Identifying latent distances with Finslerian geometry

Anonymous authors
Paper under double-blind review

Abstract

Riemannian geometry provides powerful tools to explore the latent space of generative models while preserving the inherent structure of the data. Distance and volume measures can be computed from a Riemannian metric defined by pulling back the Euclidean metric from the data to the latent manifold. With this in mind, most generative models are stochastic, and so is the pullback metric. Yet, manipulating stochastic objects is at best impractical, and at worst unachievable. To perform operations such as interpolations, or measuring the distance between data points, we need a deterministic approximation of the pullback metric. In this work, we define a new metric as the expected length derived from the stochastic pullback metric. We show this metric defines a Finsler metric. We compare it with the expected pullback metric. We show that in high dimensions, the metrics converge to each other at a rate of $\mathcal{O}\left(\frac{1}{D}\right)$.

1 Introduction

Generative models provide a convenient way to learn low-dimensional latent variables z corresponding to data observations x through a smooth function $f : \mathcal{Z} \subset \mathbb{R}^q \to \mathcal{X} \subset \mathbb{R}^D$, such that $x = f(z)$. In practice, this function can be the decoding part of a variational autoencoder (VAE) (Kingma & Welling, 2013; Rezende et al., 2014), a Gaussian Process Latent Variable Model (GPLVM) (Lawrence, 2003) or similar.

Through this learnt manifold, one can generate new data or compare observations by interpolating or computing distances. However, **we argue that doing so using the Euclidean metric in the latent space is wrong and misleading** (Hauberg, 2018). If our observations are lying near a manifold (Fefferman et al., 2016), we want to equip our latent space with a metric that preserves distance measures on it. Let us compute, for example, the infinitesimal Euclidean norm in our data space. Using the Taylor expansion, we have: $\left\|f(z+\Delta z) - f(z)\right\|_2^2 \approx \left\|f(z) + J_f(z)\Delta z - f(z)\right\|_2^2 = \Delta z^\top J_f(z)^\top J_f(z)\Delta z$. As a first approximation, the Euclidean distance can be preserved on $\mathcal{X}$ by defining, locally, a new norm on $\mathcal{Z}$ that includes the term $G(z) = J_f^\top(z)J_f(z)$, with $z \in \mathcal{Z}$. This term is called a Riemannian metric tensor, and it condenses all the geometrical information needed to describe the data space. In mathematical jargon, we say that the manifold $\mathcal{Z}$ is equipped with the Riemannian metric $g : (u, v) \to u^\top G v$ obtained by pulling back the Euclidean metric through the immersion f.

Riemannian geometry enables the exploration of the latent space in precise geometric terms, and quantities of interest such as the length, the energy or the volume can be directly derived from the pullback metric. These geometric quantities are, by construction, known to be invariant to reparametrizations of the latent space $\mathcal{Z}$, and are thus statistically identifiable. Yet, we encounter another problem: this geometric framework only deals with deterministic objects, and, often, the decoding part of generative models is stochastic. Then, **we need a deterministic approximation of the Riemannian metric pulled back through the stochastic immersion f.**

In this paper, we believe that the length derived from the pullback metric is a sensible distance measure, and we approximate it by taking its expectation. Then, we **define a new metric as the expected length derived from the stochastic pullback metric.** This extends the work of Eklund & Hauberg (2019).

The findings are the following:

1. We prove that the expected length defines a norm, and this norm is a Finsler metric. Finsler geometry is a generalisation of Riemannian geometry.

2. We prove that, in high dimensions, our Finsler metric and a previously studied Riemannian metric, introduced by Tosi et al. (2014), converge to each other at a rate of $\mathcal{O}\left(\frac{1}{D}\right)$.

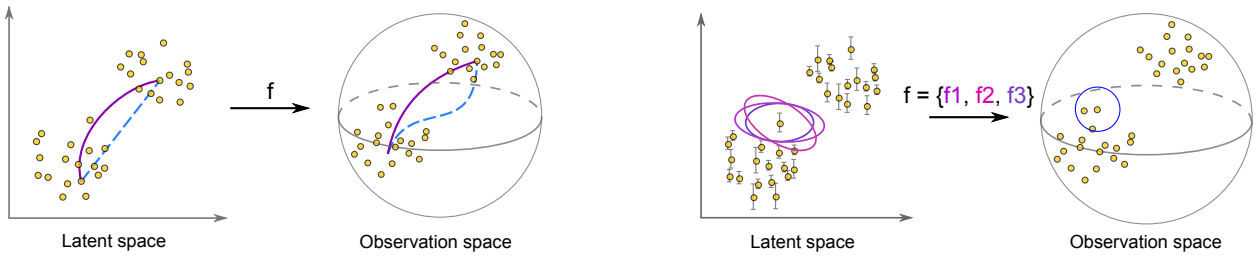

Figure 1: **Left figure**: The Euclidean distance measure, in blue, in the latent space does not take into consideration the geometry of the observational manifold, and therefore it is not identifiable and leads to misinterpretations. Instead, the length derived from the pullback metric will follow the curvature of the manifold. **Right figure**: Generative models often map the latent space to the data space using a stochastic process $f = \{f_1, f_2, \dots\}$. A stochastic Riemannian metric, whose realisations are represented by ellipses, are obtained when we pull back the Euclidean metric, represented by a unit circle, through the stochastic process f.

1.1 Outline of the paper

In Section 2, we explain the notion of random manifolds and introduce the expected Riemannian metric tensor. We then define a norm as the expectation of the stochastic length, and show that this norm is a Finsler metric. After outlining the general differences between Finsler and Riemannian geometry, we draw absolute bounds on the Finsler metric, and then relative bounds. We finally show that the metrics converge when increasing the number of dimensions of the observational space. We perform some experiments in Section 4, where we can see that the Riemannian and Finsler metrics are similar. We conclude by discussing our findings in Section 5.

1.2 Related works

Finsler geometry in Machine Learning

Our work crucially relies on Finslerian geometry, which has been well-studied mathematically, but has only seen very limited use in machine learning and statistics. We point to two notable exceptions, which are quite distinct from our work. Lopez et al. (2021) use symmetric spaces to represent graphs and endow these with a Finsler metric to capture dissimilarity structure in the observational data. Ratliff et al. (2021) discuss the role of differential geometry in motion planning for robotics. Along the way, they touch upon Finslerian geometry, but mostly as a neat tool to allow for generalizations. To the best of our knowledge, no prior work has investigated the links between stochastic and Finslerian geometry.

Strategies to deal with stochastic Riemannian geometry

In addition to the work of Tosi et al. (2014) in the case of GP-LVMs and Arvanitidis et al. (2018) for VAEs, detailed in Section 2.2, a solution to circumvent the randomness of the metric tensor is to consider that the data follows a specific probability distribution. Instead of looking at the shortest path on the data manifold, Arvanitidis et al. (2021) borrow tools from information geometry and consider the straightest paths on the manifold whose elements are probability distributions.

2 Expectation on random manifolds

The metric pulled back by a stochastic mapping is, de facto, stochastic and endows a random manifold. Unfortunately, we are not yet equipped to derive geometric objects on a random manifold. Instead, we dodge this problem by seeking a deterministic approximation of this stochastic metric.

As mentioned above, a common solution is to approximate such a metric by its expectation. In section 2.2, we study the expected Riemannian metric and summarise the main findings of Eklund & Hauberg (2019).

The other solution suggested by this paper is to approximate the expectation of the lengths instead of the random metric itself. In section 2.4, we show that this new metric is not Riemannian but Finslerian (Proposition 2.2), and it has a closed-form expression when the stochastic immersion is a Gaussian process (Proposition 2.3).

2.1 Riemannian geometry and random metric

Consider a smooth map $f : \mathcal{M} \to \mathcal{N}$, such that its derivatives exist and are injective everywhere on the manifold. This mapping defines an immersion, and so, the pullback of a Euclidean metric on $\mathcal{N}$ is a Riemannian metric on $\mathcal{M}$ (Lee, 2013, Proposition 13.9). A manifold equipped with a Riemannian metric is called a Riemannian manifold.

Definition 2.1. The pullback of the Euclidean metric through the immersion $f : \mathcal{M} \to \mathcal{N}$ is a **Riemannian metric**. It is defined as the inner product $g_x : (\mathcal{T}_x\mathcal{M}, \mathcal{T}_x\mathcal{M}) \to \mathbb{R}_+ : (u, v) \to u^\top G(x)v$, at a specific point x in the manifold $\mathcal{M}$. u and v are vectors lying in the tangent plane $\mathcal{T}_x\mathcal{M}$ of the manifold. $G = J^\top J$, with J the Jacobian of f.

Since a Riemannian metric is an inner product, it induces a norm. We can use this norm to define the curve length and curve energy on a manifold:

Definition 2.2. We consider a curve γ and its derivative $\dot{\gamma}$ on a Riemannian manifold $(\mathcal{M}, g)$. Then, we define the **curve length and curve energy**:

$$L_G(\gamma) = \int_0^1 \left\|\dot{\gamma}(t)\right\|_G dt = \int_0^1 \sqrt{\dot{\gamma}(t)^\top G \dot{\gamma}(t)}\, dt,$$
$$E_G(\gamma) = \int_0^1 \left\|\dot{\gamma}(t)\right\|_G^2 dt = \int_0^1 \dot{\gamma}(t)^\top G \dot{\gamma}(t)\, dt.$$

Locally length-minimising curves between two connecting points are called **Geodesics**.

Similar to the change of variable theorem, we can define a volume measure to estimate densities on the manifold:

Definition 2.3. We consider a Riemannian manifold $(\mathcal{M}, g)$, and G is the corresponding metric tensor. Let $f : \mathcal{M} \to \mathcal{N}$ be an immersion, and h be an integrable function over $\mathcal{N}$. Then, we have: $\int_{\mathcal{N}} h(y)\, dy = \int_{\mathcal{M}} h \circ f(x) V_G(x)\, dx$, with the V_G volume measure defined as:

$$V_G(x) = \sqrt{\det G(x)}$$

All the functionals L_G, E_G and V_G can be directly derived from the Riemannian metric tensor $G = J^\top J$. When f is a stochastic immersion, the metric tensor becomes a random matrix. In this paper, we call a manifold equipped with the stochastic pullback metric a **random manifold**. As a consequence of the stochastic aspect of the metric, all the functionals are stochastic themselves, and they are no longer trivial to manipulate.

2.2 Expected metric tensor on random manifolds

A deterministic approximation is needed, and the simplest way to approximate a random metric tensor is to take its expectation with respect to the collection of random metrics induced by the stochastic process. This has been introduced before by Tosi et al. (2014) for Gaussian Process Latent Variable Models (GP-LVMs) and Arvanitidis et al. (2018) for Variational Autoencoders (VAEs).

Definition 2.4. Let G be a stochastic Riemannian metric tensor on the manifold $\mathcal{M}$. We refer to $\mathbb{E}[G]$ as the **expected metric tensor**, and $g_x : (\mathcal{T}_x\mathcal{M}, \mathcal{T}_x\mathcal{M}) \to \mathbb{R}_+ : (u, v) \to u^\top \mathbb{E}[G(x)]v$ as the **expected Riemannian metric**, defined at a specific point $x \in \mathcal{M}$. This metric defines a Riemannian metric on $\mathcal{M}$.

Like any Riemannian metric, we can define the following functionals:

$$
\begin{aligned}
L(\gamma) &= \int_0^1 \sqrt{\dot\gamma(t)^\top \mathbb{E}[G]\dot\gamma(t)}\, dt, \\
E(\gamma) &= \int_0^1 \dot\gamma(t)^\top \mathbb{E}[G]\dot\gamma(t)\, dt = \mathbb{E}[E_G(\gamma)], \\
V(x) &= \sqrt{\det \mathbb{E}[G(x)]}.
\end{aligned}
$$

2.3 Expected paths on random manifolds

Approximating the stochastic metric by its expectation seems a natural but also ad-hoc solution. If we want to explore a manifold, we might prefer to use a representative quantity, such as the lengths between data points. The expectation of the lengths can give us an idea about how, on average, two points are connected on a random manifold:

Definition 2.5. We define the **expected curve length**, and its corresponding **curve energy** on our random manifold as:

$$
\begin{aligned}
\mathcal{L}(\gamma) &= \int_0^1 \mathbb{E}\left[\sqrt{\dot\gamma(t)^\top G_t \dot\gamma(t)}\right] dt = \mathbb{E}[L_G(\gamma)], \\
\mathcal{E}(\gamma) &= \int_0^1 \mathbb{E}\left[\sqrt{\dot\gamma(t)^\top G_t \dot\gamma(t)}\right]^2 dt.
\end{aligned}
$$

As noted by Eklund & Hauberg (2019), the length (L) derived from the expected Riemannian metric is not equal to the expected curve length ($\mathcal{L}$), and their respective energy curves differ by a variance term:

$$
\begin{aligned}
E(\gamma) - \mathcal{E}(\gamma) &= \int_0^1 \dot\gamma(t)^\top \mathbb{E}[G]\dot\gamma(t) - \mathbb{E}\left[\sqrt{\dot\gamma(t)^\top G_t \dot\gamma(t)}\right]^2 dt \\
&= \int_0^1 \mathbb{E}[\|\dot\gamma(t)\|_G^2] - \mathbb{E}[\|\dot\gamma(t)\|_G]^2\, dt = \int_0^1 \mathrm{Var}[\|\dot\gamma(t)\|_G]\, dt
\end{aligned}
$$

This term can be regarded as a regularisation term for the Riemannian energy curve: the curve energy E might be penalised when the curve goes through regions with high-variance. In practice, for a Gaussian process with a stationary kernel, this variance term is upper bounded by the posterior variance that is relatively low next to the training points and is high outside of the support of the data. Later, we will also see that the functionals agree in high dimensions, leading to the same geodesics (Section 3).

Eklund & Hauberg (2019) also observed that these quantities are bounded by the number of dimensions:

Proposition 2.1. (Eklund & Hauberg, 2019) Let $f : \mathbb{R}^q \to \mathbb{R}^D$ be a stochastic process such that the sequence: $\{f'_1, f'_2, \dots, f'_D\}$ has uniformly bounded moments. There is then a constant C such that:

$$0 \leq \frac{L - \mathcal{L}}{L} \leq \frac{C}{8D}$$

2.4 Finsler geometry: norm defined as the expected length

Instead of considering the expected Riemannian metric, we want to study a norm such that its corresponding curve length is defined as the expectation of the random lengths. We define our new norm as: $v \to \mathbb{E}\left[\sqrt{v^\top G(x) v}\right]$.

Proposition 2.2. Let G be a stochastic Riemannian metric. Then, the function:

$$F_x : \mathcal{T}_x\mathcal{M} \to \mathbb{R} : v \to \mathbb{E}\left[\sqrt{v^\top G(x) v}\right]$$

defines a Finsler metric, but it is not induced by a Riemannian metric.

Proof. If F was induced by a Riemannian metric, then this metric would be defined as: $f_x : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}_+ : (v_1, v_2) \to \mathbb{E}\left[\sqrt{v_1^\top G(x) v_2}\right]^2$. Since a Riemannian metric is an inner product, it should be symmetric, positive, definite and bilinear. Here, we can see that f_x is not bilinear, so f_x is not a Riemannian metric. However, we can prove that $F_x : \mathbb{R}^q \to \mathbb{R} : v \to \mathbb{E}\left[\sqrt{v^\top G(x) v}\right]$ is positive, homogeneous, smooth and strongly convex, and so F_x is a Finsler metric (Shen & Shen, 2016, Definition 2.1). For the full proof, see Section B.1. □

So far, we have assumed that f is an immersion and a stochastic process. If we consider f to be a Gaussian Process in particular, the Finsler norm can be rewritten in a closed form expression.

