Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 04–31

FITTING SMOOTH PATHS ON RIEMANNIAN MANIFOLDS

LUÍS MIGUEL MACHADO AND F. SILVA LEITE

Abstract: In this paper we formulate a least squares problem on a Riemannian manifold M, in order to generate smoothing spline curves fitting a given data set of points in M, $q_0, q_1, \ldots, q_N$, at given instants of time $t_0 < t_1 < \cdots < t_N$.

Using tools from Riemannian geometry, we derive the Euler-Lagrange equations associated to this variational problem and prove that its solutions are Riemannian cubic polynomials defined at each interval $[t_i, t_{i+1}[$, $i = 0, \ldots, N-1$, and satisfying some smoothing constraints at the knot points t_i. The geodesic that best fits the data, arises as a limiting process of the above.

When M is replaced by the Euclidean space $\mathbb{R}^n$, the proposed problem has a unique solution which is a natural cubic spline given explicitly in terms of the data. We prove that, in this case, the straight line obtained from the limiting process is precisely the linear regression line associated to the data.

Using tools from optimization on Riemannian manifolds we also present a direct procedure to generate geodesics fitting a given data set of time labelled points for the particular cases when M is the Lie group $SO(n)$ and the unitary $n-$sphere S^n.

Keywords: Covariant differentiation, curvature tensor, geodesics, geodesic distance, Riemannian cubic polynomials, normal equations.

1. Introduction

The most primitive and important class of functions for the purpose of fitting curves to data in Euclidean spaces is indeed the class of polynomial curves. Polynomial interpolation is the most elementary notion of curve fitting. We refer to Lancaster and Salkauskas [14] for an overview of the curve fitting problem.

The classical least squares method, introduced by Lagrange (1736-1813), can also be seen as a typical method for fitting curves. Here we are given a finite set of points in $\mathbb{R}^n$, $q_0, q_1, \ldots, q_N$, and a sequence of instants of time $t_0 < t_1 < \cdots < t_N$ and the objective is to find, in the class $\mathcal{P}_m$ of polynomial functions of degree not exceeding m $(m \leq N)$, a polynomial that best fits

Received October 29, 2004.

Work supported in part by ISR and research network contract HPMT-CT-2001-00278. Support for the first author was also provided by PRODEP 5.3 program (UE-FSE).

the given data, in the sense that the functional

$$E(\gamma) = \sum_{i=0}^{N} d^2 \left(q_i, \gamma(t_i)\right),$$

should be as small as possible, where d denotes Euclidean distance.

It is well known ([14]) that for each $m \in \mathbb{N}$, with $m \leq N$, there exists a unique polynomial in $\mathcal{P}_m$ minimizing E, whose $m+1$ coefficients can be obtained by solving a linear system of equations known in the literature as the "normal equations".

The progression from polynomial to piecewise polynomial gave rise to spline methods for curve fitting. These more advanced methods turned out to be more flexible and powerful than the classical methods.

Interpolating polynomial splines in Euclidean spaces appeared in the 1940's as the solution of some optimization problems. The most common are the cubic splines, which minimize changes in velocity and are, for that reason, particularly useful in applications. Other interpolating splines (generalized splines, L-splines,...) have been introduced and their optimal properties have been studied.

Splines on Riemannian manifolds have received much attention during the last 15 years, since the pioneer work of Noakes et al. [17]. Other relevant works, where splines are also seen as solutions of a variational problem, include Crouch and Silva Leite [6, 7]), Camarinha [2] and Krakowski [13]. Contrary to these contributions on interpolating splines, our concentration goes to spline curves that best approximate a given data on a Riemannian manifold. This might be more realistic in situations where the data is corrupted by noise and frequently also results in a significant decrease in the cost. This is also of particular importance for problems arising, for instance, in physics and engineering. Trajectory planning in aeronautics, robotics, and biomechanics are some areas which motivate the study of least squares problems on Riemannian manifolds, since the configuration spaces of most mechanical systems have components which are particular manifolds, such is the case of Lie groups or symmetric spaces. The case when the manifold is the 2-dimensional sphere, has been extensively studied in the early 80's (see, for instance, Jupp and Kent [11] and references therein). In particular, Jupp and Kent defined smoothing spherical splines and presented a method for computing them, which is based on a reduction of the problem to the plane tangent to S^2, at a particular point, via convenient diffeomorphisms. Splines fitting data on general Riemannian manifolds is the subject of this article.

