some functions on the circle, it was natural to consider using a version of the L^2-metric (a version invariant to reparametrizations). However, as we illustrate in this section, this does not give rise to a usable notation of length, because the geodesic distance following from this metric becomes 0 for every two curves in the space.

It can seem weird to go through so much trouble to define a useless distance. But firstly, the construction illustrates some of the difficulties in defining a reasonable notion on length on $\mathcal{I}$; secondly, and more importantly, the fact that the distance vanishes is a quite surprising result, which depends crucially on the formula for the length of a path given in Proposition 25. This gives us some justification for the somewhat unfruitful work of defining the L^2-metric.

The "proof" below is mostly heuristic, and we only consider the very simply case of a path transforming the circle into a larger circle, but this captures the main idea.

Theorem 26. *For all* $\mathrm{a}, \mathrm{b} \in \mathcal{I}$

$$\mathcal{D}(\mathrm{a}, \mathrm{b}) = 0.$$

Proof of Theorem 26. By the definition of the length $\mathcal{D}$ we need to show that for any two paths $\mathrm{a}, \mathrm{b} \in \mathcal{I}$ and any $\varepsilon > 0$ there exists a path q in $\mathcal{I}$ such that

$$\mathcal{L}(\mathrm{q}) < \varepsilon.$$

The trick is to construct a path in Imm that moves in zigzag and then use Proposition 25 to calculate the length of this path in $\mathcal{I}$. It turns out that when we increase the number of teeth of the zigzag path, the length of the path decreases in $\mathcal{I}$. The idea of such a zigzag path is illustrated in figure 3.

To see why this phenomenon happens, first rewrite the (5) from Proposition 25

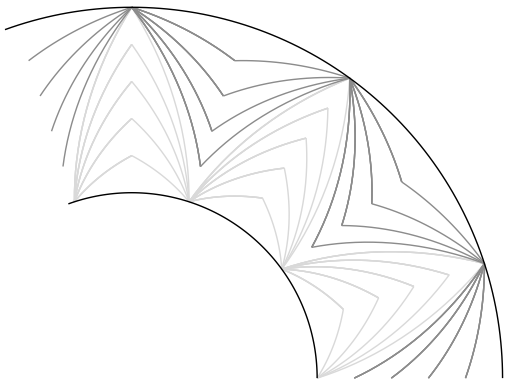

Figure 3: Illustration of a zigzag path moving the unit circle to the circle with radius 2. The path first moves the points along the path sketched in light-grey and then moves them along the dark-grey path.

as

$$\begin{aligned}\mathcal{L}(q) &= \int_0^1 \left(\int_{\mathbb{S}^1} \frac{\langle q_t, iq_\theta \rangle^2}{|q_\theta|} \,\mathrm{d}\theta \right)^{1/2} \mathrm{d}t \\ &= \int_0^1 \left(\int_{\mathbb{S}^1} \left(\frac{\langle q_t, iq_\theta \rangle}{|q_t||q_\theta|} \right)^2 |q_t|^2 |q_\theta| \,\mathrm{d}\theta \right)^{1/2} \mathrm{d}t \\ &= \int_0^1 \left(\int_{\mathbb{S}^1} \cos(\alpha(q_t, iq_\theta))^2 |q_t|^2 |q_\theta| \,\mathrm{d}\theta \right)^{1/2} \mathrm{d}t,\end{aligned}$$

with $\alpha(x, y)$ denoting the angle between x and y. When constructing a zigzag-path the angle will for large enough n be given approximately by

$$2n \approx \tan(\alpha). \tag{6}$$

Note that this does not depend on θ nor t. [Maybe not completely obvious that this holds.] See figure 4 for a visual justification of this.

We have that

$$\cos(\arctan(2n)) = (1 + (2n)^2)^{-1/2} = O(n^{-1}),$$

so we can write

$$\begin{aligned}\mathcal{L}(\mathrm{q}) &= \int_0^1 \left(\int_{\mathbb{S}^1} O(n^{-1})^2 |q_t|^2 |q_\theta| \,\mathrm{d}\theta \right)^{1/2} \mathrm{d}t \\ &= O(n^{-1}) \int_0^1 \left(\int_{\mathbb{S}^1} |q_t|^2 |q_\theta| \,\mathrm{d}\theta \right)^{1/2} \mathrm{d}t.\end{aligned} \tag{7}$$

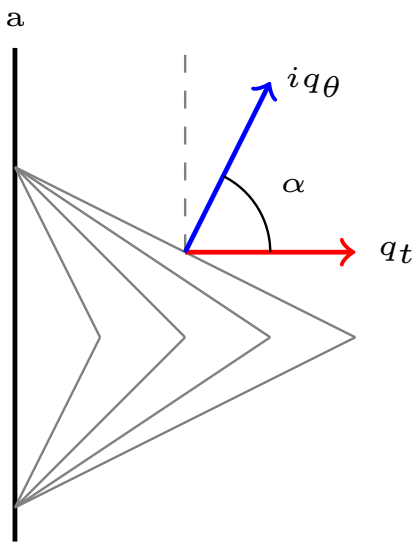

Figure 4: Relative to q_t, iq_θ is by construction a line with slope $2n$, and thus the angle between these two vectors are found by (6). Note that the starting curve, a, is here represented as a straight line to simplify the calculations; this should be a reasonable approximation when n is large.