Proposition 2.3. Let f be a Gaussian process and J its Jacobian.

Then $G = J^\top J \sim \mathcal{W}_q(D, \Sigma, \mathbb{E}[J]^\top \mathbb{E}[J])$ follows a non-central Wishart distribution. The Finsler metric can be written as:

$$F_x : \mathcal{T}_x\mathcal{M} \to \mathbb{R} : v \to \sqrt{2}\sqrt{v^\top \Sigma v}\frac{\Gamma(\frac{D}{2} + \frac{1}{2})}{\Gamma(\frac{D}{2})} {}_1F_1\left(-\frac{1}{2}, \frac{D}{2}, -\frac{\omega}{2}\right),$$

with ${}_1F_1$ as the confluent hypergeometric function of the first kind and $\omega = (v^\top \Sigma_x v)^{-1}(v^\top \mathbb{E}[J]^\top \mathbb{E}[J] v)$.

Proof. We suppose that f is a Gaussian process, and so is its Jacobian. $G = J^\top J$ follows a non-central Wishart distribution. $v^\top G v$ is a scalar and also follows a non-central Wishart distribution (Kent & Muirhead, 1984, Definition 10.3.1). The square-root of a non-central Wishart distribution follows a non-central Nakagami distribution (Hauberg, 2018) whose expectation is known. □

3 Comparison of Riemannian and Finsler metrics

3.1 Theoretical comparison

In geometry, we need to define a metric (a norm) to compute functionals. In Riemannian geometry, the metric is conveniently obtained by constructing an inner product. Because of its bilinearity, the inner product greatly simplifies subsequent computations, but it is also restrictive. A generalisation of Riemannian geometry can be obtained by relaxing this assumption. Instead of defining a metric as an inner product, we can define

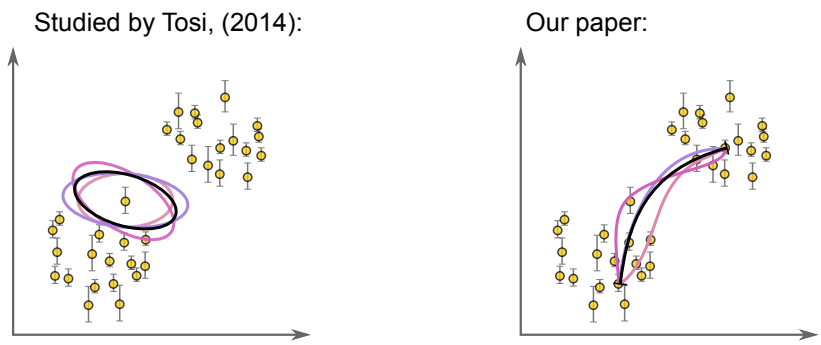

Figure 2: The plots represent a random manifold $(\mathcal{M}, \tilde{g})$, with the indicatrices that illustrate the realisations of the random Riemannian metric $\tilde{g}$ on the left, and its respective random lengths L_G on the right. **Left figure**: Tosi et al. (2014) studied the expectation of the Riemannian metric tensor g. **Right figure**: In our paper, we consider that the expectation of the lengths derived on the random manifold is a functional directly derived from the Finsler norm F.

the metric as a norm[1]. Relaxing this assumption was studied by Finsler (1918), who gave his name to this discipline.

Finsler geometry is similar to Riemannian geometry without the bilinear assumption. Most of the functionals (curve length and curve energy) are defined similarly to those obtained in Riemannian geometry. However, the volume measure is different, and there are at least two definitions of volume measure used in Finsler geometry: the Busemann-Hausdorff volume and the Holmes-Thomson volume measure (Wu, 2011). In this paper, we decided to focus on the Busemann-Hausdorff definition (Definition 3.1), which is more intuitive and easier to derive. If the Finsler metric is a Riemannian metric, the definition of volume naturally coincides with the Riemannian volume measure.

In Figure 4, a Busemann-Hausdorff and the Riemannian volume measures have been computed for the same set of data points. A Gaussian process has been trained to fit data representing a pinwheel projected onto a sphere.

Definition 3.1. For a given point x on the manifold, we define the **Finsler indicatrix** as the set of vectors in the tangent space such that the Finsler metric is equal to: $\{v \in \mathcal{T}_x M | F_x(v) = 1\})$. We call $\mathbb{B}^n(1)$ the Euclidean unit ball. In local coordinates $(e^1, \cdots, e^d)$ on a Finsler manifold $\mathcal{M}$, the **Busemann-Hausdorff volume** form is defined as $d\mathcal{V} = \mathcal{V}(x) e^1 \wedge \cdots \wedge e^d$, with:

$$\mathcal{V}(x) = \frac{\mathrm{vol}(\mathbb{B}^n(1))}{\mathrm{vol}(\{v \in \mathcal{T}_x M | F_x(v) < 1\})}.$$

In the definition above, we introduce the notion of *indicatrix*. An indicatrix is a way to represent the distortion induced by the metric on a unit circle. If our metric is euclidean, we will only have a linear transformation between the latent and the observational spaces, and the indicatrix would still be a circle. Because the Riemannian metric is quadratic, it will always generate an ellipse in the latent space. The Finsler indicatrix, however, would have a convex, even asymmetrical, shape. This difference can be observed in the indicatrix-field represented in Figure 3: The Finsler indicatrices in purple can have almost rectangular shape, while the Riemannian indicatrices, in orange, are ellipses.

There are also a few observations to note in Figure 3. First, in the area of low predictive variance (where data points lie in the latent space), the Finsler and Riemannian indicatrices are alike. This follows from the preceding comment that the metrics diverge by a variance term. If our mapping f was deterministic,

[1] A norm only needs to be definite and satisfies the triangular inequality, but is not necessary symmetric. This means that, for a vector v, we can have a non reversible Finsler metric: $F_x(v) \neq F_x(-v)$. Intuitively, this means that the path used to connect two points would be different depending on the starting point. This asymmetric property becomes valuable when studying the geometry of anisotropic media (Markvorsen, 2016), for example. In our case, our Finsler metric is reversible.

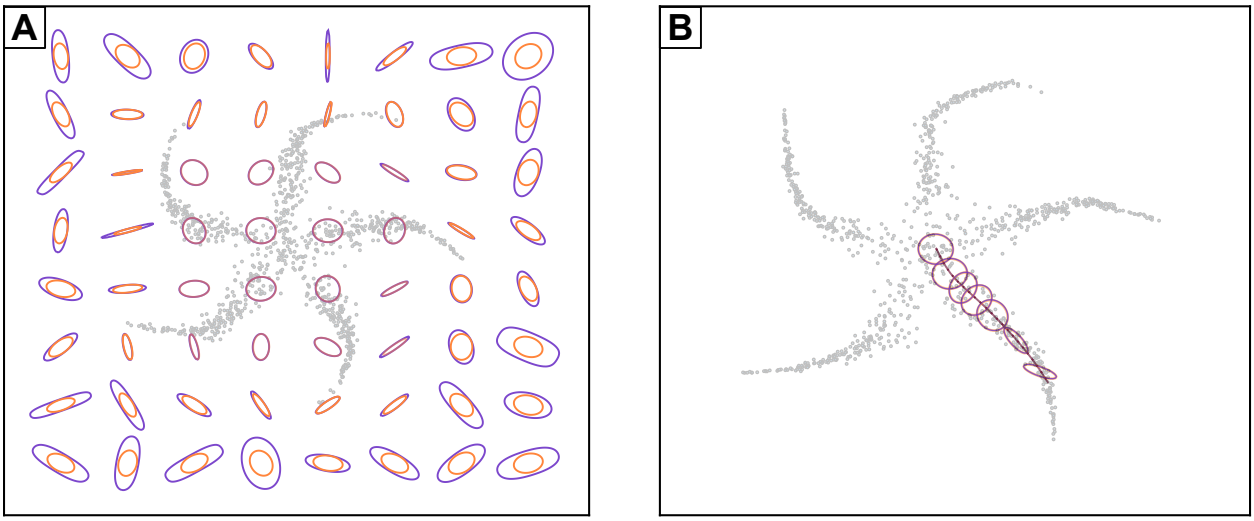

Figure 3: Indicatrice field over the latent space of the pinwheel data (in grey) representing the Riemannian (in orange) and Finslerian (in purple) metrics (See Section C). (A) The indicatrices are computed over a grid in the latent space. (B) The indicatrices are computed along a geodesic: the Riemannian and Finslerian metrics coincide.

both metrics would agree. Second, for every point, the Riemannian indicatrices are always contained by the Finslerian ones, illustrating Proposition 3.1 on our absolute bounds in the following section.

3.2 Absolute bounds on the Finsler metric

The Finsler metric is upper bounded with the norm induced by the pullback metric. It is also lower bounded:

Proposition 3.1. We define $\alpha = 2\left(\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}\right)^2$. The Finsler metric: $v \to \mathbb{E}\left[\sqrt{v^\top G_x v}\right]$ is bounded by two Riemannian metric tensors: the covariance tensor $\alpha\Sigma_x$ and the expected metric tensor $\mathbb{E}[G_x]$.

$$\forall (x, v) \in \mathcal{M} \times \mathcal{T}_x M : \ \sqrt{v^\top \alpha \Sigma_x v} \leq F_x(v) \leq \sqrt{v^\top \mathbb{E}[G_x] v}$$

Proof. The proof can be sketched as follows: the upper bound $F_x(v) \leq \sqrt{v^\top \mathbb{E}[G_x] v}$ is obtained by applying Jensen's inequality, knowing that the square root $x \to \sqrt{x}$ is a concave function. The lower bound $\sqrt{v^\top \alpha \Sigma_x v} \leq F(x, v)$ is obtained using the closed form expression of the Finsler function. □

The result is illustrated in Figure 5 (lower right). Four metric tensors (G_1, G_2, G_3, G_4), each following a non-central Wishart distribution with a specific mean and covariance matrix, have been computed. For each of them, we have drawn the indicatrices induced by the Finsler metric F, the Riemannian metric tensor $\mathbb{E}[G]$ and the covariance metric tensor $\alpha\Sigma$. As expected, we can notice that the $\alpha\Sigma$-indicatrix contains the Finsler indicatrix, itself containing $\mathbb{E}[G]$-indicatrix.

By bounding the Finsler metric, we are able to bound their respective functionals:

Corollary 3.1. The length, the energy and the Busemann-Hausdorff volume of the Finsler metric are bounded respectively by the Riemannian length, energy and volume of the covariance tensor $\alpha\Sigma$ (noted

$L_\Sigma, E_\Sigma, V_\Sigma$) and the expected metric $\mathbb{E}[G]$ (noted L_G, E_G, V_G):

$$\begin{aligned} \forall x \in \mathcal{M},\ L_\Sigma(x) \leq \mathcal{L}(x) \leq L_G(x) \\ E_\Sigma(x) \leq \mathcal{E}(x) \leq E_G(x) \\ V_\Sigma(x) \leq \mathcal{V}(x) \leq V_G(x) \end{aligned}$$

Proof. From Proposition 3.1, we need to integrate each term of the inequality to obtain the length and the energy. The volume is less trivial, since we use the Busemann-Hausdorff definition for measuring $\mathcal{V}$. We have to place ourselves in hyperspherical coordinates, and show that the Finsler indicatrix is still bounded. □

3.3 Relative bounds on the Finsler metric

Proposition 3.2. The relative difference between the Finsler metric: $F_x : v \to \mathbb{E}[\sqrt{v^\top G_x v}]$ and the Riemmanian metric $g : (v, v) \to v^\top \mathbb{E}[G_x] v$ is:

$$0 \leq \frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} \leq \frac{\mathrm{Var}[v^\top G_x v]}{2\mathbb{E}[v^\top G_x v]^2}.$$

Proof. This proposition is a direct application of the Sharpened Jensen's inequality (Liao & Berg, 2019). □

The previous proposition is valid for any stochastic immersion. We can see that the metrics become equal when the ratio of the variance over the expectation shrinks to zero. This happens in two cases: when the variance converges to zero, which is similar to having a deterministic immersion, and when the number of dimensions increases. The latter case is investigated below for a Gaussian process.

Proposition 3.3. Let f be a Gaussian process. We note $\omega = (v^\top \Sigma_x v)^{-1}(v^\top \mathbb{E}[J]^\top \mathbb{E}[J] v)$, with J_f the jacobian of f, and Σ the covariance matrix of J.
The relative ratio between the Finsler metric: $F : (x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$ and the Riemmanian metric $g : (v, v) \to v^\top \mathbb{E}[G_x] v$ is:

$$0 \leq \frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} \leq \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2}.$$

Proof. $v^\top G_x v$ follows a one-dimension non-central Wishart distribution. We use the theorem of the moments to obtain both the expectation and the variance, which leads us to the result. □

Corollary 3.2. The relative ratio between the length, the energy and the volume of the Finsler metric: $(x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$ (noted $\mathcal{L}, \mathcal{E}, \mathcal{V}$) and the Riemannian metric with the metric tensor $\mathbb{E}[G_x]$ (noted L_G, E_G, V_G) is:

$$\begin{aligned} 0 \leq & \frac{L_G(x) - \mathcal{L}(x)}{L_G(x)} \leq \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\} \\ 0 \leq & \frac{E_G(x) - \mathcal{E}(x)}{E_G(x)} \leq \max_{v \in \mathcal{T}_x M} \left\{ \frac{2}{D+\omega} + \frac{1+2\omega}{(D+\omega)^2} + \frac{2\omega}{(D+\omega)^3} + \frac{\omega^2}{(D+\omega)^4} \right\} \\ 0 \leq & \frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} \leq 1 - \left(1 - \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\} \right)^q \end{aligned}$$

Proof. We directly use Proposition 3.3. To obtain the inequalities with the lengths and the energies, we first multiply all the terms by the Riemannian metric, and we integrate every term. To obtain the inequality with the volume, similarly to Corollary 3.1, we place ourselves in hyperspherical coordinates and bound the radius of the Finsler indicatrix. □

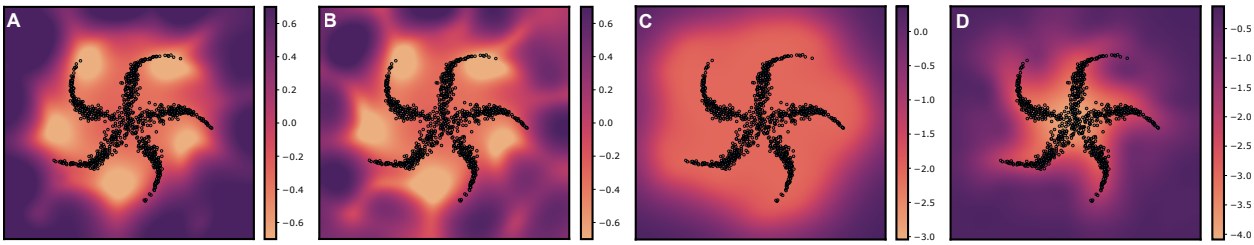

Figure 4: Difference of volume for data embedded in the latent space. (A) Riemannian volume measure, (B) Finslerian (Busemann-Hausdorff) volume measure, (A) Variance of the Gaussian process, (A) Ratio between the Riemannian and Finslerian volume: $(V_G(x) - \mathcal{V}(x))/V_G(x)$. All heatmaps are computed in logarithm scale.

