Anisotropically weighted and nonholonomically constrained evolutions on manifolds

Sommer, Stefan Horst

Published in:
Entropy

DOI:
10.3390/e18120425

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Anisotropically Weighted and Nonholonomically Constrained Evolutions on Manifolds †

Stefan Sommer

Department of Computer Science, University of Copenhagen, DK-2100 Copenhagen E, Denmark; sommer@di.ku.dk

† This paper is an extended version of our paper published in the 2nd Conference on Geometric Science of Information, Paris, France, 28–30 October 2015.

Academic Editors: Frédéric Barbaresco and Frank Nielsen

Received: 1 September 2016; Accepted: 23 November 2016; Published: 26 November 2016

Abstract: We present evolution equations for a family of paths that results from anisotropically weighting curve energies in non-linear statistics of manifold valued data. This situation arises when performing inference on data that have non-trivial covariance and are anisotropic distributed. The family can be interpreted as most probable paths for a driving semi-martingale that through stochastic development is mapped to the manifold. We discuss how the paths are projections of geodesics for a sub-Riemannian metric on the frame bundle of the manifold, and how the curvature of the underlying connection appears in the sub-Riemannian Hamilton–Jacobi equations. Evolution equations for both metric and cometric formulations of the sub-Riemannian metric are derived. We furthermore show how rank-deficient metrics can be mixed with an underlying Riemannian metric, and we relate the paths to geodesics and polynomials in Riemannian geometry. Examples from the family of paths are visualized on embedded surfaces, and we explore computational representations on finite dimensional landmark manifolds with geometry induced from right-invariant metrics on diffeomorphism groups.

Keywords: sub-Riemannian geometry; geodesics; most probable paths; stochastic development; non-linear data analysis; statistics

1. Introduction

When manifold valued data have non-trivial covariance (i.e., when anisotropy asserts higher variance in some directions than others), non-zero curvature necessitates special care when generalizing Euclidean space normal distributions to manifold valued distributions: in the Euclidean situation, normal distributions can be seen as transition distributions of diffusion processes, but on the manifold, holonomy makes transport of covariance path-dependent in the presence of curvature, preventing a global notion of a spatially constant covariance matrix. To handle this, in the diffusion principal component analysis (PCA) framework [1], and with the class of anisotropic normal distributions on manifolds defined in [2, 3], data on non-linear manifolds are modelled as being distributed according to transition distributions of anisotropic diffusion processes that are mapped from Euclidean space to the manifold by stochastic development (see [4]). The construction is connected to a non-bracket-generating sub-Riemannian metric on the bundle of linear frames of the manifold, the frame bundle, and the requirement that covariance stays covariantly constant gives a nonholonomically constrained system.

Velocity vectors and geodesic distances are conventionally used for estimation and statistics in Riemannian manifolds; for example, for estimation of the Frechét mean [5], for Principal Geodesic Analysis [6], and for tangent space statistics [7]. In contrast to this, anisotropy as modelled with anisotropic normal distributions makes a distance for a sub-Riemannian metric the natural vehicle for
estimation and statistics. This metric naturally accounts for anisotropy in a similar way as the precision matrix weights the inner product in the negative log-likelihood of a Euclidean normal distribution. The connection between the weighted distance and statistics of manifold valued data was presented in [2], and the underlying sub-Riemannian and fiber-bundle geometry, together with properties of the generated densities, was further explored in [3]. The fundamental idea is to perform statistics on manifolds by maximum likelihood (ML) instead of parametric constructions that use, for example, approximating geodesic subspaces; by defining natural families of probability distributions (in this case using diffusion processes), ML parameter estimates give a coherent way to statistically model non-linear data. The anisotropically weighted distance and the resulting family of extremal paths arises in this situation when the diffusion processes have non-isotropic covariance (i.e., when the distribution is not generated from a standard Brownian motion).

In this paper, we focus on the family of most probable paths for the semi-martingales that drives the stochastic development, and in turn the manifold valued anisotropic stochastic processes. Such paths, as exemplified in Figure 1, extremize the anisotropically weighted action functional. We present derivations of evolution equations for the paths from different viewpoints, and we discuss the role of frames as representing either metrics or cometrics. In the derivation, we explicitly see the influence of the connection and its curvature. We then turn to the relation between the sub-Riemannian metric and the Sasaki–Mok metric on the frame bundle, and we develop a construction that allows the sub-Riemannian metric to be defined as a sum of a rank-deficient generator and an underlying Riemannian metric. Finally, we relate the paths to geodesics and polynomials in Riemannian geometry, and we explore computational representations on different manifolds including a specific case: the finite dimensional manifolds arising in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) [8] landmark matching problem. The paper ends with a discussion concerning statistical implications, open questions, and concluding remarks.

**Figure 1.** (a) A most probable path (MPP) for a driving Euclidean Brownian motion on an ellipsoid. The gray ellipsis over the starting point (red dot) indicates the covariance of the anisotropic diffusion. A frame $u_t$ (black/gray vectors) representing the square root covariance is parallel transported along the curve, enabling the anisotropic weighting with the precision matrix in the action functional. With isotropic covariance, normal MPPs are Riemannian geodesics. In general situations, such as the displayed anisotropic case, the family of MPPs is much larger; (b) The corresponding anti-development in $\mathbb{R}^2$ (red line) of the MPP. Compare with the anti-development of a Riemannian geodesic with same initial velocity (blue dotted line). The frames $u_t \in \text{GL}(\mathbb{R}^2, T_x M)$ provide local frame coordinates for each time $t$.

**Background**

Generalizing common statistical tools for performing inference on Euclidean space data to manifold valued data has been the subject of extensive work (e.g., [9]). Perhaps most fundamental is the notion of Frechét or Karcher means [5,10], defined as minimizers of the square Riemannian...
distance. Generalizations of the Euclidean principal component analysis procedure to manifolds are particularly relevant for data exhibiting anisotropy. Approaches include principal geodesic analysis (PGA, [6]), geodesic PCA (GPCA, [11]), principal nested spheres (PNS, [12]), barycentric subspace analysis (BSA, [13]), and horizontal component analysis (HCA, [14]). Common to these constructions are explicit representations of approximating low-dimensional subspaces. The fundamental challenge here is that the notion of Euclidean linear subspace on which PCA relies has no direct analogue in non-linear spaces.

A different approach taken by diffusion PCA (DPCA, [1,2]) and probabilistic PGA [15] is to base the PCA problem on a maximum likelihood fit of normal distributions to data. In Euclidean space, this approach was first introduced with probabilistic PCA [16]. In DPCA, the process of stochastic development [4] is used to define a class of anisotropic distributions that generalizes the family of Euclidean space normal distributions to the manifold context. DPCA is then a simple maximum likelihood fit in this family of distributions mimicking the Euclidean probabilistic PCA. The approach transfers the geometric complexities of defining subspaces common in the approaches listed above to the problem of defining a geometrically natural notion of normal distributions.

In Euclidean space, squared distances $\|x - x_0\|^2$ between observations $x$ and the mean $x_0$ are affinely related to the negative log-likelihood of a normal distribution $N(x_0, \text{Id})$. This makes an ML fit of the mean such as performed in probabilistic PCA equivalent to minimizing squared distances. On a manifold, distances $d_g(x, x_0)^2$ coming from a Riemannian metric $g$ are equivalent to tangent space distances $\|\log_{x_0} x\|^2$ when mapping data from $M$ to $T_{x_0}M$ using the inverse exponential map $\log_{x_0}$. Assuming $\log_{x_0}$ are distributed according to a normal distribution in the linear space $T_{x_0}M$, this restores the equivalence with a maximum likelihood fit. Let $\{e_1, \ldots, e_d\}$ be the standard basis for $\mathbb{R}^d$. If $u : \mathbb{R}^d \to T_{x_0}M$ is a linear invertible map with $ue_1, \ldots, ue_d$ orthonormal with respect to $g$, the normal distribution in $T_{x_0}M$ can be defined as $uN(0, \text{Id})$ (see Figure 2).

![Figure 2](image_url)

Figure 2. (a) Normal distributions $uN(0, \text{Id})$ in the tangent space $T_{x_0}M$ with covariance $uu^T$ (blue ellipses) can be mapped to the manifold by applying the exponential map $\text{Exp}_{x_0}$ to sampled vectors $v \in T_{x_0}M$ (red vectors). This effectively linearises the geometry around $x_0$; (b) The stochastic development map $\varphi_u$ maps $\mathbb{R}^d$ valued paths $w_t$ to $M$ by transporting the covariance in each step (blue ellipses) giving a covariance $u_t$ along the entire sample path. The approach does not linearise around a single point. Holonomy of the connection implies that the covariance “rotates” around closed loops—an effect which can be illustrated by continuing the transport along the loop created by the dashed path. The anisotropic metric $g_{FM}$ weights step lengths by the transported covariance at each time $t$.