The paper is organized as follows. In section 2, we revisit the classical Euclidean least squares method and present an alternative approach using techniques from optimization on Riemannian manifolds. The linear regression line appears as a particular case.

For Riemannian manifolds where explicit formulas for geodesics are available, the least squares problem for the linear regression line can easily be generalized. Such is the case of the Lie group $SO(n)$ and the unitary $n-$sphere S^n. In section 3 we formulate the optimization problems corresponding to such generalization, and derive the counterpart of the "normal equations", which are necessary conditions for the best fitting geodesic.

In section 4 we formulate the least squares problem corresponding to fitting by cubic splines on Riemannian manifolds. The situation is now more complex since no explicit forms for cubic polynomials on a general Riemannian manifold M are known. However, these curves have been defined as critical points for the functional

$$L(\gamma) = \int_0^1 \langle \frac{D^2\gamma}{dt^2}, \frac{D^2\gamma}{dt^2} \rangle \, dt,$$

defined over the class of twice continuously differentiable curves in M, where $\langle \cdot, \cdot \rangle$ is the Riemannian metric and $\frac{D^2\gamma}{dt^2}$ denotes covariant acceleration along the curve $t \mapsto \gamma(t)$. So, the corresponding least squares problem can be formulated as: "Find a curve on M which minimizes the functional

$$J(\lambda) = E(\gamma) + \lambda L(\gamma),$$

where $\lambda > 0$ is a smoothing parameter and the distance defining E is the geodesic distance on M".

We derive necessary optimality conditions for the functional J and prove that when λ converges to $+\infty$ the curve reduces to a geodesic fitting the data. For the particular case when $M = \mathbb{R}^n$, this limiting process produces the linear regression line.

2. The Euclidean least squares method revisited

The least squares method (see, for instance, Lancaster and Salkauskas [14]), is a classical example of fitting curves to data in Euclidean spaces. In this method we are given a finite sequence of distinct points in $\mathbb{R}^n$,

$$q_0, q_1, \ldots, q_N, \tag{1}$$

and a sequence of instants of time

$$t_0 < t_1 < \cdots < t_N, \tag{2}$$

and the objective is to find a polynomial function $t \longmapsto x(t)$, of degree not exceeding m, with $m \leq N$, that best fits the given data in the sense that the functional

$$E(x) = \frac{1}{2}\sum_{i=0}^{N} \left[d\left(x(t_i), q_i\right)\right]^2, \tag{3}$$

should be as small as possible, where d denotes the Euclidean distance.We denote by $\mathcal{P}_m$ the family of polynomial functions $t \longmapsto x(t) \in \mathbb{R}^n$, with degree less than or equal to m (and assume that $m \leq N$).

Since $t \longmapsto (t-t_0)/(t_N-t_0)$ defines a bijection between the intervals $[t_0, t_N]$ and $[0, 1]$, from now on we also assume that the instants of time (2) form a partition of the interval $[0, 1]$.

Most literature solves that problem for data in $\mathbb{R}$, but the approach is easily adapted to the Euclidean space $\mathbb{R}^n$. Here we show that the least squares problem above can be reformulated as an optimization problem in the matrix space $\mathbb{R}^{(m+1)\times n}$, equipped with the Euclidean inner product

$$\langle X, Y\rangle = \mathrm{tr}(X^\top Y).$$

For that, we first have to introduce some notation.