In Figure 4, we can compare the volume measures obtained from the Riemannian and Finsler metrics, and in particular, their ratio in the top right image. When the metrics are computed next to the data points in area where the variance is very low, we can see that the ratio of the volume measure is at the order of magnitude 10^{-4}. Further away from the data points, the variance increases and so does the difference between the Riemannian and Finsler volume measures.

3.4 Results in high dimensions

Proposition 3.3 and Corrollary 3.2 indicate that the metrics become similar when the dimension (D) of the observational space increases. If we assume that the latent space is a bounded manifold, the metrics converge to each other at a rate of $\mathcal{O}\left(\frac{1}{D}\right)$, as do their functionals.

We need to assume that the latent manifold is bounded for two reasons: (1) we need to show that ω, which defines the Finsler metric, does not grow faster than the number of dimensions; and (2) we need to assume that the metrics are finite.

Corollary 3.3. In high dimensions, we have:

$$\frac{L_G(x) - \mathcal{L}(x)}{L_G(x)} = \mathcal{O}\left(\frac{1}{D}\right)$$
$$\frac{E_G(x) - \mathcal{E}(x)}{E_G(x)} = \mathcal{O}\left(\frac{1}{D}\right)$$
$$\frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} = \mathcal{O}\left(\frac{q}{D}\right)$$

And, when D converges toward infinity: $L_G \underset{+\infty}{\sim} \mathcal{L}$, $E_G \underset{+\infty}{\sim} \mathcal{E}$ and $V_G \underset{+\infty}{\sim} \mathcal{V}$.

Proof. This result follows from Corollary 3.2, assuming the latent manifold is bounded. □

Corollary 3.4. In high dimensions, the relative ratio between the Finsler metric: $F : (x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$ and the Riemmanian metric $g : (v, v) \to v^\top \mathbb{E}[G_x] v$ is:

$$\frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} = \mathcal{O}\left(\frac{1}{D}\right)$$

And, when D converges toward infinity: $\forall v \in \mathcal{T}_x M$, $\sqrt{g(v,v)} \underset{+\infty}{\sim} F_x(v)$.

Proof. Similarly, from Proposition 3.3, in a bounded manifold, both metrics converge to each other in high dimensions. □

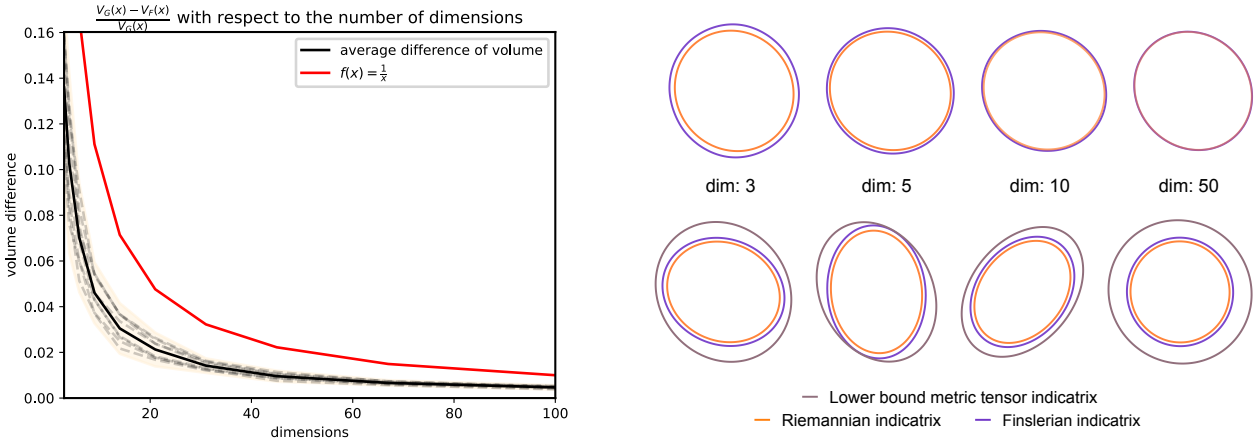

Figure 5: Left: Ratio of volumes $(V_G - \mathcal{V})/V_G$ varying with respect to the number of dimensions. The results were obtained from using a collection of matrices $\{G_i\}$ following a non-central Wishart distribution. Upper right: Finsler and Riemannian indicatrices converging towards each other as the number of dimensions increases. Lower right: $\alpha\Sigma$-indicatrices, Riemannian indicatrices and Finsler indicatrices illustrating the absolute bounds in Proposition 3.1.

4 Experiments

We want to illustrate cases where these metrics could be useful in practice for real-world data. For thi, we use three datasets: a synthetic dataset (composed of data representing a pinwheel projected onto a sphere), a font dataset Campbell & Kautz (2014), and a dataset representing single-cells stages Guo et al. (2010). We trained a GP-LVM model to learn a 2d-manifold. From the optimised Gaussian process, we can access the Riemannian and Finsler metric, and minimise their respective curve energies to obtain geodesics.

As we can see, the Finsler and Riemannian geodesics coincide in all cases. For all latent spaces (A.1. and B.1. in Figure 6, and in Figure 7), the heatmap represents the Riemannian volume measure in logarithm scale. The volume measure is low in the area of high density and high in the area of low density of data points.

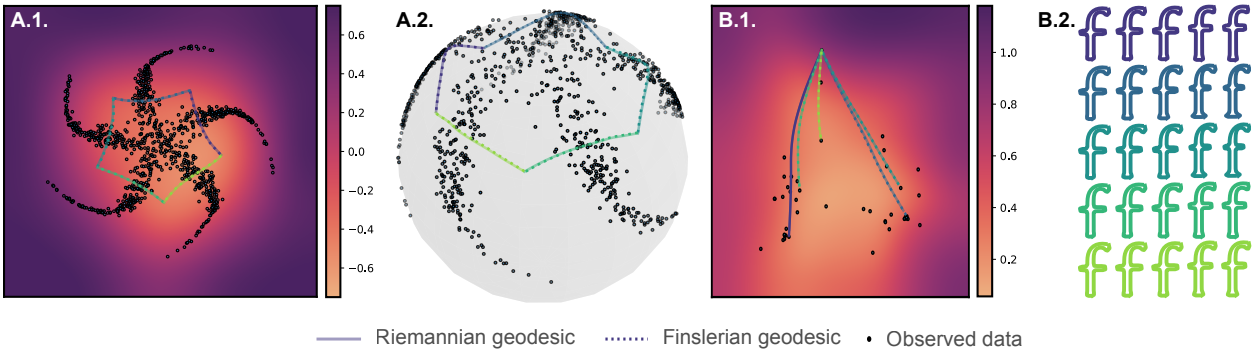

Figure 6: Geodesics computed for latent (A.1., B.1.) and observational (A.2., B.2.) spaces. (A) The dataset used was a pinwheel projected onto a sphere, as seen in A.2. (B) The dataset consists of the position of the markers parametrising the contour of the letter **f**.

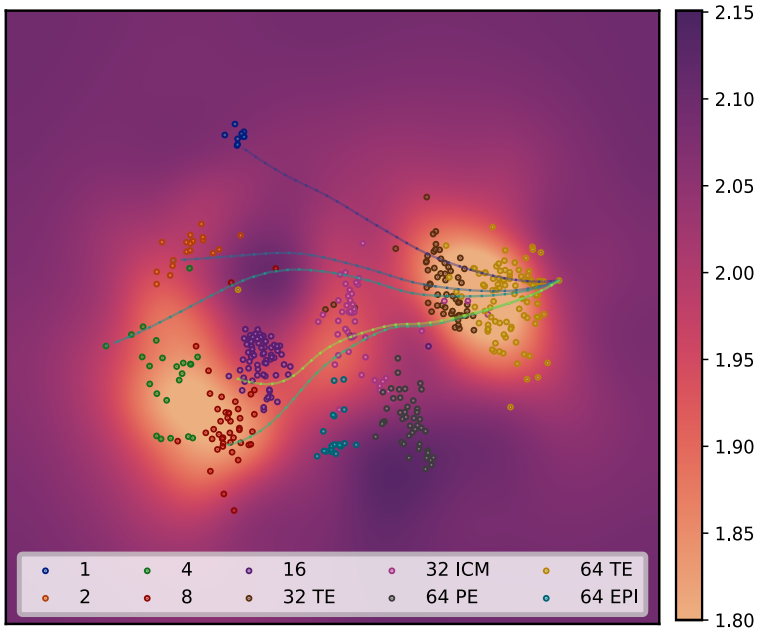

Figure 7: qPCR data

5 Discussion

Generative models are often used to reduce data dimension in order to better understand the mechanisms behind the data generating process. We consider the general setting where the mapping from latent variables to observations is driven by a smooth stochastic process, and the sample mappings span Riemannian manifolds. The Riemanniangeometry machinery has already been used in the past to explore the latent space.

In this paper, we have shown how curves and volumes can be identified by defining the length of a latent curve as its expected length measured in the observation space. This is a natural extension of classical differential geometric constructions to the stochastic realm. Surprisingly, we have shown that this does not give rise to a Riemannian metric over the latent space, even if sample mappings do. Rather, the latent representation naturally becomes equipped with a Finsler metric, implying that stochastic manifolds, such as those spanned by Latent Variable Models (LVMs), are inherently more complex than their deterministic counterparts.

The Finslerian view of the latent representation gives us a suitable general solution to explore a random manifold, but it does not immediately translate into a practical computational tool. As Riemannian manifolds are better understood computationally than Finsler manifolds, we have raised the question: How good an approximation of the Finsler metric can be achieved by a Riemannian metric? The answer turns out to be: quite good. We have shown that as data dimension increases, the Finsler metric becomes increasingly Riemannian. Since LVMs are most commonly applied to high-dimensional data (as this is where dimensionality reduction carries value), we have justification for approximating the Finsler metric with a Riemannian metric such that computational tools become more easily available. In practice we find that geodesics under the Finsler the Riemannian metric are near identical, except in regions of high uncertainty.

Notations

$\mathcal{M}$	Smooth differentiable manifold,
$\mathcal{T}_x M$	Tangent space of the manifold $\mathcal{M}$ at a point x,
D	Number of dimensions of the observational space,
q	Number dimensions of the latent space,
J_f	Jacobian of a stochastic function $f : \mathbb{R}^q \to \mathbb{R}^D$,
Σ_x	Covariance matrix of the Jacobian $J_x \sim \prod_{i=1}^{D} \mathcal{N}(\mu_i, \Sigma_x)$,
G_x	Stochastic metric tensor defined as; $G_x = J_x^\top J_x$,
F	Finsler metric: $F : \mathcal{T}_x M \to \mathbb{R} : v \to \mathbb{E}[\sqrt{v^\top G_x v}]$,
g	Riemannian metric: $g : (\mathcal{T}_x M, \mathcal{T}_x M) \to \mathbb{R} : (v, v) \to v^\top \mathbb{E}\left[G_x\right] v$,
$\mathcal{L}, \mathcal{E}, \mathcal{V}$	Length, energy and Busemann Hausdorff volume of the Finsler metric,
L_G, E_G, V_G	Length, energy and volume of the Riemannian metric with $\mathbb{E}[G]$ the expected metric tensor.

References

Sumon Ahmed, Magnus Rattray, and Alexis Boukouvalas. Grandprix: scaling up the bayesian gplvm for single-cell data. *Bioinformatics*, 35(1):47–54, 2019.

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. Latent Space Oddity: on the Curvature of Deep Generative Models. In *International Conference on Learning Representations (ICLR)*, 2018.

Georgios Arvanitidis, Miguel González-Duque, Alison Pouplin, Dimitris Kalatzis, and Søren Hauberg. Pulling back information geometry, 2021. URL `https://arxiv.org/abs/2106.05367`.

Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.

Neill D. F. Campbell and Jan Kautz. Learning a manifold of fonts. *ACM Trans. Graph.*, 33(4), jul 2014. ISSN 0730-0301. doi: 10.1145/2601097.2601212. URL `https://doi.org/10.1145/2601097.2601212`.

Nicki S. Detlefsen, Alison Pouplin, Cilie W. Feldager, Cong Geng, Dimitris Kalatzis, Helene Hauschultz, Miguel González Duque, Frederik Warburg, Marco Miani, and Søren Hauberg. Stochman. *GitHub. Note: https://github.com/MachineLearningLifeScience/stochman/*, 2021.

David Eklund and Søren Hauberg. Expected path length on random manifolds. *arXiv preprint arXiv:1908.07377*, 2019.

Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

Paul Finsler. *Ueber kurven und Flächen in allgemeinen Räumen.* Philos. Fak., Georg-August-Univ., 1918.

Guoji Guo, Mikael Huss, Guo Qing Tong, Chaoyang Wang, Li Li Sun, Neil D. Clarke, and Paul Robson. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental Cell*, 18(4):675–685, 2010. ISSN 1534-5807. doi: https://doi.org/10.1016/j.devcel.2010.02.012. URL `https://www.sciencedirect.com/science/article/pii/S1534580710001103`.

Søren Hauberg. The non-central Nakagami distribution. Technical report, Technical University of Denmark, 2018. URL `http://www2.compute.dtu.dk/~sohau/papers/nakagami2018/nakagami.pdf`.

John T. Kent and R. J. Muirhead. Aspects of Multivariate Statistical Theory. *The Statistician*, 1984. ISSN 00390526. doi: 10.2307/2987858.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in neural information processing systems*, 16, 2003.

John M Lee. Smooth manifolds. In *Introduction to smooth manifolds*, pp. 1–31. Springer, 2013.

J. G. Liao and Arthur Berg. Sharpening jensen's inequality. *The American Statistician*, 73(3):278–281, 2019. doi: 10.1080/00031305.2017.1419145.

Federico Lopez, Beatrice Pozzetti, Steve Trettel, Michael Strube, and Anna Wienhard. Symmetric spaces for graph embeddings: A finsler-riemannian approach. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7090–7101. PMLR, 18–24 Jul 2021. URL `https://proceedings.mlr.press/v139/lopez21a.html`.

Steen Markvorsen. A finsler geodesic spray paradigm for wildfire spread modelling. *Nonlinear Analysis: Real World Applications*, 28:208–228, 2016.

Pyro. Gaussian process latent variable modelű, 2022. URL `https://pyro.ai/examples/gplvm.html`.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

Nathan D. Ratliff, Karl Van Wyk, Mandy Xie, Anqi Li, and Muhammad Asif Rana. Generalized nonlinear and finsler geometry for robotics. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 10206–10212, 2021. doi: 10.1109/ICRA48506.2021.9561543.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pp. 1278–1286. PMLR, 2014.

Yi-Bing Shen and Zhongmin Shen. *Introduction to Modern Finsler Geometry*. Co-published with HEP, 2016. doi: 10.1142/9726. URL `https://www.worldscientific.com/doi/abs/10.1142/9726`.

Alessandra Tosi, Søren Hauberg, Alfredo Vellido, and Neil D Lawrence. Metrics for probabilistic geometries. *arXiv preprint arXiv:1411.7432*, 2014.

J. G. Wendel. Note on the gamma function. *The American Mathematical Monthly*, 55(9):563–564, 1948. ISSN 00029890, 19300972. URL `http://www.jstor.org/stable/2304460`.

Bingye Wu. Volume form and its applications in finsler geometry. *Publicationes Mathematicae*, 78, 04 2011. doi: 10.5486/PMD.2011.4998.

A A primer on Geometry

The main purpose of the paper is to define and compare two legitimate metrics to compute the average length between random points. Before going further, it's important to formally define the two metrics (Riemannian and Finsler metrics, respectively) which we do in sections A.2 and A.3. They are both constructed on topological manifolds, the definition of which is recalled in section A.1. We finally introduce the notion of random manifold in section A.4, which is the last notion needed to frame our problem of interest: which metric should we use to compute the average distance on a random manifold?