The map $u$ can be represented as a point in the frame bundle $FM$ of $M$. When the orthonormal requirement on $u$ is relaxed so that $uN(0, \text{Id})$ is a normal distribution in $T_{x_0}M$ with anisotropic covariance, the negative log-likelihood in $T_{x_0}M$ is related to $(u^{-1} \log_{x_0} x)^T (u^{-1} \log_{x_0} x)$ in the same way as the precision matrix $\Sigma^{-1}$ is related to the negative log-likelihood $(x - x_0)^T \Sigma^{-1} (x - x_0)$ in Euclidean space. The distance is thus weighted by the anisotropy of $u$, and $u$ can be interpreted as a square root covariance matrix $\Sigma^{1/2}$. 
However, the above approach does not specify how \( u \) changes when moving away from the base point \( x_0 \). The use of \( \log_{x_0} x \) effectively linearises the geometry around \( x_0 \), but a geometrically natural way to relate \( u \) at points nearby to \( x_0 \) will be to parallel transport it, equivalently specifying that \( u \) when transported does not change as measured from the curved geometry. This constraint is nonholonomic, and it implies that any path from \( x_0 \) to \( x \) carries with it a parallel transport of \( u \) lifting paths from \( M \) to paths in the frame bundle \( FM \). It therefore becomes natural to equip \( FM \) with a form of metric that encodes the anisotropy represented by \( u \). The result is the sub-Riemannian metric on \( FM \) defined below that weights infinitesimal movements on \( M \) using the parallel transport of the frame \( u \). Optimal paths for this metric are sub-Riemannian geodesics giving the family of most probable paths for the driving process that this paper concerns. Figure 1 shows one such path for an anisotropic normal distribution with \( M \) an ellipsoid embedded in \( \mathbb{R}^3 \).

2. Frame Bundles, Stochastic Development, and Anisotropic Diffusions

Let \( M \) be a finite dimensional manifold of dimension \( d \) with connection \( C \), and let \( x_0 \) be a fixed point in \( M \). When a Riemannian metric is present, and \( C \) is its Levi–Civita connection, we denote the metric \( g_R \). For a given interval \([0, T]\), we let \( W(M) \) denote the Wiener space of continuous paths in \( M \) starting at \( x_0 \). Similarly, \( W(\mathbb{R}^d) \) is the Wiener space of paths in \( \mathbb{R}^d \). We let \( H(\mathbb{R}^d) \) denote the subspace of \( W(\mathbb{R}^d) \) of finite energy paths.

Let now \( u = (u_1, \ldots, u_d) \) be a frame for \( T_xM \), \( x \in M \); i.e., \( u_1, \ldots, u_d \) is an ordered set of linearly independent vectors in \( T_xM \) with \( \text{span}\{u_1, \ldots, u_d\} = T_xM \). We can regard the frame as an isomorphism \( u : \mathbb{R}^d \to T_xM \) with \( u(e_i) = u_i \), where \( e_1, \ldots, e_d \) denotes the standard basis in \( \mathbb{R}^d \). Stochastic development (e.g., [4]) provides an invertible map \( \varphi_u \) from \( W(\mathbb{R}^d) \) to \( W(M) \). Through \( \varphi_u \), Euclidean semi-martingales map to stochastic processes on \( M \). When \( M \) is Riemannian and \( u \) orthonormal, the result is the Eells–Elworthy–Malliavin construction of Brownian motion [17]. We here outline the geometry behind development, stochastic development, the connection, and curvature, focusing in particular on frame bundle geometry.

2.1. The Frame Bundle

For each point \( x \in M \), let \( F_xM \) be the set of frames for \( T_xM \) (i.e., the set of ordered bases for \( T_xM \)). The set \( \{F_xM\}_{x \in M} \) can be given a natural differential structure as a fiber bundle on \( M \) called the frame bundle \( FM \). It can equivalently be defined as the principal bundle \( GL(\mathbb{R}^d, TM) \). We let the map \( \pi : FM \to M \) denote the canonical projection. The kernel of \( \pi_* : TFM \to TM \) is the sub-bundle of \( TFM \) that consists of vectors tangent to the fibers \( \pi^{-1}(x) \). It is denoted the vertical subspace \( VFM \).

We will often work in a local trivialization \( u = (x, u_1, \ldots, u_d) \in FM \), where \( x = \pi(u) \in M \) denotes the base point, and for each \( i = 1, \ldots, d \), \( u_i \in T_xM \) is the \( i \)-th frame vector. For \( v \in T_xM \) and \( u \in FM \) with \( \pi(u) = x \), the vector \( u^{-1}v \in \mathbb{R}^d \) expresses \( v \) in components in terms of the frame \( u \). We will denote the vector \( u^{-1}v \) frame coordinates of \( v \).

For a differentiable curve \( x_t \) in \( M \) with \( x = x_0 \), a frame \( u \) for \( T_{x_0}M \) can be parallel transported along \( x_t \) by parallel transporting each vector in the frame, thus giving a path \( u_t \in FM \). Such paths are called horizontal, and have zero acceleration in the sense \( C'(u_t) = 0 \). For each \( x \in M \), their derivatives form a \( d \)-dimensional subspace of the \( d + d^2 \)-dimensional tangent space \( T_xFM \).

This horizontal subspace \( HFM \) and the vertical subspace \( VFM \) together split the tangent bundle of \( FM \) (i.e., \( TFM = HFM \oplus VFM \)). The split induces a map \( \pi_* : HFM \to TM \), see Figure 3. For fixed \( u \in FM \), the restriction \( \pi_*(-)_{|H_FM} : H_uFM \to T_uM \) is an isomorphism. Its inverse is called the horizontal lift and is denoted \( \lambda_u \). Using \( h_{u} \), horizontal vector fields \( H_u \) on \( FM \) are defined for vectors \( v \in \mathbb{R}^d \) by \( H_u(v) = h_u(uv) \). In particular, the standard basis \( (e_1, \ldots, e_d) \) on \( \mathbb{R}^d \) gives \( d \) globally defined horizontal vector fields \( H_i \in HFM \), \( i = 1, \ldots, d \) by \( H_i = H_{e_i} \). Intuitively, the fields \( H_i(u) \) model infinitesimal transformations in \( M \) of \( x_0 \) in direction \( u_i = ve_i \) with corresponding infinitesimal parallel transport of the vectors \( u_1, \ldots, u_d \) of the frame along the direction \( u_i \). A horizontal lift of a
differentiable curve \( x_t \in M \) is a curve in \( FM \) tangent to \( HFM \) that projects to \( x_t \). Horizontal lifts are unique up to the choice of initial frame \( u_0 \).

\[
\begin{align*}
VFM & \xrightarrow{h + v \mapsto v} TFM & & \xrightarrow{h + v \mapsto h} HFM & & \xrightarrow{\pi} FM \\
\psi & & \pi & & \pi_{TFM} & & \pi_{FM} \\
FM \times \mathfrak{gl}(n) & \xrightarrow{T^*FM} T^*FM & & \xrightarrow{T^*M} TM & & \xrightarrow{\pi_{FM}} M
\end{align*}
\]

**Figure 3.** Relations between the manifold, frame bundle, the horizontal distribution \( HFM \), the vertical bundle \( VFM \), a Riemannian metric \( g_R \), and the sub-Riemannian metric \( g_{FM} \), defined below. The connection \( \mathcal{C} \) provides the splitting \( TFM = HFM \oplus VFM \). The restrictions \( \pi_s \vert_{HFM} \) are invertible maps \( H_uM \rightarrow T_{\pi(u)}M \) with inverse \( h_u \), the horizontal lift. Correspondingly, the vertical bundle \( VFM \) is isomorphic to the trivial bundle \( FM \times \mathfrak{gl}(n) \). The metric \( g_{FM} : T^*FM \rightarrow TFM \) has an image in the subspace \( HFM \).

### 2.2. Development and Stochastic Development

Let \( x_t \) be a differentiable curve on \( M \) and \( u_t \) a horizontal lift. If \( s_t \) is a curve in \( \mathbb{R}^d \) with components \( s_t^i \) such that \( \dot{x}_t = H_i(u_t) s_t^i \), \( x_t \) is said to be a development of \( s_t \). Correspondingly, \( s_t \) is the anti-development of \( x_t \). For each \( t \), the vector \( s_t \) contains frame coordinates of \( x_t \) as defined above. Similarly, let \( W_t \) be an \( \mathbb{R}^d \) valued Brownian motion so that sample paths \( W_t(\omega) \in W(\mathbb{R}^d) \). A solution to the stochastic differential equation \( dU_t = \sum_{i=1}^{d} H_i(U_t) \circ dW_t^i \) in \( FM \) is called a stochastic development of \( W_t \) in \( FM \). The solution projects to a stochastic development \( X_t = \pi(U_t) \) in \( M \). We call the process \( W_t \) in \( \mathbb{R}^d \), that through \( \phi \) maps to \( X_t \), the driving process of \( X_t \). Let \( \theta_u : W(\mathbb{R}^d) \rightarrow W(M) \) be the map that for fixed \( u \) sends a path in \( \mathbb{R}^d \) to its development on \( M \). Its inverse \( \theta_u^{-1} \) is the anti-development in \( \mathbb{R}^d \) of paths on \( M \) given \( u \).

