

Statistical analysis of shapes

Mads and Anders

October 25, 2018

1 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 extent “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 definitions geometrically and then deduce some properties from these definitions – 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 $\text{Im}(c) \subset \mathbb{R}^2$ (as illustrated in Figure 1). We impose some smoothness structure on the curves and define the spaces we want to consider formally as

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

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 objects $q, p \in \text{Imm}$ in \mathcal{I} if $q = p \circ \varphi$, where $\varphi \in \text{Diff}(\mathbb{S}^1)$ is a diffeomorphism on the unit circle.

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.

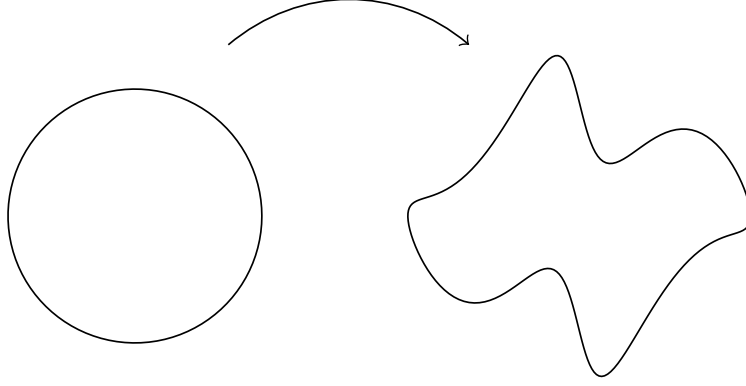


Figure 1: A simple shape. We can think of a shape as the whole mapping, or simply as the subset to the right.

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

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

$$[-\varepsilon, \varepsilon] \ni t \mapsto q(t, \cdot) := (\theta \mapsto q(t, \theta)) \in \text{Imm}, \quad q(0, \theta) = c(\theta), \quad (1)$$

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

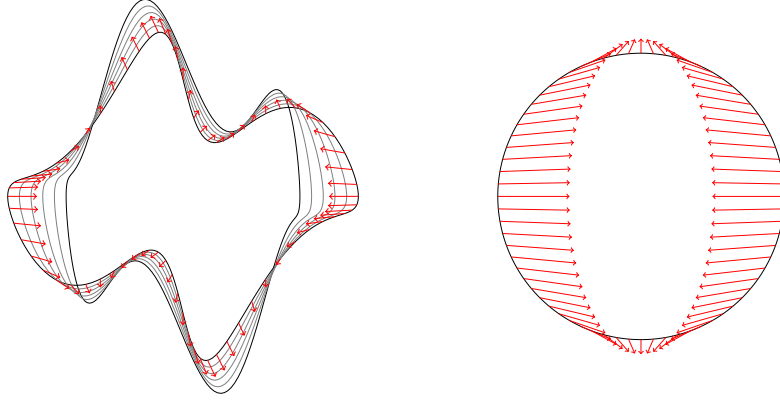


Figure 2: Illustration of tangent vectors in Imm . The left illustration shows a path in the space of curves and how to obtain a tangent vector from this. The right illustration shows how to think of this tangent vector as a vector field on the circle.

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

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

where q is some path passing through c at 0 (as defined in (1)).

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 2. The G metric or G_c metric at the point $c \in \text{Imm}$ is defined as

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

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

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 3. 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

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

The *geodesic distance* [geodesic?] between to curves $b, c \in \text{Imm}$ (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 $q(1, \cdot) = c$.

1.2 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 $c \in \text{Imm}$ we write $\pi(c) := \text{Im}(c) \in \mathcal{I}$ for the projection onto the quotient space, and we refer to $c \in \text{Imm}$ as a (*parametrized*) *representative* for $\pi(c) \in \mathcal{I}$. Similarly for a path $q = q(t, \cdot) \in \text{Imm}$ 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 $\text{Diff}(\mathbb{S}^1)$; 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 4. *For every path $t \mapsto q(t, \cdot)$ in Imm there exists a time-dependent reparametrization $t \mapsto \varphi(t, \cdot) \in \text{Diff}(\mathbb{S}^1)$ 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

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

Proof. todo or ref. □

- Remark 5** 1. When we write \tilde{q} in the following, we shall refer to a path obtained from another path q by reparametrizing with φ above.
2. Note that for every vector field $h \in T_c(\text{Imm})$, determined from the path q , we can make a pointwise decomposition of h onto $q_\theta(0, \cdot)$ and $i q_\theta(0, \cdot)$ by using the pointwise orthogonal projection. Explicitly we have that

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

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

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

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

$$h(\theta) = \frac{\partial}{\partial t} \Big|_{t=0} \tilde{q}(t, \theta) = \frac{\partial}{\partial t} \Big|_{t=0} q(t, \varphi(t, \theta)), \quad (3)$$

with \tilde{q} and φ given in accordance with remark 5.