A.1 Topological and differentiable manifolds

This section aims to define core concepts in differential geometry that will be used later to define Riemannian and Finsler manifolds. Recall that two topological spaces are called homeomorphic if there is a continuous bijection between them with continuous inverse.

Definition A.1. A d-dimensional **topological manifold** $\mathcal{M}$ is a second-countable Hausdorff topological space such that every point has an open neighbourhood homeomorphic to an open subset of $\mathbb{R}^d$.

Let $\mathcal{M}$ be a topological manifold. This means that for any $x \in \mathcal{M}$ there is an open neighbourhood U_x of x and a homeomorphism $\phi_{U_x} : U_x \to \mathbb{R}^d$ onto an open subset of $\mathbb{R}^d$. Suppose that $x, y \in \mathcal{M}$ are such that $U_x \cap U_y \neq \emptyset$, let $U = U_x$, $V = U_y$ and consider the so-called coordinate change map

$$\phi_V \circ \phi^{-1}_{U|\phi_U(U\cap V)} : \phi_U(U \cap V) \to \mathbb{R}^d.$$

We call $\mathcal{M}$ together with an open cover $\{U_x\}_{x\in\mathcal{M}}$ as above a **differentiable** or **smooth** manifold if the coordinate maps are infinitely differentiable.

Beyond these technical definitions, one can imagine a differentiable manifold as a well-behaved smooth surface that possesses *locally* all the topological properties of a Euclidean space. All the manifolds in this paper are assumed to be differentiable and connected manifolds.

Definition A.2. We also define, for a differentiable manifold $\mathcal{M}$, the **tangent space** $\mathcal{T}_xM$ as the set of all the tangent vectors at $x \in \mathcal{M}$, and the **tangent bundle** $\mathcal{TM}$ the disjoint union of all the tangent spaces: $\mathcal{TM} = \bigcup_{x\in\mathcal{M}} \mathcal{T}_xM$.

So far, we have only defined topological and differential properties of manifolds. In order to compute geometric quantities, we need to equip those with a metric that helps us derive useful quantities such as lengths, energies and volumes. A metric is a scalar valued function that is defined for each point on the topological manifold and takes as inputs one or two vectors (depending on the type of metric) from the tangent space at the specific point. Such a function can either be defined as a scalar product between two vectors, this is the case of a Riemannian metric or, in the case of a Finsler metric, it is defined similarly to the norm of a vector. We will formally define these metrics and highlight their differences in the following sections.

A.2 Riemannian manifolds

Definition A.3. Let $\mathcal{M}$ be a manifold. A **Riemannian metric** is a map assigning at each point $x \in \mathcal{M}$ a scalar product $g_x(\cdot,\cdot) : \mathcal{T}_xM \times \mathcal{T}_xM \to \mathbb{R}$, with g_x a positive definite bilinear map, which is smooth with respect to x. A smooth manifold equipped with a Riemannian metric is called a **Riemannian manifold**. We usually express the metric as a symmetric positive definite matrix G_x, where we have for two vectors $u, v \in \mathcal{T}_xM$: $g_x(u, v) = \langle u, v\rangle_{G_x} = u^\top G_x v$. We further define the induced **norm**: $v \in \mathcal{T}_xM, \|v\|_G = \sqrt{g_x(v, v)}$.

The Riemannian metric here can either refer to the scalar product g_x itself, or the associated metric tensor G_x.

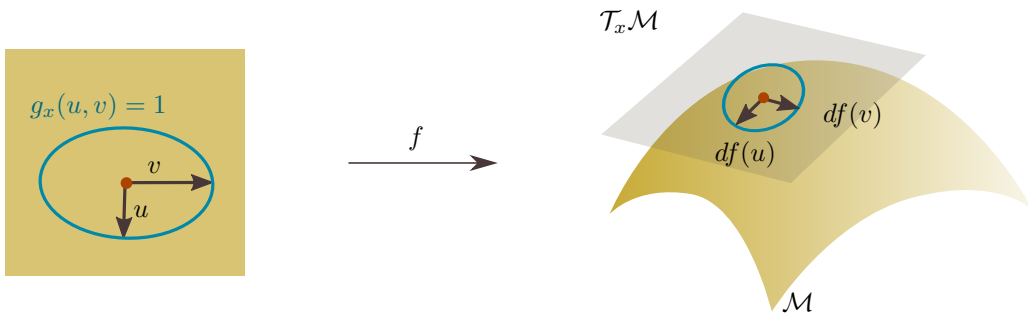

Figure 8: f is an immersion that maps a low dimensional manifold to a high dimensional manifold $\mathcal{M}$. On $\mathcal{M}$, a tangent plane $\mathcal{T}_x\mathcal{M}$ is draw at x. The indicatrice of the Euclidean metric is plotted in blue. When this metric is pulled-back through f, the low dimensional space is now equipped with the pullback metric g, which is a Riemannian metric by definition. The vectors $df(u)$ and $df(v)$ are called the push-forwards of the vectors u and v through f.

Definition A.4. We consider a curve $\gamma(t)$ and its derivative $\dot{\gamma}(t)$ on a Riemannian manifold $\mathcal{M}$ equipped with the metric g. Then, we define the **length of the curve**:

$$L_G(\gamma) = \int \left\|\dot{\gamma}(t))\right\|_G dt = \int \sqrt{g_t(\dot{\gamma}(t), \dot{\gamma}(t))}\, dt,$$

where $g_t = g_{\gamma(t)}$. Locally length-minimising curves between two connecting points are called **Geodesics**.

Definition A.5. The **curve energy** is defined as:

$$E_G(\gamma) = \int \left\|\dot{\gamma}(t))\right\|_G^2 dt = \int g_t(\dot{\gamma}(t), \dot{\gamma}(t))\, dt.$$

There are two interesting properties to note about the length of a curve and the curve energy. First, the length is parametrisation invariant: for any bijective smooth function η on the domain of γ we have that $L_G(\gamma \circ \eta) = L_G(\gamma)$. We also say the Riemannian metric gives us intrinsic coordinates to compute the length. Secondly, for a given curve γ, we have: $L_G(\gamma)^2 \leq 2E_G(\gamma)$. Because of the invariance of the curve, when we aim to minimise it, a solver can find an infinite number of solutions. On the other hand, the curve energy is convex and will lead to a unique solution. Thus, to obtain a geodesic, instead of solving the corresponding ODE equations, or directly minimising lengths, it is easier in practice to minimise the curve energy, as a minimal energy gives a minimal length.

The Riemannian metric also provides us with an infinitesimal volume element that relates our metric G to an orthonormal basis, the same way the Jacobian determinant accommodate for a change of coordinates in the change of variables theorem.

Definition A.6. In local coordinates $(e^1, \cdots, e^d)$, the **volume form** of the Riemannian manifold $\mathcal{M}$, equipped with the metric tensor G, is defined as: $dV_G = V_G(x)e^1 \wedge \cdots \wedge e^d$, with:

$$V_G(x) = \sqrt{\det(G_x)}.$$

Remark. *The symbol $\wedge$ represents the wedge product and it is used to manipulate differential k-forms. Here, the basis vectors $(e^1, \cdots, e^d)$ form a d-dimensional parallelepiped $(e^1 \wedge \cdots \wedge e^d)$ with unit volume.*

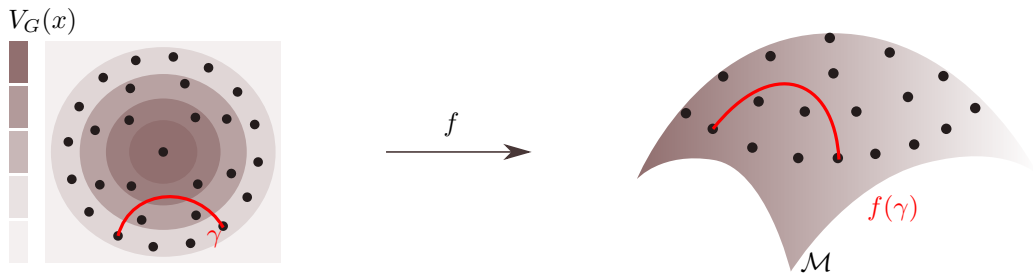

Figure 9: Once the low dimensional manifold is equipped with a metric that captures the inherent structure of the high dimensional manifold, we can compute a geodesic γ, by minimising the energy functional between two points. The geodesic $f(\gamma)$ will be the shortest path between two points on the manifold $\mathcal{M}$. The volume measure V_G can also be used to integrate functions over regions of the manifold, as we would do in the Euclidean space. It can also be linked to the density of the data: if the data points are uniformly distributed over the high-dimensional manifold, in the low-dimensional manifold, a low volume would correspond to a high density of data. It is a useful way to give more information about the distribution of the data.

A.3 Finsler manifolds

Finsler geometry is often described as an extension of Riemannian geometry, since the metric is defined in a more general way, lifting the quadratic constraint. In particular, the norm of a Riemmanian metric is a Finsler metric, but the converse is not true.

Definition A.7. Let $F : \mathcal{TM} \to \mathbb{R}_+$ be a continuous non-negative function defined on the tangent bundle $\mathcal{TM}$ of a differentiable manifold M. We say that F is a **Finsler metric** if, for each point x of $\mathcal{M}$ and v on $\mathcal{T}_xM$, we have:

1. Positive homogeneity: $\forall \lambda \in \mathbb{R}_+$, $F(\lambda v) = \lambda F(v)$.
2. Smoothness: F is a C^∞ function on the slit tangent bundle $\mathcal{TM} \setminus \{0\}$.
3. Strong convexity criterion: the Hessian matrix $g_{ij}(v) = \frac{1}{2}\frac{\partial^2 F^2}{\partial v^i v^j}(v)$ is positive definite for non-zero v.

A differentiable manifold $\mathcal{M}$ equipped with a Finsler metric is called a **Finsler manifold**.

Here, it is worth noting that, for a given point in the manifold, the Finsler metric is defined with only one vector in the tangent space, while the Riemannian metric is defined with two vectors. Moreover, from the previous definition, we can deduce that the metric is:

1. Positive definite: for all $x \in \mathcal{M}$ and $v \in \mathcal{T}_xM$, $F(v) \geq 0$ and $F(v) = 0$ if and only if $v = 0$.
2. Subadditive: $F(v + w) \leq F(v) + F(w)$ for all $x \in \mathcal{M}$ and $v, w \in \mathcal{T}_xM$.

We say that F is a Minkowski norm on each tangent space $\mathcal{T}_xM$. Furthermore, if F satisfies the reversibility property: $F(v) = F(-v)$, it defines a norm on $\mathcal{T}_xM$ in the usual sense.

Similarly to Riemannian geometry, lengths, energies and volumes can be defined directly from the Finsler metric:

Definition A.8. We consider a curve γ and its derivative $\dot{\gamma}$ on a Finsler manifold $\mathcal{M}$ equipped with the metric F. We define the **length of the curve** as follows:

$$\mathcal{L}(\gamma) = \int F(\dot{\gamma}(t))dt.$$

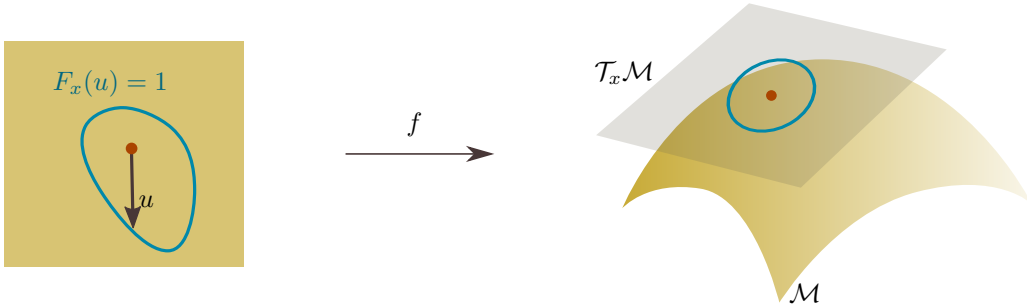

Figure 10: f is an immersion that maps a low dimensional manifold to a high dimensional manifold $\mathcal{M}$. On $\mathcal{M}$, a tangent plane $\mathcal{T}_x\mathcal{M}$ is draw at x. Compared to the Riemannian manifold, the Finsler indicatrix, which represents the all the vectors $u \in \mathcal{T}_x\mathcal{M}$ such that $F_x(u) = 1$, is not necessarily an ellipse. It can be asymmetric if the metric is asymmetric itself. It is always convex.

Definition A.9. The **curve energy** is defined as:

$$\mathcal{E}(\gamma) = \int F(\dot{\gamma}(t))^2 dt.$$

Not only are the definitions strikingly similar, they also share the same properties. The curve length is also invariant under reparametrisation, and upper bounded by the curve energy. Computing geodesics on a manifold is reduced to a variational optimisation problem. These propositions are proved in detail in Lemmas B.1.4 and B.1.5, in the appendix.

In Riemannian geometry, the volume measure defined by the metric is unique. In Finsler geometry, different definitions of the volume exist, and they all coincide with the Riemannian volume element when the metric is Riemannian. The most common choices of volume forms are the Busemann-Hausdorff measure and the Holmes-Thompson measure. According Wu (2011), depending on the Finsler metric and the topological manifold, some choices seem more legitimate than others. In this paper, we decided to only focus on the Busemann-Hausdorff volume, as its definition is the most commonly used and leads to easier derivations. We will later show that in high dimensions, our Finsler metric converges to a Riemannian metric, and thus, the results obtained for the Busemann-Hausdorff volume measure are also valid for the Holmes-Thomson volume measure.

Definition A.10. For a given point x on the manifold, we define the **Finsler indicatrix** as the set of vectors in the tangent space such that the Finsler metric is equal to one: $\{v \in \mathcal{T}_x M | F(v) = 1\})$. We denote the Euclidean unit ball in $\mathbb{R}^d$ by $\mathbb{B}^d(1)$ and for measurable subsets $S \subseteq \mathbb{R}^d$ we use $\text{vol}(S)$ to denote the standard Eulcidean volume of S. In local coordinates $(e^1, \cdots, e^d)$ on a Finsler manifold $\mathcal{M}$, the **Busemann-Hausdorff volume** form is defined as $d\mathcal{V} = \mathcal{V}(x) e^1 \wedge \cdots \wedge e^d$, with:

$$\mathcal{V}(x) = \frac{\text{vol}(\mathbb{B}^d(1))}{\text{vol}(\{v \in \mathcal{T}_x M | F(v) < 1\})}.$$

We can interpret the volume as the ratio between the euclidean ball, and a convex ball whose radius is defined as a unit Finsler metric. If the Finsler metric is replaced by a Riemannian metric, the volume of the indicatrix will be an ellipsoid whose semi-axis are equal to the inverse of the squareroot of the metric's eigenvalues. The Finsler volume then reduces to the definition of the Riemannian volume.

A.4 Random manifolds

So far, we have only considered deterministic data points lying on a manifold. If we consider our data to be random variables, we will need to define the associated random metric and manifold.