Equivalent to the fact that normal distributions \( \mathcal{N}(0, \Sigma) \) in \( \mathbb{R}^d \) can be obtained as the transition distributions of diffusion processes \( \Sigma^{1/2}W_t \) stopped at time \( t = 1 \), a general class of distributions on the manifold \( M \) can be defined by stochastic development of processes \( W_t \), resulting in \( M \)-valued random variables \( X = X_1 \). This family of distributions on \( M \) introduced in [2] is denoted **anisotropic normal distributions**. The stochastic development by construction ensures that \( U_t \) is horizontal, and the frames are thus parallel transported along the stochastic displacements. The effect is that the frames stay covariantly constant, thus resembling the Euclidean situation where \( \Sigma^{1/2} \) is spatially constant and therefore does not change as \( W_t \) evolves. Thus, as further discussed in Section 3.2, the covariance is kept constant at each of the infinitesimal stochastic displacements. The existence of a smooth density for the target process \( X_t \) and small time asymptotics are discussed in [3].

Stochastic development gives a map \( \int_{\text{Diff}} : FM \rightarrow \text{Prob}(M) \) to the space of probability distributions on \( M \). For each point \( u \in FM \), the map sends a Brownian motion in \( \mathbb{R}^d \) to a distribution \( \mu_u \) by stochastic development of the process \( U_t \) in \( FM \), starting at \( u \) and letting \( \mu_u \) be the distribution of \( X = \pi(U_t) \). The pair \( (x, u) \), \( x = \pi(u) \) is analogous to the parameters \( (\mu, \Sigma) \) for a Euclidean normal distribution: the point \( x \in M \) represents the starting point of the diffusion, and the frame \( u \) represents a square root \( \Sigma^{1/2} \) of the covariance \( \Sigma \). In the general situation where \( \mu_u \) has smooth density, the construction can be used to fit the parameters \( u \) to data by maximum likelihood. As an example, diffusion PCA fits distributions obtained through \( \int_{\text{Diff}} \) by maximum likelihood to observed samples in \( M \); i.e., it optimizes for the most likely parameters \( u = (u_1, \ldots, u_d) \) for the anisotropic diffusion process, giving a fit to the data of the manifold generalization of the Euclidean normal distribution.

### 2.3. Adapted Coordinates

For concrete expressions of the geometric constructions related to frame bundles, and for computational purposes, it is useful to apply coordinates that are adapted to the horizontal bundle.
The map \( \pi_\ast : HFM \to TM \) is in coordinates of the adapted frame \( \pi_\ast(w^iD_j) = w^i\partial_{x^j} \). Correspondingly, the horizontal lift \( h_u \) is \( h_u(w^i\partial_{x^j}) = w^iD_j \). The map \( u: \mathbb{R}^d \to T_x M \) is given by the matrix \([u^a_i]\) so that \( uv = u^a_kv^k\partial_{x^a} = u_a v^a \).

Switching between standard coordinates and the adapted frame and coframes can be expressed in terms of the component matrices \( A \) below the frame and coframe induced by the coordinates \((x^i, u^a_i)\) and the adapted frame \( D \) and coframe \( D^* \). We have

\[
(\partial_x, \partial_{u^a_i})^A_D = \begin{bmatrix} I & 0 \\ -\Gamma & I \end{bmatrix}
\]

writing \( \Gamma \) for the matrix \([\Gamma^b_{ij}]\). Similarly, the component matrices of the dual frame \( D^* \) are

\[
(\partial_x, \partial_{u^a_i})^{A_{D^*}} = \begin{bmatrix} I & \Gamma^T \\ 0 & I \end{bmatrix}
\]

and \( D^* A(\partial_x, \partial_{u^a_i}) = \begin{bmatrix} I & -\Gamma^T \\ 0 & I \end{bmatrix} \).

2.4. Connection and Curvature

The \( TM \) valued connection \( C : TM \times TM \to TM \) lifts to a principal connection \( TFM \times TFM \to VFM \) on the principal bundle \( FM \). \( C \) can then be identified with the \( \mathfrak{gl}(n) \)-valued connection form \( \omega \) on \( TFM \). The identification occurs by the isomorphism \( \psi \) between \( FM \times \mathfrak{gl}(n) \) and \( VFM \) given by \( \psi(u, v) = \frac{d}{dt} u \exp(tv)|_{t=0} \) (e.g., [19,20]).

The map \( \psi \) is equivariant with respect to the \( \mathfrak{gl}(n) \) action \( g \mapsto u^g \) on \( FM \). In order to explicitly see the connection between the usual covariant derivative \( \nabla : (\Gamma(TM) \times \Gamma(TM)) \to \Gamma(TM) \) on \( M \) determined by \( C \) and \( \Gamma \) regarded as a connection on the principal bundle \( FM \), following [19], we let \( s : M \to TM \) be a local vector field on \( M \); equivalently, \( s \in \Gamma(TM) \) is a local section of \( TM \). \( s \) determines a map \( s^{FM} : FM \to \mathbb{R}^d \) by \( s^{FM}(u) = u^{-1}s(\pi(u)) \); i.e., it gives the coordinates of \( s(x) \) in the frame \( u \) at \( x \). The pushforward \( (s^{FM})_\ast : TFM \to \mathbb{R}^d \) has in its \( i \)th component the exterior derivative \( d(s^{FM})_i \).

Let now \( w(x) \) be a local section of \( FM \). The composition \( w \circ (s^{FM})_\ast \circ h_u : TM \to TM \) is identical to the covariant derivative \( \nabla(s : TM \to TM) \). The construction is independent of the choice of \( w \) because of the \( \mathfrak{gl}(n) \)-equivariance of \( s^{FM} \). The connection form \( \omega \) can be expressed as the matrix \((s_1^{FM} \circ h_{w_1}, \ldots, s_d^{FM} \circ h_{w_d})\) when letting \( s_i^{FM}(u) = e_i \).

The identification becomes particularly simple if the covariant derivative is taken along a curve \( x_t \) on which \( w_t \) is the horizontal lift. In this case, we can let \( s_t = w_t s_i^t \). Then, \( s^{FM}(w_1) = (s_1^t, \ldots, s_d^t)^T \), and

\[
\frac{d}{dt}(s_1^t, \ldots, s_d^t)^T = \frac{d}{dt}s_i^t \nabla_{x_t} s_i = (s^{FM})_\ast(h_{w_t}(x_t)) = \frac{d}{dt}(s_1^t, \ldots, s_d^t)^T
\]

i.e., the covariant derivative takes the form of the standard derivative applied to the frame coordinates \( s_i^t \).

The curvature tensor \( R \in T^3_1(M) \) gives the \( \mathfrak{gl}(n) \)-valued curvature form \( \Omega : TFM \times TFM \to \mathfrak{gl}(n) \) on \( TFM \) by

\[
\Omega(v_{u}, w_{u}) = u^{-1}R(\pi_\ast(v_{u}), \pi_\ast(w_{u}))u, \quad v_u, w_v \in TFM.
\]
Note that \( \Omega(v_u, w_u) = \Omega(h_u(\pi_x(v_u)), h_u(\pi_x(w_u))) \), which we can use to write the curvature \( R \) as the \( g(n) \)-valued map \( R_u : T^2(T_{\pi(u)}M) \to g(n) \), \((v, w) \mapsto \Omega(h_u(\pi_x(v_u)), h_u(\pi_x(w_u))) \) for fixed \( u \).

In coordinates, the curvature is

\[
R_{ijk}^s = \Gamma_{ik}^l \Gamma_{jl}^s - \Gamma_{ik}^l \Gamma_{jl}^s + \Gamma_{ij}^s \Gamma_{lk}^l - \Gamma_{ij}^s \Gamma_{lk}^l,
\]

where \( \Gamma^s_{ik} = \frac{\partial x^l}{\partial x^i} \Gamma^s_{jk} \).

Let \( x_{t,s} \) be a family of paths in \( M \), and let \( u_{t,s} \in \pi^{-1}(x_{t,s}) \) be horizontal lifts of \( x_{t,s} \) for each fixed \( s \). Write \( \dot{x}_{t,s} = \frac{\partial}{\partial s} x_{t,s} \) and \( \partial_t = \frac{\partial}{\partial t} u_{t,s} \). The \( s \)-derivative of \( u_{t,s} \) can be regarded a pushforward of the horizontal lift and is the curve in TFM

\[
\partial_s u_{t,s} = \psi(u_{t,s}, \psi_{u_{0,s}}^{-1}(C(\partial_t u_{0,s}))) + \int_0^s \Omega(\dot{u}_{r,s}, \partial_s u_{r,s})dr + h_{u_{1,s}}(\partial_s x_{t,s})
\]

\[
= \psi(u_{t,s}, \psi_{u_{0,s}}^{-1}(C(\partial_t u_{0,s}))) + \int_0^s R_{u_{r,s}}(\dot{x}_{r,s}, \partial_s x_{r,s})dr + h_{u_{1,s}}(\partial_s x_{t,s}).
\]

This follows from the structure equation \( dw = -\omega \wedge \omega + \Omega \) (e.g., [21]). Note that the curve depends on the vertical variation \( C(\partial_t u_{0,s}) \) at only one point along the curve. The remaining terms depend on the horizontal variation or, equivalently, \( \partial_s x_{t,s} \). The \( t \)-derivative of \( \partial_s x_{t,s} \) is the curve in TTFM satisfying

\[
\partial_t h_{u_{t,s}}(\dot{x}_{t,s}) = \psi(u_{t,s}, R_{u_{t,s}}(\dot{x}_{t,s}, \partial_t x_{t,s})) + \partial_t \psi(u_{t,s}, \psi_{u_{t,s}}^{-1}(C(\partial_s u_{t,s}))) + \partial_t (h_{u_{1,s}}(\partial_s x_{t,s}))
\]

\[
= \psi(u_{t,s}, R_{u_{t,s}}(\dot{x}_{t,s}, \partial_t x_{t,s})) + \partial_t \psi(u_{t,s}, \psi_{u_{t,s}}^{-1}(C(\partial_s u_{t,s})))
\]

\[
+ h_{u_{1,s}}(\partial_t \partial_s x_{t,s}) + (\partial_t h_{u_{1,s}})(\partial_s x_{t,s}).
\]

Here, the first and third term in the last expression are identified with elements of \( T_{\partial_s x_{t,s}} \) TFM by the natural mapping \( T_{\partial_s x_{t,s}} FM \to T_{\partial_s x_{t,s}} TFM \). When \( C(\partial_s u_{t,s}) \) is zero, the relation reflects the property that the curve arises when computing brackets between horizontal vector fields. Note that the first term of (3) has values in VFM, while the third term has values in HFM.