A point $y \in \mathbb{R}^n$ will be denoted by

$$y = \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{bmatrix}. \tag{4}$$

If y_j denotes the value of x at the instant of time t_j, the functional defined by (3) may be written as

$$E(x) = \frac{1}{2}\sum_{j=0}^{N} \|y_j - q_j\|^2 = \frac{1}{2}\sum_{j=0}^{N}\sum_{k=1}^{n} \left(y_j^k - q_j^k\right)^2.$$

We are looking for a polynomial function

$$t \longmapsto x(t) = a_0 + a_1 t + \cdots + a_m t^m, \ \text{ with } a_i \in \mathbb{R}^n, \forall i = 0, \ldots, m,$$

which minimizes the functional (3). The unknown coefficients form the columns of the following matrix

$$X^\top = \begin{bmatrix} a_0^1 & a_1^1 & a_2^1 & \ldots & a_m^1 \\ a_0^2 & a_1^2 & a_2^2 & \ldots & a_m^2 \\ & & & \vdots & \\ a_0^n & a_1^n & a_2^n & \ldots & a_m^n \end{bmatrix} \in \mathbb{R}^{n\times(m+1)},$$

and the data may be used to define the matrices

$$V^\top = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ t_0 & t_1 & t_2 & \ldots & t_N \\ & & & \vdots & \\ t_0^m & t_1^m & t_2^m & \ldots & t_N^m \end{bmatrix} \in \mathbb{R}^{(m+1)\times(N+1)}$$

and

$$P^\top = \begin{bmatrix} q_0^1 & q_1^1 & q_2^1 & \ldots & q_N^1 \\ q_0^2 & q_1^2 & q_2^2 & \ldots & q_N^2 \\ & & & \vdots & \\ q_0^n & q_1^n & q_2^n & \ldots & q_N^n \end{bmatrix} \in \mathbb{R}^{n\times(N+1)}.$$

After some trivial matrix computations and properties of the trace, it follows that

$$\begin{aligned} E(x) &= \tfrac{1}{2} \sum_{j=0}^{N} \sum_{k=1}^{n} \left(y_j^k - q_j^k\right)^2 \\ &= \tfrac{1}{2} \operatorname{tr}\left(\left(X^\top V^\top - P^\top\right)^\top \left(X^\top V^\top - P^\top\right) \right) \\ &= \tfrac{1}{2} \operatorname{tr}\left(V^\top V X X^\top - 2V^\top P X^\top\right). \end{aligned}$$

Therefore, with the above notations, the classical least squares problem in the Euclidean space $\mathbb{R}^n$ can be reformulated as the following optimization problem in the Riemannian manifold $\mathbb{R}^{(m+1)\times n}$:

$$(\mathcal{P}_1) \qquad \min_{X \in \mathbb{R}^{(m+1)\times n}} F(X) = \frac{1}{2} \operatorname{tr}\left(V^\top V X X^\top - 2V^\top P X^\top\right). \tag{5}$$

We can now apply techniques of optimization of functions defined on Riemannian manifolds to prove the following result. For more details about optimization on Riemannian manifolds, we refer to Helmke and Moore [9] and Udriste [22].

Theorem 2.1. *If $q_0, q_1, \ldots, q_N$ are distinct points given in $\mathbb{R}^n$ and $0 = t_0 < t_1 < \cdots < t_N = 1$ is a given partition of the real interval $[0,1]$, then there exists a unique polynomial $x \in \mathcal{P}_m$, with $m \leq N$, that minimizes the functional E defined by (3). The matrix X whose rows are the coefficients of that polynomial is given by $X = \left(V^\top V\right)^{-1} V^\top P$. Moreover, if $m = N$, the polynomial curve $x \in \mathcal{P}_m$ that minimizes the functional E, interpolates the given data set of points at the given instants of time.*