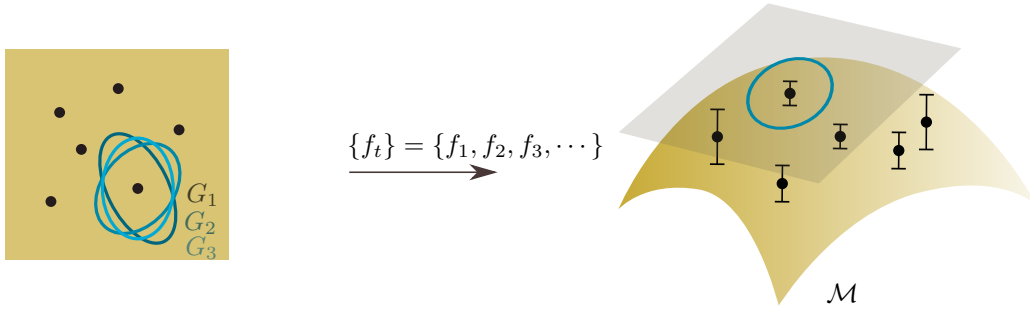

Figure 11: Usually the immersion f would be deterministic. In the case of most generative models, where f is described by a GP-LVM, or the decoder of a VAE, the immersion is stochastic. The pullback metric is stochastic de facto.

As said previously, if we have a function $f : \mathbb{R}^q \to \mathbb{R}^D$ that parametrises a manifold, then we can construct a Riemannian metric $G_x = J_f^\top J_f$, with J_f the Jacobian of the function f. In the previous cases, we assumed f to be a deterministic function, and so is the metric. We construct a stochastic Riemannian metric in the same way, with f being a stochastic process. A stochastic process $f : \mathbb{R}^q \to \mathbb{R}^D$ is a random map in the sense that samples of the process are maps from $\mathbb{R}^q$ to $\mathbb{R}^D$ (the so-called sample paths of the process).

Definition A.11. A stochastic process $f : \mathbb{R}^q \to \mathbb{R}^D$ is smooth if the sample paths of f are smooth. We call a smooth process f a **stochastic immersion** if the Jacobian matrix of its sample paths has full rank everywhere. We can then define the **stochastic Riemannian metric** $G_x = J_f^\top J_f$.

The definition of the stochastic immersion is fairly important, as it means that its Jacobian is full rank. Since the Jacobian is full rank, the random metric G_x is positive definite, a necessary condition to define a Riemannian metric. Another definition of a stochastic Riemannian metric would be the following:

Definition A.12. A **stochastic Riemannian metric** on $\mathbb{R}^q$ is a matrix-valued random field on $\mathbb{R}^q$ whose sample paths are Riemannian metrics.

Here, we insist on the fact that any matrices drawn from this stochastic metric would be a proper Riemannian metric. Finally, we define a random manifold:

Definition A.13. A **random manifold** is a differentiable manifold equipped with a random Riemannian metric.

When using the random Riemannian metric on two vectors $u, v \in \mathcal{T}_x M$, $g_x(u, v) = u^\top G_x v$ is a random variable, but both u, v are deterministic vectors. From this definition, it follows that the length, the energy and the also volume are random variables.

B Proofs

B.1 Finslerian geometry of the expected length

In this section, we will always let $f : \mathbb{R}^q \to \mathbb{R}^D$ be a stochastic immersion, J_f its Jacobian, and $G_x = J_f^\top J_f$ a metric tensor. We will first prove that the function $F : \mathcal{T}M \to \mathbb{R} : v \to \mathbb{E}\left[\sqrt{v^\top G_x v}\right]$ is a Finsler metric. Then, for the specific case where J_f follows a non-central normal distribution, the Finsler metric F defined as the expected length follows a non-central Nakagami distribution and can be expressed in closed form.

To prove that the function F is indeed a Finsler metric, we will need to verify the criteria above, among them the strong convexity criterion is less trivial to prove than the others. It will be detailed in Lemma B.1.3.

Object	Riemann	Finsler
metric	$g_{ij} : \mathcal{T}_x M \times \mathcal{T}_x M \to \mathbb{R}$	$F : \mathcal{T}M \to \mathbb{R}_+$
length structure	$L_G(\gamma) = \int \sqrt{g_t(\dot\gamma(t), \dot\gamma(t))}\, dt$	$\mathcal{L}(\gamma) = \int F(\dot\gamma(t))\, dt$
energy structure	$E_G(\gamma) = \int g_t(\dot\gamma(t), \dot\gamma(t))\, dt$	$\mathcal{E}(\gamma) = \int F(\dot\gamma(t))^2\, dt$
volume element	$V_G(x) = \sqrt{\lvert \det\{G_x\}\rvert}$	$\mathcal{V}(x) = \mathrm{vol}(\mathbb{B}^n(1))/\mathrm{vol}(\{v \in \mathcal{T}_x M \lvert F(x,v) < 1\})$ Busemann-Hausdorff volume measure

Table 1: Comparison of Riemannian and Finsler geometry.

Strong convexity means that the Hessian matrix $\frac{1}{2}\mathrm{Hess}(F(v)^2) = \frac{1}{2}\frac{\partial^2 F^2}{\partial v^i v^j}(v)$ is strictly positive definite for non-negative v. This matrix, when F is a Finsler function, is also called the fundamental form and plays an important role in Finsler geometry. To prove the strong convexity criterion, we will need the full expression of the fundamental form, detailed in Lemma B.1.1.

Lemma B.1.1. *The Hessian matrix $\frac{1}{2}Hess(F(v)^2)$ of the function $F(v) = \mathbb{E}\left[\sqrt{v^\top G_x v}\right]$ is given by*

$$\frac{1}{2}Hess(F(v)^2) = \mathbb{E}\left[(v^\top G_x v)^{\frac{1}{2}}\right]\mathbb{E}\left[(v^\top G_x v)^{-\frac{1}{2}}G_x - (v^\top G_x v)^{-\frac{3}{2}}G_x v v^\top G_x\right] + \mathbb{E}\left[(v^\top G_x v)^{-\frac{1}{2}}G_x\right]^2 vv^\top.$$

Proof. Let G be a random positive definite symmetric matrix and define $g : \mathbb{R}^q \to \mathbb{R} : v \mapsto \sqrt{v^\top G v}$, where v is considered a column vector. We would like to know the different derivatives of g with respect to v. We name by default J_g and H_g, its Jacobian and Hessian matrix. Using the chain rule, we have: $J_g = (v^\top G v)^{-\frac{1}{2}} v^\top G$ and $H_g = (v^\top G v)^{-\frac{1}{2}} G - (v^\top G v)^{-\frac{3}{2}}(G v v^\top G)$.

For the rest of the proof, we need to show that derivatives and expectation values commute.

Using the Fubini theorem, we can show that tha derivatives and the expectation values commute.

For $F : \mathbb{R}^q \to \mathbb{R} : v \mapsto \mathbb{E}[\sqrt{v^\top G v}]$,

$$\mathrm{Hess}(F) = \mathbb{E}[H_g] = \mathbb{E}\left[(v^\top G v)^{-\frac{1}{2}}G - (v^\top G v)^{-\frac{3}{2}} G v v^\top G\right]$$

$$\nabla F = \mathbb{E}[J_g] = \mathbb{E}[(v^\top G v)^{-\frac{1}{2}} G v].$$

We now consider the function $h : \mathbb{R}^q \to \mathbb{R} : v \mapsto \mathbb{E}[\sqrt{v^\top G v}]^2 = F(v)^2$. Using the chain rule and changing the order of expectation and derivatives, we have its Hessian

$$H_h = 2F \cdot \mathrm{Hess}[F] + 2\nabla F^\top \nabla F = 2\mathbb{E}[g]\mathbb{E}[H_g] + 2\mathbb{E}[J_g]^\top \mathbb{E}[J_g].$$

Finally, replacing J_g and H_g previously obtained in this expression, we conclude:

$$\frac{1}{2}H_h(x,v) = \mathbb{E}\left[(v^\top G v)^{\frac{1}{2}}\right]\mathbb{E}\left[(v^\top G v)^{-\frac{1}{2}}G - (v^\top G v)^{-\frac{3}{2}} G v v^\top G\right] + \mathbb{E}\left[(v^\top G v)^{-\frac{1}{2}}G\right]^2 vv^\top.$$

□

Remark. *Before going further, it's important to note that $G_x = J_f^\top J_f$ is a random matrix that is positive definite: it is symmetric by definition and has full rank. The later statement is justified by the assumption that the stochastic process $f : \mathbb{R}^q \to \mathbb{R}^D$ is an immersion, then J_f is full rank.*

Lemma B.1.2. *The function $F(v) = \mathbb{E}\left[\sqrt{v^\top G_x v}\right]$ is:*

1. *positive homogeneous:* $\forall \lambda \in \mathbb{R}_+$, $F(\lambda v) = \lambda F(v)$
2. *smooth:* $F(v)$ *is a* C^∞ *function on the slit tangent bundle* $\mathcal{TM} \setminus \{0\}$

Proof. 1) Let $\lambda \in \mathbb{R}$, then we have: $F(\lambda v) = \mathbb{E}\left[\sqrt{\lambda^2 v^\top G_x v}\right] = |\lambda| \left[\sqrt{v^\top G_x v}\right]$.

2) The multivariate function: $\mathbb{R}^q \backslash \{0\} \to \mathbb{R}_+^* : v \to v^\top G_x v$ is C^∞ and strictly positive, since $G_x = J_f^\top J_f$ is positive definite. The function $\mathbb{R}_+^* \to \mathbb{R}_+^* : x \to \sqrt{x}$ is also C^∞. Finally, $\mathbb{R}_+^* \to \mathbb{R}_+^* : x \to \mathbb{E}[x]$ is by definition differentiable. By composition, $F(v)$ is a C^∞ function on the slit tangent bundle $\mathcal{TM} \setminus \{0\}$. □

Lemma B.1.3. *The function* $F(v) = \mathbb{E}\left[\sqrt{v^\top G_x v}\right]$ *satisfies the strong convexity criterion.*

Proof. Proving that F satisfies the strong convexity criterion is equivalent to show that the Hessian matrix $H = \frac{1}{2}\text{Hess}(F(v)^2)$ is strictly positive definite. Thus, we need to prove that $\forall w \in \mathbb{R}^q \backslash \{0\}, w^\top H w > 0$. According to Lemma B.1.1, because the expectation is a positive function, it's straightforward to see that $\forall w \in \mathbb{R}^q \backslash \{0\}, w^\top H w \geq 0$. The tricky part of this proof is to show that $w^\top H w > 0$. This can be obtained if one of the terms ($F \cdot \text{Hess}(F)$ or $\nabla F^\top \nabla F$) is strictly positive.

First, let's decompose H as the sum of matrices: $H = F\text{Hess}(F) + \nabla F^\top \nabla F$ (Lemma B.1.1), with:

$$F \cdot \text{Hess}(F) = \mathbb{E}\left[(v^\top G v)^{\frac{1}{2}}\right] \mathbb{E}\left[(v^\top G v)^{-\frac{3}{2}} \left((v^\top G v)G - Gv(Gv)^\top\right)\right],$$

$$\nabla F^\top \nabla F = \mathbb{E}\left[(v^\top G v)^{-\frac{1}{2}} G\right]^2 vv^\top.$$

We will study two cases: when $w \in \text{span}(v)$, and when $w \notin \text{span}(v)$. We will always assume that $v \neq 0$, and so by definition: $F(v) > 0$.

Let $w \in \text{span}(v)$. We will show that $w^\top \nabla F^\top \nabla F w > 0$. We have $w = \alpha v, \alpha \in \mathbb{R}$. Because F is 1-homogeneous and using Euler theorem, we have: $\nabla F(v) v = F(v)$. Then $(\alpha v)^\top \nabla F^\top \nabla F(\alpha v) = \alpha^2 F^2$, and $\alpha^2 F(v)^2 > 0$.

Let $w \notin \text{span}(v)$. F being a scalar function, we have: $w^\top F\text{Hess}[F]w = Fw^\top \text{Hess}[F]w$. We would like to show that: $w^\top \text{Hess}[F]w > 0$. The strategy is the following: if we prove that the kernel of $\text{Hess}[F]$ is equal to the $\text{span}(v)$, then $w \notin \text{span}(v)$ is equivalent to say that $w \notin \ker(\text{Hess}[F])$ and we can conclude that: $w^\top \text{Hess}[F]w > 0$. Let's prove $\text{span}(v) \in \ker(\text{Hess}(F))$. We know that $\text{Hess}(F)v = 0$, since F is 1-homogeneous, so we have $\text{span}(v) \in \ker(\text{Hess}(F))$. To obtain the equality, we just need to prove that the dimension of the kernel is equal to 1. Let $z \in \text{span}(v^\top G)^\top$, which is $(Gv)^T z = 0$. We have $\dim(\text{span}(v^\top M)) = 1$, and thus: $\dim(\text{span}(v^\top G)^\top) = q - 1$. Furthermore, $z^\top \text{Hess}[F]z = z^\top \mathbb{E}\left[M(v^\top M v)^{-\frac{1}{2}}\right] z > 0$, so we can deduce that $\dim(\text{im}(\text{Hess}[F])) = q - 1$. Using the Rank-Nullity theorem, we conclude that $\dim(\ker(\text{Hess}(F))) = q - \dim(\text{im}(\text{Hess}[F])) = 1$, which concludes the proof.

In conclusion, $\forall w \in \mathbb{R}^q \backslash \{0\}, w^\top \frac{1}{2}\text{Hess}(F(v)^2)w > 0$. The function F satisfies the strong convexity criterion. □

Proposition 2.2. Let G be a stochastic Riemannian metric. Then, the function:

$$F_x : \mathcal{T}_x\mathcal{M} \to \mathbb{R} : v \to \mathbb{E}\left[\sqrt{v^\top G(x) v}\right]$$

defines a Finsler metric, but it is not induced by a Riemannian metric.

Proof. Let's define F as a Riemannian metric: $F : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R} : (v_1, v_2) \to \mathbb{E}\left[\sqrt{v_1^\top G_x v_2}\right]$. If F were a Riemannian metric, then it would be bilinear, which is clearly not the case. Thus, F is not a Riemannian metric. According to Lemma B.1.2 and Lemma B.1.2, F is a Finsler metric. □

Proposition 2.3. Let f be a Gaussian process and J its Jacobian.

Then $G = J^\top J \sim \mathcal{W}_q(D, \Sigma, \mathbb{E}[J]^\top \mathbb{E}[J])$ follows a non-central Wishart distribution. The Finsler metric can be written as:

$$F_x : \mathcal{T}_x\mathcal{M} \to \mathbb{R} : v \to \sqrt{2}\sqrt{v^\top \Sigma v}\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}{}_1F_1\left(-\frac{1}{2}, \frac{D}{2}, -\frac{\omega}{2}\right),$$

with ${}_1F_1$ as the confluent hypergeometric function of the first kind and $\omega = (v^\top \Sigma_x v)^{-1}(v^\top \mathbb{E}[J]^\top \mathbb{E}[J]v)$.

Proof. The objective of the proof is to show that, if the Jacobian J_f follows a non-central normal distribution, then, $\forall v \in \mathbb{R}^q$, the expectation $\mathbb{E}[v^\top J_f^\top J_f v]$ will follow a non-central Nakagami distribution. This is a particular case of the derivation of moments of non-central Wishart distributions, previously shown and studied by Kent & Muirhead (1984); Hauberg (2018).

By hypothesis, J_f follows a non-central normal distribution: $J_f \sim \mathcal{N}(\mathbb{E}[J], I_D \otimes \Sigma_x)$. Then, $G_x = J_f^\top J_f$ follows a non-central Wishart distribution: $G_x \sim \mathcal{W}_d(D, \Sigma_x, \Sigma_x^{-1}\mathbb{E}[J]^\top \mathbb{E}[J])$. According to (Kent & Muirhead, 1984, Theorem 10.3.5.), $v^\top G_x v$ will also follow a non-central Wishart distribution: $v^\top G_x v \sim \mathcal{W}_1(D, v^\top \Sigma_x v, \omega)$, with: $\omega = (v^\top \Sigma_x v)^{-1}(v^\top \mathbb{E}[J]^\top \mathbb{E}[J]v)$.