3. The Anisotropically Weighted Metric

For a Euclidean driftless diffusion process with spatially constant stochastic generator \( \Sigma \), the log-probability of a sample path can formally be written

\[
\ln \hat{p}_\Sigma(x_t) \propto -\int_0^1 \| \dot{x}_t \|^2 dt + c_\Sigma
\]

with the norm \( \| \cdot \|_\Sigma \) given by the inner product \( \langle v, w \rangle_\Sigma = \langle \Sigma^{-1/2} v, \Sigma^{-1/2} w \rangle = v \Sigma^{-1} w \); i.e., the inner product weighted by the precision matrix \( \Sigma^{-1} \). Though only formal, as the sample paths are almost surely nowhere differentiable, the interpretation can be given a precise meaning by taking limits of piecewise linear curves [21]. Turning to the manifold situation with the processes mapped to \( M \) by stochastic development, the probability of observing a differentiable path can either be given a precise meaning in the manifold by taking limits of small tubes around the curve, or in \( \mathbb{R}^d \) by considering infinitesimal tubes around the anti-development of the curves. With the former formulation, a scalar curvature correction term must be added to (4), giving the Onsager–Machlup function [22]. The latter formulation corresponds to defining a notion of path density for the driving \( \mathbb{R}^d \)-valued process \( W_t \). When \( M \) is Riemannian and \( \Sigma \) unitary, taking the maximum of (4) gives geodesics as most probable paths for the driving process.

Let now \( U_t \) be a path in \( FM \), and choose a local trivialization \( u_t = (x_t, u_{1,t}, \ldots, u_{d,t}) \) such that the matrix \( [u_{a,t}] \) represents the square root covariance matrix \( \Sigma^{1/2} \) at \( x_t \). Since \( u_t \) being a frame defines an
invertible map $\mathbb{R}^d \to T_x M$, the norm $\| \cdot \|_{\Sigma}$ above has a direct analogue in the norm $\| \cdot \|_u$ defined by the inner product

$$\langle v, w \rangle_u = \left\langle u^{-1}v, u^{-1}w \right\rangle_{\mathbb{R}^d}$$

for vectors $v, w \in T_x M$. The transport of the frame along paths in effect defines a transport of inner product along sample paths: the paths carry with them the inner product weighted by the precision matrix, which in turn is a transport of the square root covariance $u^{-1}$. The inner product can equivalently be defined as a metric $g_u : T_x^* M \to T_x^* M$. Again using that $u$ can be considered a map $\mathbb{R}^d \to T_x M$, $g_u$ is defined by $\xi \mapsto u_1(\xi \circ u)^2$, where $\xi$ is the standard identification $(\mathbb{R}^d)^* \to \mathbb{R}^d$. The sequence of mappings defining $g_u$ is illustrated below:

$$T_x^* M \to (\mathbb{R}^d)^* \to \mathbb{R}^d \to T_x^* M \quad \xi \mapsto \xi \circ u \mapsto (\xi \circ u)^2 \mapsto u(\xi \circ u)^2.$$  

This definition uses the $\mathbb{R}^d$ inner product in the definition of $\xi$. Its inverse gives the cometric $g_u^{-1} : T_x M \to T_x^* M$; i.e., $v \mapsto \langle u^{-1}v \rangle^\circ \circ u^{-1}$.

3.1. Sub-Riemannian Metric on the Horizontal Distribution

We now lift the path-dependent metric defined above to a sub-Riemannian metric on $HFM$. For any $w, \tilde{w} \in H_u FM$, the lift of (5) by $\pi_*$ is the inner product

$$\langle w, \tilde{w} \rangle = \left\langle u^{-1} \pi_* w, u^{-1} \pi_* \tilde{w} \right\rangle_{\mathbb{R}^d}.$$

The inner product induces a sub-Riemannian metric $g_{FM} : TFM^* \to HFM \subset TFM$ by

$$\langle w, g_{FM}(\tilde{w}) \rangle = \langle \xi | w \rangle, \quad \forall w \in H_u FM$$

with $\langle \xi | w \rangle$ denoting the evaluation $\xi(w)$ for the covector $\xi \in T^* FM$. The metric $g_{FM}$ gives $FM$ a non-bracket-generating sub-Riemannian structure [23] on $FM$ (see also Figure 3). It is equivalent to the lift

$$\xi \mapsto h_u(g_u(\xi \circ h_u)), \quad \xi \in H_u FM$$

of the metric $g_u$ above. In frame coordinates, the metric takes the form

$$u^{-1} \pi_* g_{FM}(\xi) = \begin{pmatrix} \xi(H_1(u)) \\ \vdots \\ \xi(H_d(u)) \end{pmatrix}. $$

In terms of the adapted coordinates for $TFM$ described in Section 2.3, with $w = \omega^j D_j$ and $\tilde{w} = \tilde{\omega}^i D_i$, we have

$$\langle w, \tilde{w} \rangle = \left\langle \omega^j D_j, \tilde{\omega}^i D_i \right\rangle = \left\langle u^{-1} \omega^j \partial_x^j, u^{-1} \tilde{\omega}^i \partial_{x^i} \right\rangle = \left\langle \omega^a \omega^b \partial \omega^a \partial \omega^b \right\rangle_{\mathbb{R}^d} = \delta_{ab} \omega^a \tilde{\omega}^b \tilde{u}^a u^a \tilde{u}^b = W_{ij} \omega^i \tilde{\omega}^j$$

where $[u^a]$ is the inverse of $[u^a]$ and $W_{ij} = \delta_{\alpha \beta} u_\alpha^a u_\beta^b$. Define now $W^{ij} = \delta^{ij} u_\alpha^a u_\beta^b$, so that $W^{ir} W_{ij} = \delta^i_j$ and $W_{ir} W^{ij} = \delta^i_j$. We can then write the metric $g_{FM}$ directly as

$$g_{FM}(\xi_h D_h + \xi_{h^*} D_{h^*}) = W^{ij} \xi_h D_i,$$
because $\langle w, g_{FM}(\xi) \rangle = \langle w, W^h \xi_h D_l \rangle = W_{ij} w^i W^h \xi_h = w^i \xi_i D_l (w^i D_l) = \xi(w)$. One clearly recognizes the dependence on the horizontal $H^* FM$ part of $T^* FM$ only, and the fact that $g_{FM}$ has image in $HFM$. The sub-Riemannian energy of an almost everywhere horizontal path $u_t$ is

$$I_{FM}(u_t) = \int g_{FM}(u_t, u_t) dt,$$

i.e., the line element is $ds^2 = W_{ij} D^i D^j$ in adapted coordinates. The corresponding distance is given by

$$d_{FM}(u_1, u_2) = \inf \{ I_{FM}(\gamma) \mid \gamma(0) = u_1, \gamma(1) = u_2 \}.$$

If we wish to express $g_{FM}$ in canonical coordinates on $T^* FM$, we can switch between the adapted frame and the coordinates $(x^i, u^i, \xi^i, \xi_h)$. From (11), $g_{FM}$ has $D, D^*$ components

$$d g_{FM,D*} = \left[ \begin{array}{cc} W^{-1} & 0 \\ 0 & 0 \end{array} \right].$$

Therefore, $g_{FM}$ has the following components in the coordinates $(x^i, u^i, \xi^i, \xi_h)$

$$(\partial_{x^i}, \partial_{u^i}) g_{FM,(\partial_{x^i}, \partial_{u^i})} = (\partial_{x^i}, \partial_{u^i}) A_D D g_{FM,D*} D^* A_{(\partial_{x^i}, \partial_{u^i})} = \left[ \begin{array}{cc} W^{-1} & -W^{-1} \Gamma^T \\ -\Gamma W^{-1} & \Gamma W^{-1} \Gamma^T \end{array} \right],$$

or $g_{ij}^{ij} = W^{ij}, g_{ij}^{ij} = -W^{ij} \Gamma^T, g_{ij}^{ij} = -\Gamma_i^{ij} W^{ij},$ and $g_{ij}^{ij} = \Gamma_i^{ij} W^{ij} \Gamma^T$.  