Proof: Since minimizing the functional E over $\mathcal{P}_m$ is equivalent to minimizing the functional F over $\mathbb{R}^{(m+1)\times n}$, we first find the critical points of the latter, that is, the points $X \in \mathbb{R}^{(m+1)\times n}$ such that

$$T_X F(W) = 0, \ \forall W \in \mathbb{R}^{(m+1)\times n},$$

where $T_X F$ denotes the tangent map of F at the point X. But

$$T_X F(W) = \left.\frac{d}{ds}\right|_{s=0} F(\Gamma(s)),$$

where $s \mapsto \Gamma(s)$ is a curve in $\mathbb{R}^{(m+1)\times n}$ satisfying $\Gamma(0) = X$ and $\dot{\Gamma}(0) = W$. We may take $\Gamma(s) = X + sW$. So,

$$\begin{aligned} T_X F(W) &= \frac{1}{2} \operatorname{tr}\left(V^\top V W X^\top + V^\top V X W^\top - 2V^\top P W^\top\right) \\ &= \frac{1}{2} \operatorname{tr}\left(\left(X^\top V^\top V + X^\top V^\top V - 2P^\top V\right) W\right) \\ &= \operatorname{tr}\left(\left(V^\top V X - V^\top P\right)^\top W\right). \end{aligned}$$

Consequently, X is a critical point of the functional F if and only if

$$V^\top V X = V^\top P. \tag{6}$$

Since the instants of time t_i are all distinct, the matrix V has full rank. Hence $V^\top V$ is symmetric and positive definite, so, there exists a unique critical point of F, given by

$$X = \left(V^\top V\right)^{-1} V^\top P.$$

Clearly, F takes its minimal value at this critical point. Now, to prove the last part of the statement, notice that, when $m = N$, V is a nonsingular Vandermonde matrix of order $N+1$ and, therefore, the equation (6) reduces to $VX = P$ or, equivalently to $X^\top V^\top = P^\top$. Now, a trivial computation shows that, for $i = 0, \ldots, N$, the $(i+1)^{th}$ column of the matrix $X^\top V^\top$ is equal to $x(t_i)$, while the corresponding column of $P^\top$ is equal to q_i. Therefore $x(t_i) = q_i$, $\forall i = 0, \ldots, N$, and the proof is complete. ■

Remark 2.1. The linear system of equations (6) is known as the <u>normal equations</u> associated to the least squares method in $\mathbb{R}^n$.

As a particular case of the above, one obtains the straight line that best fits the given data (1) at the given instants of time (2). This is known as the linear regression line and corresponds to the case $m = 1$.

Theorem 2.2. *The straight line $t \mapsto x(t)$ in $\mathbb{R}^n$ that best fits the given data set of points (1) at the given instants of time (2) is unique and given explicitly by*

$$x(t) = \frac{\displaystyle\sum_{i=0}^{N} t_i^2 \sum_{i=0}^{N} q_i - \sum_{i=0}^{N} t_i \sum_{i=0}^{N} t_i q_i}{\displaystyle (N+1)\sum_{i=0}^{N} t_i^2 - \left(\sum_{i=0}^{N} t_i\right)^2} + \frac{\displaystyle (N+1)\sum_{i=0}^{N} t_i q_i - \sum_{i=0}^{N} t_i \sum_{i=0}^{N} q_i}{\displaystyle (N+1)\sum_{i=0}^{N} t_i^2 - \left(\sum_{i=0}^{N} t_i\right)^2}\, t. \quad (7)$$

Proof: We just need to solve the normal equations (6), when $m = 1$. In this case,

$$\begin{aligned}
(V^\top V)^{-1} &= \frac{1}{\displaystyle (N+1)\sum_{i=0}^{N} t_i^2 - \left(\sum_{i=0}^{N} t_i\right)^2} \begin{bmatrix} \displaystyle\sum_{i=0}^{N} t_i^2 & \displaystyle -\sum_{i=0}^{N} t_i \\ \displaystyle -\sum_{i=0}^{N} t_i & N+1 \end{bmatrix}, \\
V^\top P &= \begin{bmatrix} \displaystyle\sum_{i=0}^{N} q_i \\ \displaystyle\sum_{i=0}^{N} t_i q_i \end{bmatrix},
\end{aligned}$$