To compute $\mathbb{E}[\sqrt{v^\top G_x v}]$, we shall look at the derivation of moments. (Kent & Muirhead, 1984, Theorem 10.3.7.) states that: if $X \sim \mathcal{W}_q(D, \Sigma, \Omega')$, with $q \leq D$, then $\mathbb{E}[(\det(X))^k] = (\det\{\Sigma\})^k 2^{qk}\frac{\Gamma_q(\frac{D}{2}+k)}{\Gamma_q(\frac{D}{2})}{}_1F_1(-k, \frac{D}{2}, -\frac{1}{2}\Omega')$. We directly apply the theorem to our case, knowing that $v^\top G_x v$ is a scalar term, so $\det\big(v^\top G_x v\big) = v^\top G_x v$, $q = 1$, and $k = \frac{1}{2}$:

$$\mathbb{E}[\sqrt{v^\top G_x v}] = \sqrt{2}\sqrt{v^\top \Sigma v}\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}{}_1F_1(-\frac{1}{2}, \frac{D}{2}, -\frac{1}{2}\omega)$$

□

Lemma B.1.4. *The length of a curve using a Finsler metric is invariant by reparametrisation.*

Proof. The proof is similar to the one obtained on a Riemannian manifold (Lee (2013), Proposition 13.25), where we make use of the homogeneity property of the Finsler metric.

Let $(\mathcal{M}, F)$ be a Finsler manifold and $\gamma : [a, b] \to \mathcal{M}$ a piecewise smooth curve segment. We call $\tilde{\gamma}$ a reparametrisation of γ, such that $\tilde{\gamma} = \gamma \circ \phi$ with $\phi : [c, d] \to [a, b]$ a diffeomorphism. We want to show that $\mathcal{L}(\gamma) = \mathcal{L}(\tilde{\gamma})$.

$$\begin{aligned}\mathcal{L}(\tilde{\gamma}) &= \int_c^d F(\dot{\tilde{\gamma}}(t))\,dt = \int_c^d F(\frac{d}{dt}(\gamma \circ \phi(t)))\,dt \\ &= \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} |\dot{\phi}(t)| F(\dot{\gamma} \circ \phi(t))\,dt = \int_a^b F(\dot{\gamma}(t))\,dt = \mathcal{L}(\gamma)\end{aligned}$$

□

Lemma B.1.5. *If a curve globally minimizes its energy on a Finsler manifold, then it also globally minimizes its length and the Finsler function F of the velocity vector along the curve is constant.*

Proof. The curve energy and the curve length are defined as: $\mathcal{E}(\gamma) = \int_0^1 F^2(\dot{\gamma}(t))dt$ and $\mathcal{L}(\gamma) = \int_0^1 F(\dot{\gamma}(t))dt$, with $\gamma : [0, 1] \to \mathbb{R}^d$. Let's define f and g two real-valued functions such that: $f : \mathbb{R} \to \mathbb{R} : t \mapsto F(\dot{\gamma}(t))$ and $g : \mathbb{R} \to \mathbb{R} : t \mapsto 1$. Applying Cauchy-Schwartz inequality, we directly obtain:

$$\left(\int_0^1 F(\dot{\gamma}(t))dt\right)^2 \leq \int_0^1 F(\dot{\gamma}(t))^2 dt \cdot \int_0^1 1^2 dt, \quad \text{which means:} \quad \mathcal{L}(\gamma)^2 \leq \mathcal{E}(\gamma).$$

The equality is obtained exactly when the functions f and g are proportional, hence, when the Finsler function is constant. $\square$

B.2 Comparison of Riemannian and Finsler metrics

We have defined both a Riemannian ($g : (v_1, v_2) \to v_1^\top \mathbb{E}[G_x] v_2$) and a Finsler ($F : (x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$) metric, in the hope to compute the average length between two points on a random manifold created by the random field f: $G_x = J_f^\top J_f$. The main idea of this section is to better compare those two metrics and in what extend they differ in terms of length, energy and volume. From now on, $f : \mathbb{R}^q \to \mathbb{R}^D$ will always be defined as a stochastic non-central gaussian process. Its Jacobian J_f also follows a non-central gaussian distribution, $G_x = J_f^\top J_f$ a non-central Wishart distribution, and $F : (x, v) = \mathbb{E}[\sqrt{v^\top G_x v}]$ a non-central Nakagami distribution (Proposition 2.3). The Finsler metric can be written in closed form.

In section B.2.1, we will see that the Finsler metric is upper and lower bounded by two Riemannian tensors (Proposition 3.1), and we can deduce an upper and lower bound for the length, the energy and the volume (Corollary 3.1). Then, in section B.2.2, we will show that the relative difference ($\Delta(g, F) = g(v, v) - F(v)/g(v, v)$) between the Finsler metric and the Riemannian metric is always positive and upper bounded a term that is inversely proportional to the number of dimensions D (Proposition 3.3). Similarly, we will deduce the same for the length, the energy and the volume (Corollary 3.2). From this last results, we can directly conclude in section B.2.3 that both metrics are equal in high dimensions (Corollary 3.4). A possible interpretation is that in high dimensions the data distribution obtained on those manifolds becomes more and more concentrated around the mean, reducing the variance term to zero. The manifold becoming deterministic, both metrics become equal.

Remark. *Most of the following proofs will be a bit technical, as they rely on the closed form expression of the non-central Nakagami distribution. Once proving the main propositions, obtaining the corollaries is straightforward. While we do not have closed form expression of the indicatrix, we will show that it's a monotoneous function which can upper and lower bounded.*

B.2.1 Bounds on the Finsler metric

Proposition 3.1. We define $\alpha = 2\left(\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}\right)^2$. The Finsler metric: $v \to \mathbb{E}\left[\sqrt{v^\top G_x v}\right]$ is bounded by two Riemannian metric tensors: the covariance tensor $\alpha\Sigma_x$ and the expected metric tensor $\mathbb{E}[G_x]$.

$$\forall (x, v) \in \mathcal{M} \times \mathcal{T}_x M : \ \sqrt{v^\top \alpha \Sigma_x v} \leq F_x(v) \leq \sqrt{v^\top \mathbb{E}[G_x] v}$$

Proof. Let's first recall that the Finsler function can be written as:$F(v) = \sqrt{2}\sqrt{v^\top \Sigma v}\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})} \cdot {}_1F_1(-\frac{1}{2}, \frac{D}{2}, -\frac{1}{2}\omega)$. The confluent hypergeometric function is defined as: ${}_1F_1(a, b, z) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k}\frac{z^k}{k!}$, with $(a)_k$ and $(b)_k$ being the Pochhammer symbols. By definition, we have: $\frac{(a)_k}{(b)_k} = \frac{\Gamma(a+k)}{\Gamma(b+k)}\frac{\Gamma(a)}{\Gamma(b)}$. We can use the Kummer transformation to obtain: ${}_1F_1(a, b, -z) = e^{-z}{}_1F_1(b - a, b, z)$. Replacing $a = -\frac{1}{2}$, $b = \frac{D}{2}$ and $z = \frac{1}{2}\omega$, we finally get:

$$F(v) = \sqrt{2}\sqrt{v^\top \Sigma v} \cdot e^{-z} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{D}{2} + \frac{1}{2} + k)}{\Gamma(\frac{D}{2} + k)} \frac{z^k}{k!}.$$

1) Let's show that: $\forall v \in \mathcal{T}_x M : \ \sqrt{v^\top \alpha \Sigma v} \leq F(v)$, with $\alpha = 2\left(\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}\right)^2$.

The Pochhammer symbole is defined as $(x)_k = x(x + 1) \ldots (x + k - 1) = \frac{\Gamma(x+k)}{\Gamma(x)}$. For $x \in \mathbb{R}_+^*$, we have: $(x)_k \leq (x + \frac{1}{2})_k$. Thus, $\frac{\Gamma(\frac{D}{2}+k)}{\Gamma(\frac{D}{2})} \leq \frac{\Gamma(\frac{D}{2}+\frac{1}{2}+k)}{\Gamma(\frac{D}{2}+\frac{1}{2})}$. The Gamma function being strictly positive on $\mathbb{R}_+$, we obtain:

$$\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})} \leq \frac{\Gamma(\frac{D}{2}+\frac{1}{2}+k)}{\Gamma(\frac{D}{2}+k)}$$
$$\sqrt{2}\sqrt{v^\top \Sigma v}\ \frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})} \cdot e^{-z} \sum_{k=0}^{\infty} \frac{z^k}{k!} \leq \sqrt{2}\sqrt{v^\top \Sigma v} \cdot e^{-z} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{D}{2}+\frac{1}{2}+k)}{\Gamma(\frac{D}{2}+k)} \frac{z^k}{k!}$$
$$\sqrt{2}\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}\sqrt{v^\top \Sigma v} \leq \sqrt{2}\sqrt{v^\top \Sigma v} \cdot e^{-z} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{D}{2}+\frac{1}{2}+k)}{\Gamma(\frac{D}{2}+k)} \frac{z^k}{k!}$$
$$\sqrt{v^\top \alpha \Sigma v} \leq F(v).$$

2) Let's show that: $\forall v \in \mathcal{T}_x M :\ F(v) \leq \sqrt{v^\top \mathbb{E}[G_x] v}$.

Wendel (1948) proved: $\frac{\Gamma(x+y)}{\Gamma(x)} \leq x^y$, for $x > 0$ and $y \in [0,1]$. With $x = \frac{D}{2}+k$, $y = \frac{1}{2}$, we obtained $\frac{\Gamma(\frac{D}{2}+\frac{1}{2}+k)}{\Gamma(\frac{D}{2}+k)} \leq \sqrt{\frac{D}{2}+k}$, which leads to: $F(v) \leq \sqrt{2v^\top \Sigma v} \cdot e^{-z} \sum_{k=0}^{\infty} \sqrt{\frac{D}{2}+k} \frac{z^k}{k!}$.

Furthermore, $\sum_{k=0}^{\infty} e^{-z} \frac{z^k}{k!} = 1$ and the function $x \to \sqrt{\frac{D}{2}+x}$ is concave. Then by Jensen's inequality: $e^{-z} \sum_{k=0}^{\infty} \sqrt{\frac{D}{2}+k} \frac{z^k}{k!} \leq \sqrt{\frac{D}{2}+e^{-z} \sum_{k=0}^{\infty} \frac{z^k}{k!} k}$. Knowing that $\sum_{k=0}^{\infty} \frac{z^k}{k!} = ze^z$, we have: $e^{-z} \sum_{k=0}^{\infty} \sqrt{\frac{D}{2}+k} \frac{z^k}{k!} \leq \sqrt{\frac{D}{2}+z}$.
And with $z = \frac{\Omega}{2}$, we obtain: $F(v) \leq \sqrt{v^\top \Sigma (D+\Omega) v}$.

From (Kent & Muirhead, 1984, p. 442), the expectation of a non-central Wishart distribution ($G_x \sim \mathcal{W}_q(D, \Sigma, \Omega)$) is: $\mathbb{E}[G_x] = D\Sigma + \Sigma\Omega$. This finally leads to:

$$F(v) \leq \sqrt{v^\top \mathbb{E}[M] v}.$$

□

Remark. *As a side note, the second part of the inequality $F(v) \leq \sqrt{v^\top \mathbb{E}[G_x] v}$ can be obtained using directly Proposition 3.2.*

Corollary 3.1. The length, the energy and the Busemann-Hausdorff volume of the Finsler metric are bounded respectively by the Riemannian length, energy and volume of the covariance tensor $\alpha\Sigma$ (noted $L_\Sigma, E_\Sigma, V_\Sigma$) and the expected metric $\mathbb{E}[G]$ (noted L_G, E_G, V_G):

$$\forall x \in \mathcal{M},\ L_\Sigma(x) \leq \mathcal{L}(x) \leq L_G(x)$$
$$E_\Sigma(x) \leq \mathcal{E}(x) \leq E_G(x)$$
$$V_\Sigma(x) \leq \mathcal{V}(x) \leq V_G(x)$$

Proof. From Proposition 3.1, we have $\forall (x,v) \in \mathcal{M} \times \mathcal{T}_x M :\ \sqrt{h(v)} \leq F(v) \leq \sqrt{g(v)}$, with $h : v \to v^\top \alpha \Sigma_x v$ and $g : v \to v^\top \mathbb{E}[G_x] v$ Riemannian metrics. We also define the parametric curve: $\forall t \in \mathbb{R}, \gamma(t) = x$ and $\dot{\gamma}(t) = v$.

1) Let's show that $L_\Sigma(x) \leq \mathcal{L}(x) \leq L_G(x)$. Because of the monotonicity of the Lebesgue integrals, we directly have: $\int \sqrt{h(\dot{\gamma}(t))} dt \leq \int F(\dot{\gamma}(t)) dt \leq \int \sqrt{g(\dot{\gamma}(t))} dt$.

2) Let's show that $E_\Sigma(x) \leq \mathcal{E}(x) \leq E_G(x)$. Since all the functions are positive, we can raise them to the power two, and again, with the monotonicity of the Lebesgue integrals, we have: $\int h(\dot{\gamma}(t)) dt \leq \int F^2(\gamma(t), \dot{\gamma}(t)) dt \leq \int \sqrt{g(\dot{\gamma}(t))} dt$.

3) Let's show that $V_\Sigma(x) \leq \mathcal{V}(x) \leq V_G(x)$. We write the vectors $v \in \mathcal{T}_x M$ in hyperspherical coordinates: $v = re$, with $r = \|v\|$ the radial distance and e the angular coordinates. With $v = re$, we have: $r \cdot \sqrt{h(e)} \leq r \cdot F(e) \leq r \cdot g(e) \iff \sqrt{h(e)}^{-1} \geq F(e)^{-1} \geq \sqrt{g(e)}^{-1}$.

We want to identify an inequality between the indicatrices, noted $\text{vol}(I_h)$, $\text{vol}(I_g)$, $\text{vol}(I_F)$, formed by the functions h, g and F. Let's define: $r_g\sqrt{h(e)} = r_h\sqrt{g(e)} = r_F F(e) = 1$. For every angular coordinate e, we obtain: $r_h \geq r_F \geq r_g$. Intuitively, this means that the finsler indicatrix will always be bounded by the indicatrices formed by h and g. The Busemann-Hausdorf volume of a function f is defined as: $\sigma_B(f) = \text{vol}(\mathbb{B}^n(1))/\text{vol}(I_f)$, with $\text{vol}(\mathbb{B}^n(1))$ the volume of the unit ball and $\text{vol}(I_f)$ the volume of the indicatrix formed by f. The previous inequality and the definition of the Busemann-Hausdorff volume implies that: $\text{vol}(I_h) \geq \text{vol}(I_F) \geq \text{vol}(I_g) \Rightarrow \sigma_B(h) \leq \sigma_B(F) \leq \sigma_B(g)$. The functions g and h being Riemannian, we have: $\sigma_B(h) = \sqrt{det(\alpha\Sigma_x)}$ and $\sigma_B(g) = \sqrt{det(\mathbb{E}[G_x])}$, which concludes the proof.