### 3.2. Covariance and Nonholonomicity

The metric $g_{FM}$ encodes the anisotropic weighting given the frame $u$, thus up to an affine transformation measuring the energy of horizontal paths equivalently to the negative log-probability of sample paths of Euclidean anisotropic diffusions as formally given in (4). In addition, the requirement that paths must stay horizontal almost everywhere enforces that $C(\dot{u}_t) = 0$ a.e., i.e., that *no change of the covariance is measured by the connection*. The intuitive effect is that covariance is covariantly constant as seen by the connection. Globally, *curvature* of $C$ will imply that the covariance changes when transported along closed loops, and *torsion* will imply that the base point “slips” when travelling along covariantly closed loops on $M$. However, the zero acceleration requirement implies that the covariance is as close to spatially constant as possible with the given connection. This is enabled by the parallel transport of the frame, and it ensures that the model closely resembles the Euclidean case with spatially constant stochastic generator.

With non-zero curvature of $C$, the horizontal distribution is non-integrable (i.e., the brackets $[H_i, H_j]$ are non-zero for some $i, j$). This prevents integrability of the horizontal distribution $HFM$ in the sense of the Frobenius theorem. In this case, the horizontal constraint is *nonholonomic* similarly to nonholonomic constraints appearing in geometric mechanics (e.g., [24]). The requirement of covariantly constant covariance thus results in a nonholonomic system.

### 3.3. Riemannian Metrics on $FM$

If the horizontality constraint is relaxed, a related Riemannian metric on $FM$ can be defined by pulling back a metric on $gl(n)$ to each fiber using the isomorphism $\phi(u, \cdot)^{-1} : V_u FM \to gl(n)$. Therefore, the metric on $HFM$ can be extended to a Riemannian metric on $FM$. Such metrics incorporate the anisotropically weighted metric on $HFM$, however, allowing vertical variations and thus that covariances can change unrestricted.

When $M$ is Riemannian, the metric $g_{FM}$ is in addition related to the Sasaki–Mok metric on $FM$ [18] that extends the Sasaki metric on $TM$. As for the above Riemannian metric on $FM$, the Sasaki–Mok metric allows paths in $FM$ to have derivatives in the vertical space $VFM$. On $HFM$, the Riemannian
metric $g_R$ is here lifted to the metric $g_{SM} = (v_u, w_u) = g_R(\pi_*(v_u), \pi_*(w_u))$ (i.e., the metric is not anisotropically weighted). The line element in this case $ds^2 = g_{ij}dx^idx^j + X_{\beta,g}\xi_\beta D\hat\theta_i$.

Geodesics for $g_{SM}$ are lifts of Riemannian geodesics for $g_R$ on $M$, in contrast to the sub-Riemannian normal geodesics for $g_{FM}$ which we will characterize below. The family of curves arising as projections to $M$ of normal geodesics for $g_{FM}$ includes Riemannian geodesics for $g_R$ (and thus projections of geodesics for $g_{SM}$), but the family is in general larger than geodesics for $g_R$.

4. Constrained Evolutions

Extremal paths for (5) can be interpreted as most probable paths for the driving process $W_t$ when $u_0$ defines an anisotropic diffusion. This is captured in the following definition [3]:

**Definition 1.** A most probable path for the driving process (MPP) from $x = \pi(u_0) \in M$ to $y \in M$ is a smooth path $x_t : [0,1] \rightarrow M$ with $x_0 = x$ and $x_1 = y$ such that its anti-development $\varphi_{u_0}^{-1}(x_t)$ is a most probable path for $W_t$; i.e.,

$$x_t \in \arg \min_{x_t \in x, x_t = y} \int_0^1 -L_{\mathbb{R}^d}(\varphi_{u_0}^{-1}(\gamma_t), \frac{d}{dt}\varphi_{u_0}^{-1}(\gamma_t)) \, dt$$

with $L_{\mathbb{R}^d}$ being the Onsager–Machlup function for the process $W_t$ on $\mathbb{R}^d$ [22].

The definition uses the one-to-one relation between $W(\mathbb{R}^d)$ and $W(M)$ provided by $\varphi_{u_0}$ to characterize the paths using the $\mathbb{R}^d$ Onsager–Machlup function $L_{\mathbb{R}^d}$. When $M$ is Riemannian with metric $g_R$, the Onsager–Machlup function for a $\xi$-Brownian motion on $M$ is $L(x_t, x_i) = -\frac{1}{2}||x_t||^2_{g_R} + \frac{1}{12}S_{g_R}(x_t)$ with $S_{g_R}$ denoting the scalar curvature. This curvature term vanishes on $\mathbb{R}^d$, and therefore $L_{\mathbb{R}^d}(\gamma_t, \gamma_t) = -\frac{1}{2}||\gamma_t||^2$ for a curve $\gamma_t \in \mathbb{R}^d$.

By pulling $x_t \in M$ back to $\mathbb{R}^d$ using $\varphi_{u_0}^{-1}$, the construction removes the $\frac{1}{12}S_{g_R}(x_t)$ scalar curvature correction term present in the non-Euclidean Onsager–Machlup function. It thereby provides a relation between geodesic energy and most probable paths for the driving process. This is contained in the following characterization of most probable paths for the driving process as extremal paths of the sub-Riemannian distance [3] that follows from the Euclidean space Onsager–Machlup theorem [22].

**Theorem 1 ([3]).** Let $Q(u_0)$ denote the principal sub-bundle of $FM$ of points $z \in FM$ reachable from $u_0 \in FM$ by horizontal paths. Suppose the Hörmander condition is satisfied on $Q(u_0)$, and that $Q(u_0)$ has compact fibers. Then, most probable paths from $x_0$ to $y \in M$ for the driving process of $X_t$ exist, and they are projections of sub-Riemannian geodesics in $FM$ minimizing the sub-Riemannian distance from $u_0$ to $\pi^{-1}(y)$.

Below, we will derive evolution equations for the set of such extremal paths that correspond to normal sub-Riemannian geodesics.

4.1. Normal Geodesics for $g_{FM}$

Connected to the metric $g_{FM}$ is the Hamiltonian

$$H(z) = \frac{1}{2}(|z|_{g_{FM}(z)})$$

(12)

on the symplectic space $T^*FM$. Letting $\hat{\gamma}$ denote the projection on the bundle $T^*FM \rightarrow FM$, (8) gives

$$H(z) = \frac{1}{2} \langle g_{FM}(z)|g_{FM}(z) \rangle = \frac{1}{2}||z \circ h_{R(z)} \circ \pi(z)||^2_{[\mathbb{R}^d]} = \frac{1}{2} \sum_{i=1}^d \xi_i(H_i(u))^2.$$

Normal geodesics in sub-Riemannian manifolds satisfy the Hamilton–Jacobi equations [23] with Hamiltonian flow

$$\dot{z}_t = X_H = \Omega^\#dH(z)$$

(13)
where $\Omega$ here is the canonical symplectic form on $T^*FM$ (e.g., [25]). We denote (13) the MPP equations, and we let projections $x_t = \pi_{T^*FM}(z_t)$ of minimizing curves satisfying (13) be denoted normal MPPs. The system (13) has $2(d + d^2)$ degrees of freedom, in contrast to the usual $2d$ degrees of freedom for the classical geodesic equation. Of these, $d^2$ describes the current frame at time $t$, while the remaining $d^2$ allows the curve to “twist” while still being horizontal. We will see this effect visualized in Section 6.

In a local canonical trivialization $z = (u, \xi)$, (13) gives the Hamilton–Jacobi equations

$$ u = \partial_\xi H(u, \xi) = g_{FM}(u, \xi) = h_u(u(\xi(H_1(u), \ldots, \xi(H_d(u))))^T) $$

$$ \dot{\xi} = -\partial_u H(u, \xi) = -\partial_u \frac{1}{2} \| \xi \circ h_u \circ u \|^2_{(\mathbb{R}^d)} = -\partial_u \frac{1}{2} \sum_{i=1}^d \xi(H_i(u))^2. \tag{14} $$

Using (3), we have for the second equation

$$ \dot{\xi} = -\sum_{i=1}^d \xi(H_i(u)) \partial_u h_u(ue_i) $$

$$ = -\sum_{i=1}^d \xi(H_i(u)) \xi(\partial_u h_u(ue_i)) $$

$$ = -\xi(\psi(u, R_u(\pi_s(u), \pi_s(\partial_u))) + \partial_u \psi(u, \psi^{-1}(C(\partial_u))) + \partial_u h_u(\pi_s(\partial_u))). \tag{15} $$

Here $\partial_u$ denotes $u$-derivative in the direction $u$, equivalently $\partial_u h_u(v) = \partial_t(h_u)(v)$. While the first equation of (14) involves only the horizontal part of $\dot{\xi}$, the second equation couples the vertical part of $\dot{\xi}$ through the evaluation of $\dot{\xi}$ on the term $\psi(u, R_u(\pi_s(u), \pi_s(\partial_u)))$. If the connection is curvature-free, which in non-flat cases implies that it carries torsion, this vertical term vanishes. Conversely, when $M$ is Riemannian, $C$ the $g_R$ Levi–Civita connection, and $u_0$ is $g_R$ orthonormal, $g_{FM}(h_u(v), h_u(w)) = g_R(v, w)$ for all $v, w \in T_{\pi(u)}M$. In this case, a normal MPP $\pi(u_t)$ will be a Riemannian $g_R$ geodesic.