and the result follows straightforward. ■

Remark 2.2. Let $\bar{q}$, $\bar{t}$ and $\bar{\bar{q}}$, $\bar{\bar{t}}$ be defined by

$$\bar{q} = \frac{1}{N+1}\sum_{i=0}^{N} q_i, \qquad \bar{t} = \frac{1}{N+1}\sum_{i=0}^{N} t_i,$$
$$\bar{\bar{q}} = \frac{1}{\sum_{i=0}^{N} t_i}\sum_{i=0}^{N} t_i q_i, \qquad \bar{\bar{t}} = \frac{1}{\sum_{i=0}^{N} t_i}\sum_{i=0}^{N} t_i^2.$$

We note that $\bar{\bar{q}}$ is the center of mass of $N+1$ points $q_i \in \mathbb{R}^n$, having attached the masses t_i and, similarly, $\bar{\bar{t}}$ is the center of mass of $N+1$ points $t_i \in \mathbb{R}$ having attached the masses t_i. When all masses are equal, $\bar{\bar{q}}$ and $\bar{\bar{t}}$ reduce to $\bar{q}$ and $\bar{t}$ respectively.

Now, a simple computation shows that if $t \mapsto x(t)$ is the geodesic given in the previous theorem, then $x(\bar{t}) = \bar{q}$ and $x(\bar{\bar{t}}) = \bar{\bar{q}}$, which has an interesting geometric interpretation in terms of center of masses.

3. Fitting geodesics to data on Riemannian manifolds

The techniques used in the previous section to generate the straight line that best fits a given data in Euclidean spaces, may be generalized to Riemannian manifolds. In this context, the Riemannian metric plays the role of the Euclidean metric, the straight line is replaced by a geodesic and the distance defining the functional (3) is now the geodesic distance. This requires that one knows explicit formulas for geodesics. Such is the case for the rotation group $SO(n)$, where geodesics are either one-parameter subgroups or their translations, and for the unit sphere S^n, where geodesics are great circles. The problem of best fitting a geodesic to data, for these two particular Riemannian manifolds, are studied next.

3.1. Fitting geodesics to data on $SO(n)$

Before the statement of the corresponding optimization problem, we introduce some notation and recall some results which will be necessary later on. For more details concerning the theory of Lie groups we refer to [8].

Let $GL(n, \mathbb{R})$ be the set of all real $n \times n$ invertible matrices. Then,

$$SO(n) = \{\Theta \in GL(n, \mathbb{R}) : \Theta^\top \Theta = I \text{ and } \det\Theta = 1\}.$$

Since $SO(n)$ is a compact and connected Lie group, geodesics on $SO(n)$ are translations of one parameter subgroups. They are charaterized by a point in $SO(n)$ and a vector tangent to $SO(n)$ at the identity I. That is, by a rotation matrix and a matrix belonging to $so(n)$, the Lie algebra of $SO(n)$, that consists of all real $n \times n$ skew-symmetric matrices.

It is well known (see, for instance, Horn and Johnson [10]) that logarithms of an invertible matrix B are the solutions of the matrix equation $\mathrm{e}^X = B$, and when B is real and doesn't have eigenvalues in the closed negative real line, there exists only one real logarithm of B whose spectrum lies in the infinite horizontal strip $\{z \in \mathbb{C} : -\pi < Im(z) < \pi\}$ of the complex plane. From now on we will consider this logarithm only, and will denote it by $\log B$. A very useful property is that, for every non-singular matrix C,

$$C^{-1} \log(B) C = \log\left(C^{-1} B C\right).$$

For $\alpha \in \mathbb{R}$ and B a non-singular matrix not having eigenvalues in the closed negative real line, we define B^α as being the nonsingular matrix

$$B^\alpha = \mathrm{e}^{\alpha \log B}. \tag{8}$$

Geometrically, B^α is the point, corresponding to $t = \alpha$, on the geodesic that passes through the identity (at $t = 0$) with initial velocity $\log B$.