□

B.2.2 Relative bounds between the Finsler and the Riemannian metric

Proposition 3.2. The relative difference between the Finsler metric: $F_x : v \to \mathbb{E}[\sqrt{v^\top G_x v}]$ and the Riemmanian metric $g : (v, v) \to v^\top \mathbb{E}[G_x]v$ is:

$$0 \leq \frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} \leq \frac{\text{Var}[v^\top G_x v]}{2\mathbb{E}[v^\top G_x v]^2}.$$

Proof. We will directly use a sharpen version of Jensen's inequality obtained by Liao & Berg (2019): Let X be a one-dimensional random variable with mean μ and $P(X \in (a, b)) = 1$, where $-\infty \leq a \leq b \leq +\infty$. Let ϕ a twice derivable function on (a, b). We further define: $h(x, \mu) = \frac{\phi(x)-\phi(\mu)}{(x-\mu)^2} - \frac{\phi'(\mu)}{x-\mu}$. Then:

$$\inf_{x\in(a,b)} \{h(x, \mu)\}\text{Var}[X] \leq \mathbb{E}[\phi(x)] - \phi(\mathbb{E}[x]) \leq \sup_{x\in(a,b)} \{h(x, \mu)\}\text{Var}[X].$$

In our case, we will chose $\phi : z \to \sqrt{z}$ with z a one-dimensional random variable defined as $z = v^\top G_x v$. $a = 0$, $b = +\infty$ and $\mu = \mathbb{E}[z]$. $h(z, \mu) = (\sqrt{z} - \sqrt{\mu})(z - \mu)^{-2} - (2(z - \mu)\sqrt{\mu})^{-1}$. Because its first derivative ϕ' is convex, the function $x \to h(x, \mu)$ is monotonically increasing. Thus:

$$\inf_{z\in(0,+\infty)} \{h(x, \mu)\} = \lim_{z\to 0} = -\frac{\sqrt{\mu}}{2\mu^2} \quad \text{and} \quad \sup_{z\in(0,+\infty)} \{h(x, \mu)\} = \lim_{z\to+\infty} = 0.$$

It finally gives:

$$-\frac{\sqrt{\mu}}{2\mu^2}\text{Var}[z] \leq \mathbb{E}[\sqrt{z}] - \sqrt{\mathbb{E}[z]} \leq 0.$$

Replacing $F(v) = \mathbb{E}[\sqrt{z}]$ and $\sqrt{g(v)} = \sqrt{\mathbb{E}[z]} = \sqrt{\mu}$ concludes the proof. □

Lemma B.2.1. *Let $z \sim \mathcal{W}_1(D, \sigma, \Omega)$ following a one-dimensional non-central Wishart distribution. Then:*

$$\frac{Var[z]}{2\mathbb{E}[z]^2} = \frac{1}{D + \Omega} + \frac{\Omega}{(D + \Omega)^2}$$

Proof. (Kent & Muirhead, 1984, Theorem 10.3.7.) states that if $z \sim \mathcal{W}_1(D, \sigma, \omega)$ then $\mathbb{E}[z^k] = \sigma^k 2^k \frac{\Gamma(\frac{D}{2}+k)}{\Gamma(\frac{D}{2})} {}_1F_1(-k, \frac{D}{2}, -\frac{1}{2}\Omega)$. In particular, for $k = 1$ and $k = 2$, we have ${}_1F_1(-1, b, c) = 1 - \frac{c}{b}$ and ${}_1F_1(-2, b, c) = 1 - \frac{2c}{b} + \frac{c^2}{b(b+1)}$. We also have $\frac{\Gamma(\frac{D}{2}+1)}{\Gamma(\frac{D}{2})} = \frac{D}{2}$ and $\frac{\Gamma(\frac{D}{2}+2)}{\Gamma(\frac{D}{2})} = \frac{D}{2}\left(\frac{D}{2} + 1\right)$, which leads to: $\mathbb{E}[z] = \sigma(D + \Omega)$ and $\mathbb{E}[z^2] = \sigma^2(2\omega + 2(D + \omega) + (D + \omega)^2)$. Finally, we conclude:

$$\frac{\text{Var}[z]}{\mathbb{E}[z]^2} = \frac{\mathbb{E}[z^2]}{\mathbb{E}[z]^2} - 1 = \frac{2\omega}{(D + \omega)^2} + \frac{2}{D + \omega}.$$

□

Proposition 3.3. Let f be a Gaussian process. We note $\omega = (v^\top \Sigma_x v)^{-1}(v^\top \mathbb{E}[J]^\top \mathbb{E}[J]v)$, with J_f the jacobian of f, and Σ the covariance matrix of J.
The relative ratio between the Finsler metric: $F : (x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$ and the Riemmanian metric $g : (v, v) \to v^\top \mathbb{E}[G_x]v$ is:

$$0 \leq \frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} \leq \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2}.$$

Proof. The result is directly obtained using Proposition 3.2 and Lemma B.2.1. □

Corollary 3.2. The relative ratio between the length, the energy and the volume of the Finsler metric: $(x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$ (noted $\mathcal{L}, \mathcal{E}, \mathcal{V}$) and the Riemannian metric with the metric tensor $\mathbb{E}[G_x]$ (noted L_G, E_G, V_G) is:

$$\begin{aligned}
0 \leq & \frac{L_G(x) - \mathcal{L}(x)}{L_G(x)} \leq \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\} \\
0 \leq & \frac{E_G(x) - \mathcal{E}(x)}{E_G(x)} \leq \max_{v \in \mathcal{T}_x M} \left\{ \frac{2}{D+\omega} + \frac{1+2\omega}{(D+\omega)^2} + \frac{2\omega}{(D+\omega)^3} + \frac{\omega^2}{(D+\omega)^4} \right\} \\
0 \leq & \frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} \leq 1 - \left(1 - \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\} \right)^q
\end{aligned}$$

Proof. Let's call $M = \max_{v \in \mathcal{T}_x M} \{ \frac{\omega}{(D+\omega)^2} + \frac{1}{D+\omega} \}$ We have:

$$0 \leq \sqrt{g(v)} - F(v) \leq \sqrt{g(v)} M, \quad \text{with} \quad M = \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\}$$

1) By the monocity of the Lesbesgue integral, we have: $0 \leq \int \sqrt{g(\gamma(\dot{t}))} - F(\gamma(\dot{t}))dt \leq M \int \sqrt{g(\gamma(\dot{t}))}dt$, which leads to: $0 \leq L_G(x) - \mathcal{L}(x) \leq M L_G(x)$.

2) Since all the functions are positive: $0 \leq F(v) \leq \sqrt{g(v)} \leq \sqrt{g(v)}M + F(v)$ leads to: $F^2(x, v) \leq g(v) \leq M^2 g(v) + 2MF(v)\sqrt{g(v)} + F^2(x, v)$, and we also know from Proposition 3.1 that $F(v) \leq \sqrt{g(v)}$, thus: $F^2(x, v) \leq g(v) \leq (M^2 + 2M)g(v) + F^2(x, v)$, and finally: $0 \leq g(v) - F^2(x, v) \leq (M^2 + 2M)g(v)$. Again, by continuity of the Lebesgue integral, we directly obtain: $0 \leq E_G(x) - \mathcal{E}(x) \leq (M^2 + 2M)E_G(x)$.

3) In order to compare the volume between the Finsler and the Riemannian metric, we need to compare the volume of their indicatrices, noted: $\text{vol}(I_g)$ and $\text{vol}(I_F)$ respectively. We write the vectors $v \in \mathcal{T}_x M$ in hypershperical coordinates, with $v = re$, $r = \|v\|$ the radial distance and e the angular coordinates. The volume of the indicatrices obtained in dimension q (dimension of the latent space) can be written as: $d^q V = r^{q-1} dr d\Phi$, with Φ defining the different angles. We will note r_F and r_g the radial distances of the Finsler and Riemann metrics such that: $F(v) = r_F F(e) = 1$ and $\sqrt{g(v)} = r_g \sqrt{g(e)} = 1$ obtained for a specific angle e.

$$\text{vol}(I_F) - \text{vol}(I_g) = \int_\Phi \left(\int_0^{r_f} r^{q-1} dr - \int_0^{r_g} r^{q-1} dr \right) d\Phi = \int_\Phi \frac{r_f^q}{q} \left(1 - \left(\frac{r_g}{r_f} \right)^q \right) d\Phi \leq \int_\Phi \frac{r_f^q}{q} d\Phi \cdot \left(1 - \left(\frac{r_g}{r_f} \right)^q \right),$$

and by definition: $\text{vol}(I_F) = \int_\Phi (r_f^q / q) d\Phi$. Furthermore, for a specific angle e, we have: $r_g / r_F = \sqrt{g(e)}/F(e) \geq 1 - M$, from Proposition 3.3. We have:

$$0 \leq \frac{\mathrm{vol}(I_F) - \mathrm{vol}(I_g)}{\mathrm{vol}(I_F)} \leq \cdot 1 - \left(\frac{r_g}{r_f}\right)^q \leq 1 - (1-M)^q \,,$$

and by the definition of the Busemann Hausdorff volume: $\frac{\mathcal{V}(x) - V_G(x)}{\mathcal{V}(x)} = \frac{\mathrm{vol}(I_F) - \mathrm{vol}(I_g)}{\mathrm{vol}(I_F)}$, we conclude the proof.

□

B.2.3 Implications in High Dimensions

In this section, we want to show that the difference between the Finsler norm and the Riemannian induced norm, as well as their respective functionals, tend to zero at a rate of $\mathcal{O}(\frac{1}{D})$. We need to be sure that ω doesn't grow faster than D, in other terms: $\omega = \mathcal{O}(D)$. This can be obtained if we assume that every element of the expectation of Jacobian is upper bounded ($\exists m \in \mathbb{R}_+^*, \forall i, j \ \mathbb{E}[J_{ij}] \leq m$). This happens in at least two cases: (1) $\mathbb{E}[f]$ is somehow Lipschitz continuous; or (2) if f is a Gaussian Process and its covariance is upper bounded. The latter case happens when the process is defined over a bounded domain.

Lemma B.2.2. *Our Finsler metric $v \to \mathbb{E}[\sqrt{v^\top G_x v}]$ is defined with $v^\top G_x v \sim \mathcal{W}_1(D, v^\top \Sigma_x v, \omega)$, and $\omega = (v^\top \Sigma_x v)^{-1}(v^\top \mathbb{E}[J]^\top \mathbb{E}[J] v)$.*

If the Finsler manifold is bounded, then: $\omega \leq DM$, with $M \in \mathbb{R}_+$.

Proof. By definition, Σ does not depend on D. We assume the manifold is bounded. f being a Gaussian Process defined on this manifold, we deduce that every element of the expected Jacobian is upper bounded: $\mathbb{E}[J]_{ij} \leq m$, with $m \in \mathbb{R}_+^*$. We call $\sigma = v^\top \Sigma_x v \in \mathbb{R}_+^*$.

We have:

$$\omega = \sigma^{-1} \sum_{k=1}^{D} \sum_{i=1}^{q} \sum_{j=1}^{q} v_i \mathbb{E}[J]_{ki} \mathbb{E}[J]_{kj} v_j \leq \sigma^{-1} \sum_{k=1}^{D} m^2 \|v\|^2 \leq DM,$$

with $M = \sigma^{-1} m^2 \|v\|^2 \in \mathbb{R}_+^*$, and M does not depend on D. □

Corollary 3.3. In high dimensions, we have:

$$\frac{L_G(x) - \mathcal{L}(x)}{L_G(x)} = \mathcal{O}\left(\frac{1}{D}\right)$$
$$\frac{E_G(x) - \mathcal{E}(x)}{E_G(x)} = \mathcal{O}\left(\frac{1}{D}\right)$$
$$\frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} = \mathcal{O}\left(\frac{q}{D}\right)$$

And, when D converges toward infinity: $L_G \underset{+\infty}{\sim} \mathcal{L}$, $E_G \underset{+\infty}{\sim} \mathcal{E}$ and $V_G \underset{+\infty}{\sim} \mathcal{V}$.

Proof. From Corollary 3.2, we directly obtained the results in high dimensions.

We assume that our latent space is bounded, then by B.2.2, we have: $0 \leq \omega \leq MD$, with $M \in \mathbb{R}+$.

For the length, we have:

$$\begin{aligned} \frac{L_G(x) - \mathcal{L}(x)}{L_G(x)} &\leq \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\} \\ &\leq \frac{1+M}{D} \end{aligned}$$

For the energy functional, we have:

$$\begin{aligned}
\frac{E_G(x) - \mathcal{E}(x)}{E_G(x)} &\leq \max_{v \in \mathcal{T}_x M} \left\{ \frac{2}{D+\omega} + \frac{1+2\omega}{(D+\omega)^2} + \frac{2\omega}{(D+\omega)^3} + \frac{\omega^2}{(D+\omega)^4} \right\} \\
&\leq \frac{2+2M}{D} + \frac{1+2M+M^2}{D^2} \\
\limsup_{D\to\infty} D \times \frac{E_G(x) - \mathcal{E}(x)}{E_G(x)} &\leq \limsup_{D\to\infty} 2(1+M) + \frac{M^2+2M+1}{D^2} \to 2(1+M)
\end{aligned}$$

For the volume, we have:

$$\begin{aligned}
\frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} &\leq 1 - \left(1 - \max_{v \in \mathcal{T}_x M} \left\{ \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \right\} \right)^q \\
&\leq 1 - \left(1 - \frac{1+M}{D}\right)^q
\end{aligned}$$

Using Taylor series expansion, when $x \sim 0$, we have: $1-(1-x)^q = qx + o(x^2)$. Let's call $\varepsilon \ll 1$, and rewrite the Taylor series:

$$\begin{aligned}
\frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} &\leq q\frac{1+M}{D} + \varepsilon q \frac{1+M}{D} \\
\limsup_{D\to\infty} \frac{D}{q} \times \frac{V_G(x) - \mathcal{V}(x)}{V_G(x)} &\leq (1+M)(1+\varepsilon)
\end{aligned}$$

The difference between the functionals can converge to zero if they are similar in high dimensions, or if they all diverge to infinity. This latter case does not happen as we assume the latent manifold being bounded, and so the metrics are then finite, which concludes the proof.

□

Corollary 3.4. In high dimensions, the relative ratio between the Finsler metric: $F : (x, v) \to \mathbb{E}[\sqrt{v^\top G_x v}]$ and the Riemmanian metric $g : (v, v) \to v^\top \mathbb{E}[G_x] v$ is:

$$\frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} = \mathcal{O}\left(\frac{1}{D}\right)$$

And, when D converges toward infinity: $\forall v \in \mathcal{T}_x M, \sqrt{g(v,v)} \underset{+\infty}{\sim} F_x(v)$.

Proof. Similar to the 3.3, assuming that our latent space is bounded, from B.2.2, we have $0 \leq \omega \leq MD$. From 3.3, we deduce:

$$\begin{aligned}
0 \leq \frac{\sqrt{g(v,v)} - F_x(v)}{\sqrt{g(v,v)}} &\leq \frac{1}{D+\omega} + \frac{\omega}{(D+\omega)^2} \\
&\leq \frac{1+M}{D}
\end{aligned}$$

In a bounded manifold, the metric are finite. We can deduce that they converge to each other in high dimensions. □

C Experiments

C.1 Datasets

C.1.1 Font data

The dataset represents 46 different font for each letter (upper and lower case) whose contour is parametrised by a spline (or two splines, depending on the letter used) obtained from at least 500 points Campbell & Kautz (2014).

In our case, we choose to learn the manifold of the letter **f**. The dataset is composed of 46 different fonts, each letter being drawn by 1024 points. We reduce this number from 1024 to 256 by sampling one point every 4. The dimension of the observational space is then 256.