4.2. Evolution in Coordinates

In coordinates $u = (x^i, u^i, \xi_i, \xi_{ik})$ for $T^*FM$, we can equivalently write

$$ \dot{x}^i = g^{ij} \dot{x}_j + g^{ijkl} \dot{\xi}_j \xi_{kl} = W^{ij} \dot{x}_j - W^{ijkl} \xi_j \xi_{kl} $$

$$ \dot{\xi}_i = g^{ijkl} \dot{\xi}_j \xi_{kl} = -\Gamma^i_{jk} W^{jk} \dot{x}_j + \Gamma^i_{jk} W^{jk} \Gamma^j_{lk} \xi_l $$

$$ \dot{\xi}_k = -\frac{1}{2} \left( \partial_y \partial_y g^{hk} \xi_k \xi_h + \partial_y \partial_y g^{hk} \xi_h \xi_k - \partial_y \partial_y g^{hk} \xi_h \xi_k + \partial_y \partial_y g^{hk} \xi_h \xi_k \right) $$

$$ \dot{\xi}_{ik} = -\frac{1}{2} \left( \partial_y \partial_y g^{hk} \xi_k \xi_h + \partial_y \partial_y g^{hk} \xi_h \xi_k - \partial_y \partial_y g^{hk} \xi_h \xi_k + \partial_y \partial_y g^{hk} \xi_h \xi_k \right) $$

with $\Gamma^i_{jk}$ for $\partial_y \Gamma^i_{jk}$, and where

$$ \partial_y g^{ij} = 0, \quad \partial_y g^{ijLP} = -W^{ijLP} h_{LJP}, \quad \partial_y g^{ijLP} = -W^{ijLP} h_{LJP}, \quad \partial_y g^{ijLP} = -W^{ijLP} h_{LJP}, \quad \partial_y g^{ijLP} = -W^{ijLP} h_{LJP}, \quad \partial_y g^{ijLP} = -W^{ijLP} h_{LJP}, $$

$$ \partial_y \dot{g}^{ij} = W_{ijl} \dot{x}_l, \quad \partial_y \dot{g}^{ijLP} = -W_{ijL} h_{LJP}, \quad \partial_y \dot{g}^{ijLP} = -W_{ijL} h_{LJP}, \quad \partial_y \dot{g}^{ijLP} = -W_{ijL} h_{LJP}, $$

$$ \partial_y \dot{g}^{ijLP} = -W_{ijL} h_{LJP}, \quad \Gamma^i_{jk} = \Gamma^i_{jk} h_{LJP} + \Gamma^i_{jk} \Gamma^j_{lp} h_{LJP}, \quad \Gamma^i_{jk} = \partial_y \left( \Gamma^i_{jk} u^k_a \right) = \partial_y \Gamma^i_{jk}, \quad \Gamma^i_{jk} = \partial_y \Gamma^i_{jk} + \partial_y \Gamma^i_{jk}. $$
Combining these expressions, we obtain

\[ \dot{x}^i = W^{ij} \dot{y}^j - W^{ih} \Gamma^j_h y^h, \quad \dot{\xi}^i = \dot{\xi}^i = \Gamma^i_h y^h, \quad \dot{\xi}^h = -\Gamma^h_i y^i + \Gamma^h_i W^{jk} \Gamma^i_j y^k \]

For curves satisfying the MPP equations, using (16) and (15), we have

\[ u_t^{-1} \nabla_{\dot{x}_t} \dot{x}_t = \frac{d}{dt} \begin{pmatrix} \xi(h_{u_1}(u_1e_1)) \\ \vdots \\ \xi(h_{u_1}(u_1e_d)) \\ \xi(h_{u_1}(u_1e_d)) \end{pmatrix} = \begin{pmatrix} \xi(h_{u_1}(u_1e_1)) \\ \vdots \\ \xi(h_{u_1}(u_1e_d)) \end{pmatrix} + \begin{pmatrix} \xi(h_{u_1}(u_1e_1)) \\ \vdots \\ \xi(h_{u_1}(u_1e_d)) \end{pmatrix} \]

\[ = -\begin{pmatrix} \xi(\partial h_{u_1}(u_1e_1)h_{u_1}(\pi_\tau(\dot{u}_t))) \\ \vdots \\ \xi(\partial h_{u_1}(u_1e_1)h_{u_1}(\pi_\tau(\dot{u}_t))) \end{pmatrix} + \begin{pmatrix} \xi(\partial h_{u_1}(u_1e_1)h_{u_1}(\pi_\tau(\dot{u}_t))) \\ \vdots \\ \xi(\partial h_{u_1}(u_1e_1)h_{u_1}(\pi_\tau(\dot{u}_t))) \end{pmatrix} \]

\[ = \begin{pmatrix} \xi(\psi(\dot{u}_t, R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t))) \\ \vdots \\ \xi(\psi(\dot{u}_t, R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t))) \end{pmatrix} \]

The fact that the covariant derivative vanishes for classical geodesic leads to a definition of higher-order polynomials through the covariant derivative by requiring \((\nabla_{\dot{x}_t})^k \dot{x}_t = 0\) for a kth order polynomial (e.g., [26,27]). As discussed above, compared to classical geodesics, curves satisfying the MPP equations have extra \(d^2\) degrees of freedom, allowing the curves to twist and deviate from being geodesic with respect to \(C\) while still satisfying the horizontality constraint on FM. This makes it natural to ask if normal MPPs relate to polynomials defined using \(C\). For curves satisfying the MPP equations, using (16) and (15), we have

\[ u_t^{-1} (\nabla_{\dot{x}_t})^2 \dot{x}_t = \frac{d}{dt} \begin{pmatrix} \xi(\psi(\dot{u}_t, R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t))) \\ \vdots \\ \xi(\psi(\dot{u}_t, R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t))) \end{pmatrix} = \begin{pmatrix} \xi(\psi(\dot{u}_t, R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t))) \\ \vdots \\ \xi(\psi(\dot{u}_t, R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t)) \end{pmatrix} \]

Thus, in general, normal MPPs are not second order polynomials in the sense \((\nabla_{\dot{x}_t})^2 \dot{x}_t = 0\) unless the curvature \(R_{u_1}(u_1e_1, \pi_\tau(\dot{u}_t))\) is constant in \(t\).

For comparison, in the Riemannian case, a variational formulation placing a cost on covariant acceleration [28,29] leads to cubic splines

\[ (\nabla_{\dot{x}_t})^2 \dot{x}_t = -R(\nabla_{\dot{x}_t} \dot{x}_t, x_t) \dot{x}_t \]

In (16), the curvature terms appear in the covariant acceleration for normal MPPs, while cubic splines lead to the curvature term appearing in the third order derivative.

5. Cometric Formulation and Low-Rank Generator

We now investigate a cometric \(g^\perp_M + \lambda g_R\), where \(g_R\) is Riemannian, \(g^\perp_M\) is a rank k positive semi-definite inner product arising from \(k\) linearly independent tangent vectors, and \(\lambda > 0\) a weight. We assume that \(g^\perp_M\) is chosen so that \(g^\perp_M + \lambda g_R\) is invertible, even though \(g^\perp_M\) is rank-deficient.
The situation corresponds to extracting the first \( k \) eigenvectors in Euclidean space PCA. If the eigenvectors are estimated statistically from observed data, this allows the estimation to be restricted to only the first \( k \) eigenvectors. In addition, an important practical implication of the construction is that a numerical implementation need not transport a full \( d \times d \) matrix for the frame, but a potentially much lower dimensional \( d \times k \) matrix. This point is essential when dealing with high-dimensional data, examples of which are landmark manifolds as discussed in Section 6.

When using the frame bundle to model covariances, the sum formulation is natural to express as a cometric compared to a metric because, with the cometric formulation, \( g_{\mathcal{F}^k M} + \lambda g_R \) represents a sum of covariance matrices instead of a sum of precision matrices. Thus, \( g_{\mathcal{F}^k M} + \lambda g_R \) can be intuitively thought of as adding isotropic noise of variance \( \lambda \) to the covariance represented by \( g_{\mathcal{F}^k M} \).

To pursue this, let \( \mathcal{F}^k M \) denote the bundle of rank \( k \) linear maps \( \mathbb{R}^k \to T_x M \). We define a cometric by

\[
\langle \xi, \xi \rangle = \delta^{\alpha \beta} (\xi | h_{ui}(u_\alpha)) (\xi | h_{ui}(u_\beta)) + \lambda \langle \xi, \xi \rangle_{g_R}
\]

for \( \xi, \xi \in T_x \mathcal{F}^k M \). The sum over \( \alpha, \beta \) is for \( \alpha, \beta = 1, \ldots, k \). The first term is equivalent to the lift (9) of the cometric \( \langle \xi, \xi \rangle = \langle \xi | g_{\mathcal{F}^k M}(\xi) \rangle \) given \( u : \mathbb{R}^k \to T_x M \). Note that in the definition (6) of \( g_{ui} \), the map \( u \) is not inverted; thus, the definition of the metric immediately carries over to the rank-deficient case.