If B belongs to the Lie group $SO(n)$, then $\log(B)$ belongs to its Lie algebra $so(n)$.

We now state a few results that will be used later on.

Lemma 3.1. *(Moakher [16]) Let $B(t)$ be a differentiable matrix valued function and assume that, for each t in the domain, $B(t)$ is a non-singular matrix not having eigenvalues in the closed negative real line. Then,*

$$\frac{d}{dt} \operatorname{tr}\left(\log^2 B(t)\right) = 2 \operatorname{tr}\left(\log(B(t)) B^{-1}(t) \frac{d}{dt} B(t)\right). \tag{9}$$

Lemma 3.2. *(Sattinger and Weaver [20]) Let $X(t)$ be a differentiable matrix valued function. Then,*

$$\frac{d}{dt} \mathrm{e}^{X(t)} = f\left(\mathrm{ad}_{X(t)}\right)\left(\dot{X}(t)\right) \mathrm{e}^{X(t)},$$

where $f(z) = \dfrac{\mathrm{e}^z - 1}{z}$ *stands for the sum of the series* $\displaystyle\sum_{k=0}^{+\infty} \frac{z^k}{(k+1)!}$, *and* ad *denotes the adjoint operator defined by* $\mathrm{ad}_X(Y) = [X, Y] = XY - YX$.

The following result, can be easily proved and will be very important for the derivation of the results presented along this section.

Lemma 3.3. $\operatorname{tr}(Yf(\mathrm{ad}_X)(Z)) = \operatorname{tr}(f(-\mathrm{ad}_X)(Y)Z)$, *for all* $n \times n$ *matrices* X, Y *and* Z.

Now, assume that we are given a finite set of time labelled points in $SO(n)$

$$\mathcal{Q} = \{(t_i, Q_i)\ : i = 0, \ldots, N\}, \tag{10}$$

where $Q_i \in SO(n)$, $i = 0, \ldots, N$, and $0 = t_0 < t_1 < \ldots < t_N = 1$ is a partition of the time interval $[0, 1]$, and the objective is to find the geodesic that best fits this data.

Since geodesics on $SO(n)$ can be parameterized explicitly as

$$\begin{array}{rccl} \gamma : & \mathbb{R} & \longrightarrow & SO(n) \\ & t & \longmapsto & \gamma(t) = \Theta \mathrm{e}^{tX}, \end{array} \tag{11}$$

where $\Theta \in SO(n)$ and $X \in so(n)$, the corresponding optimization problem may be formulated as:

$$(\mathcal{P}_2) \qquad \min_{(\Theta, X) \in SO(n) \times so(n)} \frac{1}{2} \sum_{i=0}^{N} d^2\left(Q_i, \Theta \mathrm{e}^{t_i X}\right), \tag{12}$$

where d denotes geodesic distance. Geodesic distance between 2 points in $SO(n)$ is the length of the shortest geodesic curve joining them. So, since the shortest geodesic curve joining Q_i to $\Theta \mathrm{e}^{t_i X}$ can be parameterized explicitly by

$$c(s) = Q_i \mathrm{e}^{s \log\left(Q_i^\top \Theta \mathrm{e}^{t_i X}\right)}, \ \ s \in [0, 1],$$

it follows that

$$d^2\left(Q_i, \Theta \mathrm{e}^{t_i X}\right) = -\operatorname{tr}\left(\log^2\left(Q_i^\top \Theta \mathrm{e}^{t_i X}\right)\right).$$

Consequently, a necessary condition for (Θ, X) to be a solution for the optimization problem $(\mathcal{P}_2)$ is that (Θ, X) is a critical point of the following function:

$$\begin{array}{rccl} F : & SO(n) \times so(n) & \longrightarrow & \mathbb{R} \\ & (\Theta, X) & \longmapsto & F(\Theta, X) = -\dfrac{1}{2} \displaystyle\sum_{i=0}^{N} \operatorname{tr}\left(\log^2\left(Q_i^\top \Theta \mathrm{e}^{t_i X}\right)\right). \end{array}$$

Now, to solve the optimization problem $(\mathcal{P}_2)$, we first have to compute the first variation of F.