To show that our zigzag path has arbitrary small length we thus just need to show that the remaining integral does not grow faster than n. Figure 5 illustrates how the angle α increases towards $\pi/2$, making $\cos(\alpha) \to 0$; at the same time, it is seen how q_θ grows – not fast enough, however, to kill the $\cos(\alpha)$ term, it turns out.

As an example, we take the simply case where we expand the circle $e^{i\pi\theta}$ to $2e^{i\pi\theta}$. The zigzag path in then concretely given as

$$\varphi(t,\theta) = e^{i\pi\theta} \sum_{k=0}^{n-1} h^{n,k}(t,\theta) + g^{n,k}(t,\theta)$$

where

$$h^{n,k}(t,\theta) := 1_{[\frac{k}{n}, \frac{k}{n}+\frac{1}{2n})}(\theta)\,(1 + 2t(n\theta - k))\,,$$
$$g^{n,k}(t,\theta) := 1_{[\frac{k}{n}+\frac{1}{2n}, \frac{k+1}{n})}(\theta)\,(1 + 2t(1 - n\theta - k))\,.$$

We have that

$$|\varphi_t| = \sum_{k=0}^{n-1} |h_t^{n,k}| + \sum_{k=0}^{n-1} |g_t^{n,k}|,$$
$$|\varphi_\theta| = \sum_{k=0}^{n-1} |h_\theta^{n,k} + h^{n,k}| + \sum_{k=0}^{n-1} |g_\theta^{n,k} + g^{n,k}|,$$

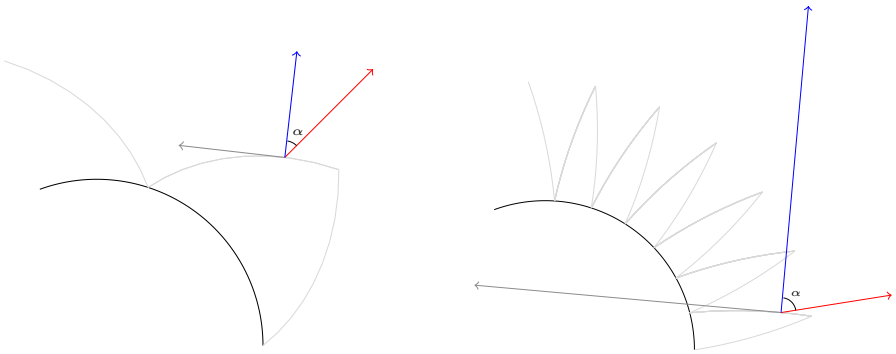

Figure 5: Illustration of how the angle α increases with the number of teeth of the zigzag path. The red vector gives the time derivative, the grey the path derivative, and the blue the normal of the path derivative. The zigzag path is shown in light-grey in the background. (Note that the length of the vectors are scaled to fit the drawing and does not necessarily reflect an exact relationship between the lengths of the two.)

so by symmetry

$$\begin{aligned}\int_0^1 |\varphi_t|^2|\varphi_\theta|\,\mathrm{d}\theta &= 2n\int_0^{\frac{1}{2n}} |h_t^{n,0}|^2|h_\theta^{n,0}+h^{n,0}|\,\mathrm{d}\theta \\ &= 2n\int_0^{\frac{1}{2n}} (2n\theta)^2(2tn+1+t2n\theta)\,\mathrm{d}\theta \\ &= \int_0^1 u^2(2tn+1+tu)\,\mathrm{d}\theta \\ &= O(n),\end{aligned}$$

for $t\in[0,1]$. Plugging this into (7) gives the result. □

4 Statistical concepts on shape space

If we wish to define statistical concepts like mean and variance on our shape space - the manifold of unparametrized curves - we are faced with some challenges. To begin with our manifold fold is infinite-dimensional, and all theory developed so far in this exposition has been on finite-dimensional manifolds. This is an important part of the generalizations of mean and variance, since we can quite easily define a measure on M. If a point $x\in M$ can be written in finitely many local coordinates, a measure can be defined via the actions of the metric on the induced (finite) basis of T_xM. It is not at all obvious how this method of constructing a measure on M generalizes to the infinite-dimensional

case.

Secondly, all results regarding existence and uniqueness of mean points have relied on either a global assumption or local assumptions on the Riemannian curvature of the manifold. The behaviour of the curvature of our shape space is not at all well understood, especially since the choice of metric is non-trivial to begin with. One might think that the curvature of $B_e(S^1, \mathbb{R}^2)$ equipped with the L^2-metric is the least difficult to examine, but since the distance function induced by this metric is vanishing, the resulting variance, $\sigma_X^2(y)$ would be 0 for all random variables X and shapes y (if it is even possible to generalize the variance-formulas in Section 2 to the infinite dimensional case.). Thus every point would be a mean point of X.

References

[1] John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.