Let \( (\bar{x}', u_\alpha') \), \( \alpha = 1, \ldots, k \) be a coordinate system on \( \mathcal{F}^k M \). The vertical distribution is in this case spanned by the \( d \) vector fields \( D_{\bar{u}_\beta} = \partial_{u_\beta} \). Except for index sums being over \( k \) instead of \( d \) terms, the situation is thus similar to the full-rank case. Note that \( \langle \xi | \pi^{-1}_{\bar{x}} \rangle = \langle \xi | W_{\bar{x}} (D_{\bar{y}}) \rangle = w^l \xi_l \). The cometric in coordinates is

\[
\langle \xi, \xi \rangle = \delta^{\alpha \beta} u_\alpha^l \xi_l u_\beta^l + \lambda g_R^{ij} \xi_i \xi_j = \bar{\xi}^i \left( \delta^{\alpha \beta} u_\alpha^l u_\beta^l + \lambda g_R^{ij} \right) \bar{\xi}^j = \bar{\xi}^i W^{ij} \bar{\xi}^j
\]

with \( W^{ij} = \delta^{\alpha \beta} u_\alpha^l u_\beta^l + \lambda g_R^{ij} \). We can then write the corresponding sub-Riemannian metric \( g_{\mathcal{F}^k M} \) in terms of the adapted frame \( D \)

\[
g_{\mathcal{F}^k M}(\xi_h D^h + \xi_h \gamma D^h) = W^{ih} \xi_h D_i
\]

because \( \langle \xi | g_{\mathcal{F}^k M}(\xi) \rangle = \langle \xi, \xi \rangle = \bar{\xi}^i W^{ij} \bar{\xi}^j \). That is, the situation is analogous to (11), except the term \( \lambda g_R^{ij} \) is added to \( W^{ij} \).

The geodesic system is again given by the Hamilton–Jacobi equations. As in the full-rank case, the system is specified by the derivatives of \( g_{\mathcal{F}^k M} \):

\[
\partial_{y^i} g_{\mathcal{F}^k M} = W^{ij}_l, \quad \partial_{y^j} g_{\mathcal{F}^k M} = -W^{ih}_j \Gamma^j_h - W^{ih} \Gamma^j_{h,i}, \quad \partial_{y^j} g_{\mathcal{F}^k M} = -\Gamma^h_i - \Gamma^h_l W^{ih}_j - \Gamma^l_i W^{j h}_j,
\]

\[
\partial_{y^i} g_{\mathcal{F}^k M} = \Gamma^i_k W^{k h} \Gamma^j_h + \Gamma^i_k W^{k h} \Gamma^j_{h,l}, \quad \partial_{y^j} g_{\mathcal{F}^k M} = \Gamma^i_k W^{k h} \Gamma^j_h + \Gamma^i_k W^{k h} \Gamma^j_{h,l},
\]

\[
\partial_{y^i} g_{\mathcal{F}^k M} = W^{-ij}_l, \quad \partial_{y^j} g_{\mathcal{F}^k M} = -W^{il}_j \Gamma^j_h - W^{il} \Gamma^j_{h,i}, \quad \partial_{y^j} g_{\mathcal{F}^k M} = -\Gamma^i_k W^{i h}_j - \Gamma^i_k W^{j h}_j,
\]

\[
\partial_{y^i} g_{\mathcal{F}^k M} = \Gamma^i_k W^{k h} \Gamma^j_h + \Gamma^i_k W^{k h} \Gamma^j_{h,l}, \quad \partial_{y^j} g_{\mathcal{F}^k M} = \Gamma^i_k W^{k h} \Gamma^j_h + \Gamma^i_k W^{k h} \Gamma^j_{h,l},
\]

\[
\Gamma^i_k = \partial_{y^i} \left( \Gamma^i_{k,l} + \lambda g_R^{ij} \right) = \delta^{i \alpha} \Gamma^j_{h \alpha} , \quad W^{ij} = \lambda g_R^{ij} , \quad W^{ij}_l = \delta^{i \alpha} u^{l \alpha} + \delta^{i \alpha} u^{l \alpha}.
\]

Note that the introduction of the Riemannian metric \( g_R \) implies that \( W^{ij} \) are now dependent on the manifold coordinates \( x' \).

6. Numerical Experiments

We aim at visualizing most probable paths for the driving process and projections of curves satisfying the MPP Equation (13) in two cases: On 2D surfaces embedded in \( \mathbb{R}^3 \) and on finite dimensional landmark manifolds that arise from equipping a subset of the diffeomorphism group with
a right-invariant metric and letting the action descend to the landmarks by a left action. The surface examples are implemented in Python using the Theano [30] framework for symbolic operations, automatic differentiation, and numerical evaluation. The landmark equations are detailed below and implemented in Numpy using Numpy’s standard ODE integrators. The code for running the experiments is available at http://bitbucket.com/stefansommer/mpps/.

6.1. Embedded Surfaces

We visualize normal MPPs and projections of curves satisfying the MPP Equation (13) on surfaces embedded in \( \mathbb{R}^3 \) in three cases: The sphere \( S^2 \), on an ellipsoid, and on a hyperbolic surface. The surfaces are chosen in order to have both positive and negative curvature, and to have varying degree of symmetry. In all cases, an open subset of the surfaces are represented in a single chart by a map \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \). For the sphere and ellipsoid, this gives a representation of the surface, except for the south pole. The metric and Christoffel symbols are calculated using the symbolic differentiation features of Theano. The integration are performed by a simple Euler integrator.

Figures 4–6 show families of curves satisfying the MPP equations in three cases: (1) With fixed starting point \( x_0 \in M \) and initial velocity \( \dot{x}_0 \in T_xM \) but varying anisotropy represented by changing frame \( u \) in the fiber above \( x_0 \); (2) minimizing normal MPPs with fixed starting point and endpoint \( x_0, x_1 \in M \) but changing frame \( u \) above \( x_0 \); (3) fixed starting point \( x_0 \in M \) and frame \( u \) but varying \( V^*FM \) vertical part of the initial momentum \( \zeta_0 \in T^*FM \). The first and second cases thus show the effect of varying anisotropy, while the third case illustrates the effect of the “twist” that the \( \theta^2 \) degrees in the vertical momentum allows. Note the displayed anti-developed curves in \( \mathbb{R}^2 \) that for classical \( C \) geodesics would always be straight lines.

![Figure 4](Image)

**Figure 4.** Curves satisfying the MPP equations (top row) and corresponding anti-development (bottom row) on three surfaces embedded in \( \mathbb{R}^3 \): (a) An ellipsoid; (b) a sphere; (c) a hyperbolic surface. The family of curves is generated by rotating by \( \pi/2 \) radians the anisotropic covariance represented in the initial frame \( u_0 \) and displayed in the gray ellipse.
Figure 5. Minimizing normal MPPs between two fixed points (red/cyan). From isotropic covariance (top row, (a)) to anisotropic (top row, (c)) on $S^2$. Compare with minimizing Riemannian geodesic (black curve). The MPP travels longer in the directions of high variance. Families of curves (middle row, (d–f)) and corresponding anti-development (bottom row, (g–i)) on the three surfaces in Figure 4. The family of curves is generated by rotating the covariance matrix as in Figure 4. Notice how the varying anisotropy affects the resulting minimizing curves, and how the anti-developed curves end at different points in $\mathbb{R}^2$. 
Figure 6. (a–l) With the setup of Figures 4 and 5, generated families of curves by varying the vertical 
$V^*FM$ part of the initial momentum $\xi_0 \in T^*FM$ but keeping the base point and frame $u_0$ fixed. 
The vertical part allows varying degree of “twisting” of the curve.

6.2. LDDMM Landmark Equations

We here give an example of the MPP equations using the finite dimensional landmark manifolds 
that arise from right invariant metrics on subsets of the diffeomorphism group in the Large Deformation 
Diffeomorphic Metric Mapping (LDDMM) framework [8]. The LDDMM metric can be conveniently 
expressed as a cometric, and, using a rank-deficient inner product $g_{FM}$ as discussed in Section 5, we 
can obtain a reduction of the system of equations to $2(2N + 2Nk)$ compared to $2(2N + (2N)^2)$ with 
$N$ landmarks in $\mathbb{R}^2$.

Let $\{p_1, \ldots, p_N\}$ be landmarks in a subset $\Omega \subset \mathbb{R}^d$. The diffeomorphism group $\text{Diff}(\Omega)$ 
acts on the left on landmarks with the action $\varphi \cdot \{p_1, \ldots, p_N\} = \{\varphi(p_1), \ldots, \varphi(p_N)\}$. In LDDMM, 
a Hilbert space structure is imposed on a linear subspace $V$ of $L^2(\Omega, \mathbb{R}^d)$ using a self-adjoint operator 
$L : V \to V^* \subset L^2(\Omega, \mathbb{R}^d)$ and defining the inner product $\langle \cdot, \cdot \rangle_V$ by 

$$\langle v, w \rangle_V = \langle Lv, w \rangle_{L^2}.$$
The derivatives of the cometric is simply \( g \). Vertical direction, respectively, allows the minimizing normal MPP to vary more in these directions because the anisotropically-weighted metric penalizes high-covariance directions less.

Let \( M \) be the manifold \( M = \{(p_1^1, \ldots, p_d^1), \ldots, (p_n^1, \ldots, p_d^n)\} \subseteq \mathbb{R}^d \). The LDDMM metric on the landmark manifold \( M \) is directly related to the kernel \( K \) when written as a cometric \( g_p(\xi, \eta) = \sum_{i,j=1}^n \xi^i K(p_i, p_j) \eta^j \). Letting \( i^k \) denote the index of the \( k \)th component of the \( i \)th landmark, the cometric is in coordinates \( g_p^{i,i} = K(p_i, p_j) \). The Christoffel symbols can be written in terms of derivatives of the cometric \( g^{ij} \) \[ \Gamma_{ij}^k = \frac{1}{2} \partial_a \left( g^{ik} \partial_a g^{kj} - g^{aj} \partial_a g^{ik} \right) \] (18).