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

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

Proof. This follows from Definition 6 and the property of the reparametrization φ . \square

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 3 on the reparametrized path \tilde{q} ; and though we don't bother to go through an 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 8. The *length* in \mathcal{I} of a path $q = \pi(q)$ is

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

with $\tilde{q}(t, \theta) = q(t, \varphi(\theta))$ as in remark 5. The *geodesic distance* between two shapes $b, c \in \mathcal{I}$ represented by $c, b \in \text{Imm}$, is

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

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

Proposition 9. \mathcal{L} is well-defined, and for any representative $t \mapsto q(t, \cdot) \in \text{Imm}$ of the path $t \mapsto \tilde{q}(t) \in B$ the length can be calculated as

$$\mathcal{L}(\tilde{q}) = \int_0^1 \left(\int_{\mathbb{S}^1} \frac{\langle q_t, iq_\theta \rangle^2}{|q_\theta|} d\theta \right)^{\frac{1}{2}} dt. \quad (4)$$

Proof. For ease of notation, write $q \circ \varphi$ to mean $q(t, \varphi(t, \theta))$ and so on during this proof. First, we show that (4) implies that \mathcal{L} is well-defined; so assume (4) 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

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

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 5, decompose $q_t \circ \varphi$ by projecting pointwise onto $q_t \circ \varphi$ and $i q_t \circ \varphi$. Then we get

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

and by Proposition 4 we see that the last term cancels with $\varphi_t(q_\theta \circ \varphi)$, so

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

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

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

from which (4) follows immediately by definition. \square

Some concluding text...

1.3 Comparing this to the formal construction as a Frechet manifold

... a lot of references ...

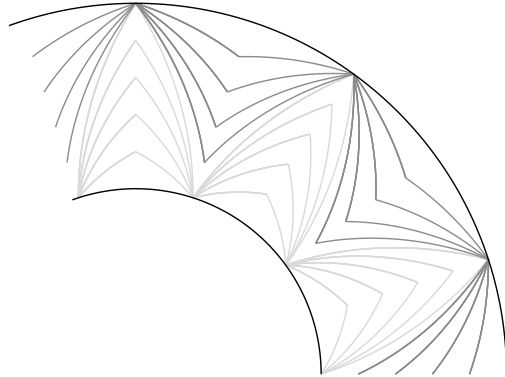


Figure 3: Illustration of a zigzag path moving the unit circle to the circle with radius 2. The path first moves the points along the path sketched in light-gray and then moves them along the dark-gray path.

1.4 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 9. This gives us some justification for the somewhat unfruitfull 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 10. *For all $a, b \in \mathcal{I}$*

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

Proof of Theorem 10. By the definition of the length \mathcal{D} we need to show that

for any two paths $a, b \in \mathcal{I}$ and any $\varepsilon > 0$ there exists a path q in \mathcal{I} such that

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

The trick is to construct a path in Imm that moves in zigzag and then use Proposition 9 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 (4) from Proposition 9 as

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

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 by

$$2n = \tan(\alpha), \quad (5)$$

i.e., independently of θ and 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

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

To show that our zigzag path has arbitrary small length we just need to show that the remaining integral does not grow faster than n^2 . Figure 5 illustrates how the angle α increases towards $\pi/2$, making $\cos(\alpha) \rightarrow 0$; 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 is then concretely given as

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

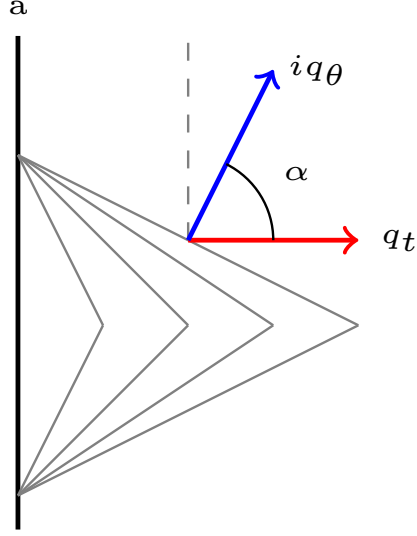


Figure 4: Relative to q_t , iq_θ is by construction a line with slope $2n$, and thus the angle between these two vectors are found by (5). Note that the starting curve, a , is here represented as a straight line to simplify the calculations; this should be a reasonable approximation when n is large.

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

$$|\varphi_t| = \sum_{k=0}^{n-1} |h_t^{n,k}| + \sum_{k=0}^{n-1} |g_t^{n,k}|,$$

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

so by symmetry

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

for $t \in [0, 1]$ which gives the result.

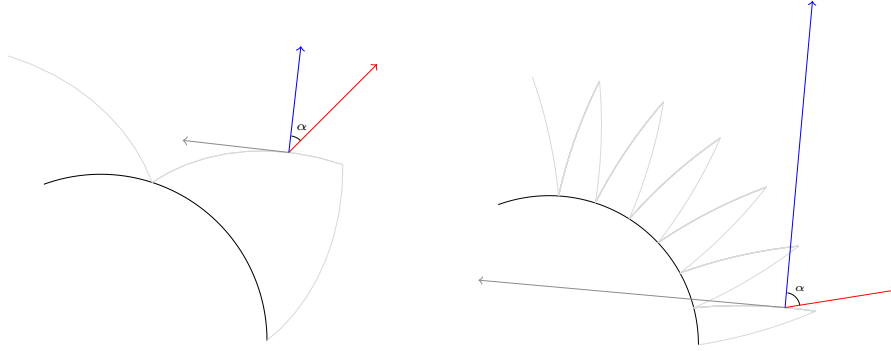


Figure 5: Illustration of how the angle α increases with the number of teeth of the zigzag path. The red vector gives the time derivate, the gray the path derivative, and the blue the normal of the path derivative. The zigzag path is shown in light-gray in the background. (Note that the length of the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

□