The tangent space of $SO(n) \times so(n)$ at the point (Θ, X) is given by

$$T_{(\Theta, X)}\left(SO(n) \times so(n)\right) = \{(\Theta Y, Z)\ :\ Y,\, Z \in so(n)\}. \tag{13}$$

Therefore, we can endow the manifold $SO(n) \times so(n)$ with a Riemannian metric, by defining the following inner product on each tangent space $T_{(\Theta,X)}\left(SO(n) \times so(n)\right)$:

$$\ll (\Theta Y_1, Z_1), (\Theta Y_2, Z_2) \gg = \operatorname{tr}\left(Y_1^\top Y_2\right) + \operatorname{tr}\left(Z_1^\top Z_2\right), \tag{14}$$

where Y_1, Y_2, Z_1, $Z_2 \in so(n)$. (We note that $\langle X, Y\rangle = \operatorname{tr}(X^\top Y) = -\operatorname{tr}(XY)$ is the Frobenius inner product on $so(n)$.)

By definition, the tangent map of F at a point $(\Theta, X) \in SO(n) \times so(n)$ is given by

$$T_{(\Theta,X)}F(\Theta Y, Z) = \left.\frac{d}{ds}\right|_{s=0} F(\alpha(s)),$$

where α is any smooth curve on $SO(n) \times so(n)$, satisfying the initial conditions

$$\alpha(0) = (\Theta, X), \quad \dot{\alpha}(0) = (\Theta Y, Z). \tag{15}$$

In particular,

$$\begin{array}{rccl} \alpha: & \mathbb{R} & \longrightarrow & SO(n) \times so(n) \\ & s & \longmapsto & \alpha(s) = \left(\Theta e^{sY}, X + sZ\right), \end{array}$$

fulfills the initial conditions (15).

Then, attending to lemmas 3.1, 3.2 and 3.3, we can write successively the following identities

$$\begin{aligned} & T_{(\Theta,X)}F(\Theta Y, Z) \\ = \; & -\frac{1}{2}\sum_{i=0}^{N} \left.\frac{d}{ds}\right|_{s=0} \operatorname{tr}\left(\log^2\left(Q_i^\top \Theta e^{sY} e^{t_i(X+sZ)}\right)\right) \\ = \; & -\sum_{i=0}^{N} \operatorname{tr}\left(\log\left(Q_i^\top \Theta e^{t_i X}\right) e^{-t_i X}\Theta^\top Q_i \left.\frac{d}{ds}\right|_{s=0}\left(Q_i^\top \Theta e^{sY} e^{t_i(X+sZ)}\right)\right) \\ = \; & -\sum_{i=0}^{N} \operatorname{tr}\left(\log\left(Q_i^\top \Theta e^{t_i X}\right) e^{-t_i X}\Theta^\top Q_i \left(Q_i^\top \Theta Y e^{t_i X} + Q_i^\top \Theta f\left(\operatorname{ad}_{t_i X}\right)\left(t_i Z\right) e^{t_i X}\right)\right) \\ = \; & -\sum_{i=0}^{N} \operatorname{tr}\left(\log\left(e^{t_i X} Q_i^\top \Theta\right) Y + \log\left(e^{t_i X} Q_i^\top \Theta\right) f\left(\operatorname{ad}_{t_i X}\right)\left(t_i Z\right)\right) \\ = \; & -\sum_{i=0}^{N} \operatorname{tr}\left(\log\left(e^{t_i X} Q_i^\top \Theta\right) Y + t_i f\left(-\operatorname{ad}_{t_i X}\right)\left(\log\left(e^{t_i X} Q_i^\top \Theta\right)\right) Z\right) \end{aligned}$$

$$= \langle \sum_{i=0}^{N} \log\left(\mathrm{e}^{t_i X} Q_i^\top \Theta\right), Y\rangle + \langle \sum_{i=0}^{N} t_i f\left(-\mathrm{ad}_{t_i X}\right)\left(\log\left(\mathrm{e}^{t_i X} Q_i^\top \Theta\right)\right), Z\rangle. \tag{16}$$