C.1.2 qPCR

The qPCR data, gathered from Guo et al. (2010), was used to illustrate the training of a GPLVM in Pyro Pyro (2022) and is available at the Open Data Science repository Ahmed et al. (2019). It consists of 437 single-cell qPCR data for which the expression of 48 genes has been measured during 10 different cell stages. We then have 437 data points, 48 observations, and 10 classes. Before training the GP-LVM, the data is grouped by the capture time, as illustrated in the Pyro documentation.

C.1.3 Pinwheel on a sphere

A pinwheel in 2-dimension is created and then projected onto a sphere using a stereographic projection method. The final dataset is composed of 1000 points with their coordinates in 3-dimensions.

C.2 GP-LVM training

We learn our two-dimensional latent space by training a GP-LVM Lawrence (2003) with Pyro Bingham et al. (2019). The Gaussian Process used is a Sparse GP, defined with a kernel (RBF, or Matern) composed of a one-dimensional lengthscale and variance. The parameters are learnt with the Adam optimiser Kingma & Ba (2014). The number of steps and the initialisation of the latent space vary with the dataset.

datasets	pinwheel	font data	qPCR
Number of data points	500	46	437
Number of observations	3	256	48
initialisation	PCA	PCA	custom
kernel	RBF	Matern52	Matern52
steps	17000	5000	5000
learning rate	1e-3	1e-4	1e-4
lengthscale	0.24	0.88	0.15
variance	0.95	0.30	0.75
noise	1e-4	1e-3	1e-3

Table 2: Description of the datasets trained with a GP-LVM.

C.3 Computing indicatrices

An indicatrix of a function g at a point x is defined such that: $v \in \mathcal{T}_x M | g_x(v) < 1$. In other terms, the indicatrix is the representation of a unit ball in our latent space. If we use an euclidean metric, our indicatrix in our 2-dimensional latent space would be a unit ball, as we need to solve: $v \in \mathcal{T}_x M$, $\|v\| < 1$. For a Riemannian metric, our indicatrix is necessarily an ellipse, whose semi axis are the square-roots of the eigenvalues of the metric tensor G:: $v \in \mathcal{T}_x M$, $v^\top G v < 1$. For our Finsler metric, we don't have an analytical solution, and so it's difficult to predict the shape of the convex polygon.

In this paper, the indicatrices are drawn the following way: for a single point in our latent space, we compute the value of $v^\top G v$ and $F(x, v)$ for v varying over the space. We then extract the contour when $v^\top G v$ and $F(x, v)$ are equal to 1. Computing the area of the indicatrices will be used in the section C.4 to compute the volume measures.

C.4 Computing the volume forms

For the figures used in this paper, by default, the background of the latent space represents the volume measure of the expected Riemannian metric ($V_G = \sqrt{\mathbb{E}[G]}$) on a logarithm scale. In figure 4, the volume measure of the Finsler metric is also computed.

C.4.1 Finsler metric

To compute the volume measure of our Finsler metric, we choose the Busemann-Hausdorff definition, which is the ratio of a unit ball over the volume of its indicatrix: $\mathcal{V} = \text{vol}(\mathbb{B}^n(1))/\text{vol}(\{v \in \mathcal{T}_x M | F(x, v) < 1\})$. While our Finsler function has an analytical form, its expression doesn't allow to directly solve the equation: $v \in \mathcal{T}_x M, F(x, v) < 1$. Instead we approximate its indicatrix as describe in C.3, using a contour plot and extracting the paths vertices. We can then compute the area of the obtained polygon, and divide with the volume of a unit ball: $\text{vol}(\mathbb{B}^2(1)) = \pi$.

The volume measure can then be computed for each point over a grid (32 x 32, in figure 4), and we interpolate all the other points. Note that this method can only be used when our latent space is of dimension 2.

C.4.2 Expected Riemannian metric

There is two ways to compute the volume measure of the expected Riemannian metric. One way is to directly use the metric tensor: $V_G = \sqrt{\mathbb{E}[G]}$. Another one is to remember that any Riemannian metric is a Finsler metric, and thus, the Busemann-Hausdorff definition also applied for our metric: $V_G = \text{vol}(\mathbb{B}^n(1))/\text{vol}(\{v \in \mathcal{T}_x M | v^\top \mathbb{E}[G] v < 1\})$. Solving $v^\top \mathbb{E}[G] v < 1$ for $v \in \mathcal{T}_x M$ is equivalent to solving the area of an ellipse.

For the first method, we can either sample multiple times the metric, which is computationally expensive, or use the fact that our metric tensor is a non-central Wishart matrix: $G = J^\top J \sim \mathcal{W}_q(D, \Sigma, \Sigma^{-1}\mathbb{E}[J]^\top \mathbb{E}[J])$, with Σ the covariance of the Jacobian J and D the dimension of the observational space. In this case, its expectation is: $\mathbb{E}[G] = \mathbb{E}[J]^\top \mathbb{E}[J] + D\Sigma$. We can access the derivatives of the function f (detailed in section D.2), and compute both quantities $\mathbb{E}[J]$ and Σ needed to estimate the expected metric and its determinant.

For the second method, we can compute the area of the ellipse in the same way we compute the Finsler volume measure.

C.5 Experiments when increasing the number of dimensions

In Figure 5, we computed the volume ratio and draw indicatrices while varying the number of dimensions, to illustrate that both our Finsler metric and the expected Riemannian metric seem to converge when D increases.

The main issue with this experiment is to vary only one factor, the number of dimensions D, while keeping the other factors unchanged. This is difficult for two reasons: 1) in real toy examples as seen in Figure 6, even with a very low dimensional observational such as in the pinwheel on the sphere, both metrics are already very similar to each other. It would be difficult to illustrate a convergence while increasing the number of dimensions. 2) the function f needs to be learnt again each time we increase the number of dimensions of the observational space, and the parameters of the Gaussian Process will change too.

Instead, we try to illustrate our results by computing empirically the stochastic metric tensor $G = J^\top J$, using its Jacobian $J \sim \mathcal{N}(\mathbb{E}[J], \Sigma)$, a $D \times q$ matrix. The number of dimensions is modified by simply truncatenating the Jacobian J. In Figure 5, the volume ratio is computed for 12 Jacobians obtained with different random parameters $\mathbb{E}[J]$ and Σ. The Finsler and Riemannian indicatrices (lower right) are drawn for only one Jacobian selected randomly.

D Computations

D.1 Computing geodesics with Stochman and minimising the curve energy functionals

An essential task is to compute shortest paths, or geodesics, between data points in the latent space. Those shortest paths can be obtained in two ways: either by solving a corresponding system of ODEs, or by minimising the curve energy in the latent space. The former being computationally expensive, we favour the second approach which consists in optimising a parameterised spline on the manifold. This method is already implemented in Stochman Detlefsen et al. (2021), where we can easily optimise splines by redefining the curve energy function of a manifold class.

We need two curvge energy functional: one for the expected Riemannian metric and one for the Finsler metric.

D.1.1 Curve energy for the Riemannian metric

We know that the stochastic metric tensor G_t defined on a point t follows a non-central Whishart distribution. Thus, we can compute its expectation $\mathbb{E}[G_t]$ knowing the Jacobian covariance and expectation: $\mathbb{E}[J_t]$ and Σ. The next Section D.2 explains how to compute those quantities.

Assuming the spline is discretized into N points, we can compute the curve energy with:

$$E_G(\gamma(t)) = \int_0^1 \dot{\gamma}(t)^\top \mathbb{E}[G_t]\dot{\gamma}(t)\, dt \approx \sum_{i=1}^{N} \dot{\gamma}_i^\top \left(\mathbb{E}[J_i]^T\mathbb{E}[J_i] + D\Sigma_i\right)\dot{\gamma}_i.$$

D.1.2 Curve energy for the Finsler metric

In order to compute of the curve energy $\mathcal{E}(\gamma)$, we must first derive the expectation $\mathbb{E}[J_t]$ and covariance Σ of the Jacobian of f, which should follow a normal distribution: $J_i = \partial f_i \sim \mathcal{N}\left(\mathbb{E}[J], \Sigma\right)$. We assume the points ∂f_i are independent samples with the same variance drawn from a normal distribution. We can then compute the Finsler metric which follows a non-central Nakagami distribution (See Proposition 2.2):

$$\mathcal{E}(\gamma(t)) = \int_0^1 F(t, \dot{\gamma}(t))^2\, dt \approx \sum_{i=1}^{N} 2\dot{\gamma}_i^\top \Sigma_i \dot{\gamma}_i \left(\frac{\Gamma(\frac{D}{2}+\frac{1}{2})}{\Gamma(\frac{D}{2})}\right)^2 {}_1F_1\left(-\frac{1}{2}, \frac{D}{2}, -\frac{\omega_i}{2}\right)^2,$$

with ${}_1F_1$ the confluent hypergeometric function of the first kind and $\omega_i = (\dot{\gamma}_i^\top \Sigma_i \dot{\gamma}_i)^{-1}(\dot{\gamma}_i^\top \Omega_i \dot{\gamma}_i)$ and $\Omega_i = \Sigma_i^{-1}\mathbb{E}[J_i]^\top \mathbb{E}[J_i]$.

This function has been implemented in Pytorch using the known gradients for the hypergeometric function: $\frac{\partial}{\partial x}{}_1F_1(a, b, x) = \frac{a}{b}{}_1F_1(a+1, b+1, x)$.

D.2 Accessing the posterior derivatives

We assume that the probabilitic mapping f from the latent variables X to the observational variables Y follows a normal distribution. We would like to obtain the posterior kernel Σ_* and expectation μ_* such that $p(\partial_t f | Y, X) \sim \mathcal{N}\left(\mu_*, \Sigma_*\right)$.

We make the hypothesis the observed variables are modelled with a gaussian noise ϵ whose variance is the same in every dimension. In particular, for the nth latent (x) and observed (y) variable in the jth dimension: $y_{n,j} = f_j(x_{n,:}) + \epsilon_n$. Thus, the output variables have the same variance, and the posterior kernel Σ_* is then isotropic with respect to the output dimensions: $\Sigma_* = \sigma_*^2 \cdot I_D$.

There are two ways of obtaining the posterior variance and expectation:

- We use the gaussian processes to predict the derivative $(\partial_c f)$ of the mapping function f, and we multiply the obtained posterior kernel by the curve derivative $(\partial_t c)$, following the chain rule: $\frac{df(c(t))}{dt} = \frac{df}{dc} \cdot \frac{dc}{dt}$ (Section: D.2.1)

- We discretize the derivative of the mapping function as the difference of this function evaluated at two close points. We use a linear operation to obtain the posterior variance and expectation: $\partial_t f(c(t)) \sim f(c(t_{i+1})) - f(c(t_i))$. (Section: D.2.2)

D.2.1 Closed-form expressions

We assume that f is a Gaussian process. Hence, because the differentiation is a linear operation, the derivative of a Gaussian Process is also a Gaussian Process Rasmussen & Williams (2005).

The data $Y \in \mathcal{R}^{N \times D}$ follows a normal distribution, so we can infer the partial derivative of one data point $(J^T)_{ji} = \frac{\partial y_i}{\partial x_j}$, with $i = 1 \dots D$ and $j = 1 \dots d$. We have:

$$\begin{bmatrix} Y \\ (J)^\top \end{bmatrix} = \prod_{i=1}^{D} \mathcal{N}\left(\begin{bmatrix} \mu_y \\ \mu_{\partial y} \end{bmatrix}, \begin{bmatrix} K(x,x) & \partial K(x,x_*) \\ \partial K(x_*,x) & \partial^2 K(x_*,x_*) \end{bmatrix} \right).$$

and $J^\top$ can be predicted:

$$p(J^\top | Y, X) = \prod_{i=1}^{D} \mathcal{N}\left(\mu_*, \Sigma_*\right),$$

with:

$$\mu_* = \partial K(x_*,x) \cdot K(x,x)^{-1} \cdot (y - \mu_y) + \mu_{\partial y}$$
$$\Sigma_* = \partial^2 K(x_*,x_*) - \partial K(x_*,x) \cdot K(x,x)^{-1} \cdot \partial K(x,x_*).$$

Finally, $\partial_t f$ is obtained:

$$p(\partial_t f(c(t)) | f(x), x) = \prod_{i=1}^{D} \mathcal{N}\left(\dot{c}\mu_*, \dot{c}^\top \Sigma_* \dot{c} \cdot I_D\right).$$

D.2.2 Discretization

One can notice that: $\partial_t f(c(t)) \sim f(c(t_{i+1})) - f(c(t_i))$. We know that $f(c(t_{i+1}))$ and $f(c(t_i))$ both follows a normal distribution.

$$\begin{bmatrix} f(c(t_i)) \\ f(c(t_{i+1})) \end{bmatrix} = \prod_{j=1}^{D} \mathcal{N}\left(\begin{bmatrix} \mu_i \\ \mu_{i+1} \end{bmatrix}, \begin{bmatrix} \sigma_{ii}^2 & \sigma_{i,i+1}^2 \\ \sigma_{i+1,i}^2 & \sigma_{i+1,i+1}^2 \end{bmatrix} \right).$$

If $Y = AX$ affine transformation of a multivariate Gaussian $X \sim \mathcal{N}(\mu, \sigma^2)$, then Y is also a multivariate Gaussian with: $Y \sim \mathcal{N}(A\mu, A^T \sigma^2 A)$. In our case, we choose $A^T = [-1, 1]$. We have:

$$f(c(t_{i+1})) - f(c(t_i)) \sim \mathcal{N}(\mu_*, \sigma_*^2 \cdot I_D),$$

with:

$$\mu_* = \mu_{i+1} - \mu i$$
$$\sigma_*^2 = \sigma_{ii}^2 + \sigma_{i+1,i+1}^2 - 2\sigma_{i,i+1}^2$$

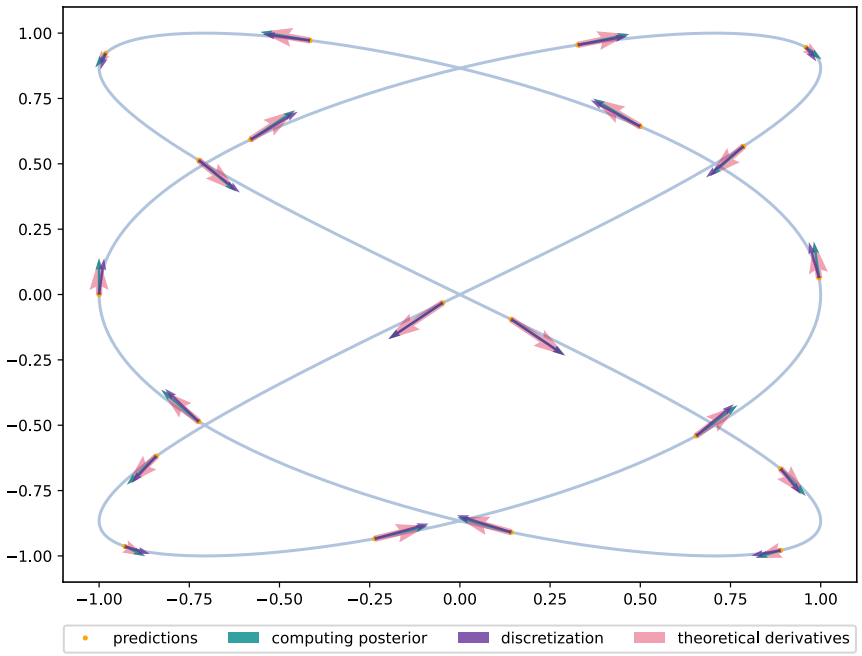

Figure 12: Illustration of the derivatives obtained with a trained GP on a simple parametrised function: both methods give the correct derivatives if enough points are sampled.