This relation comes from the fact that \( g_{i,m,k} = -g_p^{rs} g_{ij} g_{kr} \) gives the derivative of the metric. The derivatives of the cometric is simply \( g^{i,i} = \left( \delta_i^a + \delta_i^r \right) \partial_a K(p_i, p_j) \). Using (18), derivatives of the Christoffel symbols can be computed

\[
\Gamma_{ij}^k = \frac{1}{2} \delta_{ir} \left( g^{kl} g_{ls}^{rs} - g^{sl} g_{rs}^{lk} - g^{rl} g_{rs}^{jk} \right) g_{sj} + \frac{1}{2} \delta_{ir} \left( g^{kl} g_{ls}^{rs} - g^{sl} g_{rs}^{lk} - g^{rl} g_{rs}^{jk} \right) g_{sj}.
\]

This provides the full data for numerical integration of the evolution equations on \( T^2 \).

In Figure 7 (top row), we plot minimizing normal MPPs on the landmark manifold with two landarks and varying covariance in the \( \mathbb{R}^2 \) horizontal and vertical direction. The plot shows the landmark equivalent of the experiment in Figure 5. Note how adding covariance in the horizontal and vertical direction, respectively, allows the minimizing normal MPP to vary more in these directions because the anisotropically-weighted metric penalizes high-covariance directions less.

![Figure 7](image-url)

Figure 7. (Top row) Matching of two landmarks (green) to two landmarks (red) by (a) computing a minimizing Riemannian geodesic on the landmark manifold, and (b-e) minimizing MPPs with added covariance (arrows) in \( \mathbb{R}^2 \) horizontal direction (b,c) and vertical (d,e). The action of the corresponding diffeomorphisms on a regular grid is visualized by the deformed grid which is colored by the warp strain. The added covariance allows the paths to have more movement in the horizontal and vertical direction, respectively, because the anisotropically weighted metric penalizes high-covariance directions less. (bottom row, (f-j)) Five landmark trajectories with fixed initial velocity and anisotropic covariance but varying \( V^0 FM \) vertical initial momentum \( \xi_0 \). Changing the vertical momentum “twists” the paths.
Figure 7 (bottom row) shows five curves satisfying the MPP equations with varying vertical \( V^* \) initial momentum similarly to the plots in Figure 6. Again, we see how the extra degrees of freedom allows the paths to twist, generating a higher-dimensional family than classical geodesics with respect to \( C \).

7. Discussion and Concluding Remarks

Incorporating anisotropy in models for data in non-linear spaces via the frame bundle as pursued in this paper leads to a sub-Riemannian structure and metric. A direct implication is that most probable paths to observed data in the sense of sequences of stochastic steps of a driving semi-martingale are not related to geodesics in the classical sense. Instead, a best estimate of the sequence of steps \( w_t \in \mathbb{R}^d \) that leads to an observation \( x = \varphi_u(w_t) \mid t=1 \) is an MPP in the sense of Definition 1. As shown in the paper, these paths are generally not geodesics or polynomials with respect to the connection on the manifold. In particular, if \( M \) has a Riemannian structure, the MPPs are generally neither Riemannian geodesics nor Riemannian polynomials. Below, we discuss the statistical implications of this result.

7.1. Statistical Estimators

Metric distances and Riemannian geodesics have been the traditional vehicle for representing observed data in non-linear spaces. Most fundamentally, the sample Frechét mean

\[
\hat{x} = \arg\min_{x \in M} \sum_{i=1}^{N} g_R(x, x_i)^2
\]

of observed data \( x_1, \ldots, x_N \in M \) relies crucially on the Riemannian distance \( g_R \) connected to the metric \( g_R \). Many PCA constructs (e.g., Principal Geodesics Analysis [6]) use the Riemannian Exp. and Log maps to map between linear tangent spaces and the manifold. These maps are defined from the Riemannian metric and Riemannian geodesics. Distributions modelled as in the random orbit model [32] or Bayesian models [15,33] again rely on geodesics with random initial conditions.

Using the frame bundle sub-Riemannian metric \( g_{FM} \), we can define an estimator analogous to the Riemannian Frechét mean estimator. Assuming the covariance is a priori known, the estimator

\[
\hat{x} = \arg\min_{u \in s(M)} \sum_{i=1}^{N} d_{FM}(u, \pi^{-1}(x_i))^2
\]

acts correspondingly to the Frechét mean estimator (19). Here \( s \in \Gamma(FM) \) is a (local) section of \( FM \) that to \( x \in M \) connects the known covariance represented by \( s(x) \in FM \). The distances \( d_{FM}(u, \pi^{-1}(x_i)) \), \( u = s(x) \) are realized by MPPs from the mean candidate \( x \) to the fibers \( \pi^{-1}(x_i) \). The Frechét mean problem is thus lifted to the frame bundle with the anisotropic weighting incorporated in the metric \( g_{FM} \). This metric is not related to \( g_R \), except for its dependence on the connection \( C \) that can be defined as the Levi–Civita connection of \( g_R \). The fundamental role of the distance \( d_{FM} \) and \( g_R \) geodesics in (19) is thus removed.

Because covariance is an integral part of the model, sample covariance can also be estimated directly along with the sample mean. In [3], the estimator

\[
\hat{u} = \arg\min_{u \in FM} \sum_{i=1}^{N} d_{FM}(u, \pi^{-1}(x_i))^2 - N \log(\det g_R u)
\]

is suggested. The normalizing term \(-N \log(\det g_R u)\) is derived such that the estimator exactly corresponds to the maximum likelihood estimator of mean and covariance for Euclidean Gaussian distributions. The determinant is defined via \( g_R \), and the term acts to prevent the covariance from approaching infinity. Maximum likelihood estimators of mean and covariance for normally distributed Euclidean data have unique solutions in the sample mean and sample covariance matrix, respectively.
Uniqueness of the Frechét mean (19) is only ensured for sufficiently concentrated data. For the estimator (21), existence and uniqueness properties are not immediate, and more work is needed in order to find necessary and sufficient conditions.

7.2. Priors and Low-Rank Estimation

The low-rank cometric formulation pursued in Section 5 gives a natural restriction of (21) to $u \in I^k M$, $1 \leq k \leq d$. As for Euclidean PCA, most variance is often captured in the span of the first $k$ eigenvectors with $k \ll d$. Estimates of the remaining eigenvectors are generally ignored, as the variance of the eigenvector estimates increases as the noise captured in the span of the last eigenvectors becomes increasingly uniform. The low-rank cometric restricts the estimation to only the first $k$ eigenvectors, and thus builds the construction directly into the model. In addition, it makes numerical implementation feasible, because a numerical representation need only store and evolve $d \times k$ matrices. As a different approach for regularizing the estimator (21), the normalizing term $-N \log(\det g_u)$ can be extended with other priors (e.g., an $L^1$-type penalizing term). Such priors can potentially partly remove existence and uniqueness issues, and result in additional sparsity properties that can benefit numerical implementations. The effects of such priors have yet to be investigated.

In the $k = d$ case, the number of degrees of freedom for the MPPs grows quadratically in the dimension $d$. This naturally increases the variance of any MPP estimate given only one sample from its trajectory. The low-rank cometric formulation reduces the growth to linear in $d$. The number of degrees of freedom is however still $k$ times larger than for Riemannian geodesics. With longitudinal data, more samples per trajectory can be obtained, reducing the variance and allowing a better estimate of the MPP. However, for the estimators (20) and (21) above, estimates of the actual optimal MPPs are not needed—only their squared length. It can be hypothesized that the variance of the length estimates is lower than the variance of the estimates of the corresponding MPPs. Further investigation regarding this will be the subject of future work.

7.3. Conclusions

The underlying model of anisotropy used in this paper originates from the anisotropic normal distributions formulated in [2] and the diffusion PCA framework [1]. Because many statistical models are defined using normal distributions, this approach to incorporating anisotropy extends to models such as linear regression. We expect that finding most probable paths in other statistical models such as regressions models can be carried out with a program similar to the program presented in this paper.

The difference between MPPs and geodesics shows that the geometric and metric properties of geodesics, zero acceleration, and local distance minimization are not directly related to statistical properties such as maximizing path probability. Whereas the concrete application and model determines if metric or statistical properties are fundamental, most statistical models are formulated without referring to metric properties of the underlying space. It can therefore be argued that the direct incorporation of anisotropy and the resulting MPPs are natural in the context of many models of data variation in non-linear spaces.

Acknowledgments: The author wishes to thank Peter W. Michor and Sarang Joshi for suggestions for the geometric interpretation of the sub-Riemannian metric on $FM$ and discussions on diffusion processes on manifolds. The work was supported by the Danish Council for Independent Research, the CSGB Centre for Stochastic Geometry and Advanced Bioimaging funded by a grant from the Villum foundation, and the Erwin Schrödinger Institute in Vienna.

Conflicts of Interest: The author declares no conflict of interest.

References


© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).