We are now in conditions to formulate the main theorem of this section, which contains the counterpart to the "normal equations" derived in the last section.

Theorem 3.4. *A necessary condition for the geodesic $t \mapsto \gamma(t) = \Theta \mathrm{e}^{tX}$ to be a solution of the optimization problem $(\mathcal{P}_2)$, is that the pair $(\Theta, X) \in SO(n) \times so(n)$ satisfies the following set of equations:*

$$\begin{cases} \displaystyle\sum_{i=0}^{N} \log\left(\mathrm{e}^{t_i X} Q_i^\top \Theta\right) = 0 \\ \displaystyle\sum_{i=0}^{N} t_i f\left(-\mathrm{ad}_{t_i X}\right)\left(\log\left(\mathrm{e}^{t_i X} Q_i^\top \Theta\right)\right) = 0 \end{cases}. \tag{17}$$

Proof: By definition, $(\Theta, X) \in SO(n) \times so(n)$ is a critical point for F if

$$T_{(\Theta,X)} F(\Theta Y, Z) = 0, \ \forall Y, Z \in so(n).$$

The result is now a consequence of the non-degeneracy of the Frobenius inner product in $so(n)$ and the above considerations. ■

Contrary to the Euclidean case, no explicit solutions of the normal equations (17) have been found, except for the situation when $N = 1$. Indeed, in this case $t_0 = 0$, $t_1 = 1$ and, as expected, the geodesic that joins Q_0 (at $t = 0$) to Q_1 (at $t = 1$) and given by $\gamma(t) = Q_0 \mathrm{e}^{t \log(Q_0^\top Q_1)}$ satisfies equations (17), and also minimizes the functional (12), since in this case $d^2(Q_i, \gamma(t_i)) = 0$, $i = 0, 1$.

At this point, a natural question is to know whether or not the geodesic that best fits the given data in $SO(n) \times [0, 1]$ passes through the analogue of the center of mass. We believe that the answer is no, although when the given points lie in a connected, compact and abelian subgroup G of $SO(n)$, the geodesic in G that best fits the data satisfies similar properties to those mentioned in remark 2.2, as we now explain, after some considerations about the analogue of the center of mass.

The concept of center of mass, of a finite set of points in a manifold with equal masses, has been generalized to Riemannian manifolds and is called "Riemannian mean" in the literature (see, for instance, [13, 16]) . Contrary to the Euclidean situation, only a necessary condition for a point $Q \in SO(n)$ to be the Riemannian mean of the points $Q_0, \ldots, Q_N$ is known. That condition, which corresponds to the Euler-Lagrange equation associated to the optimization problem giving rise to the Riemannian mean, is written as

$$\sum_{i=0}^{N} \log(Q_i^\top Q) = 0. \tag{18}$$

A slight modification of the previous may be done in order to contemplate the situation when different masses are attached to the points. In particular, if the mass t_i is attached to the point Q_i, the necessary condition for $Q \in SO(n)$ to be the "weighted" Riemannian mean is that

$$\sum_{i=0}^{N} t_i \log(Q_i^\top Q) = 0. \tag{19}$$

Now assume that the optimization problem $(\mathcal{P}_2)$ has been formulated on a connected, compact and abelian subgroup G of $SO(n)$ (typically a torus $SO(2) \times \cdots \times SO(2)$). In this case, every pair of elements in the Lie algebra $\mathcal{L}$ of G commute, so $\mathrm{e}^X \mathrm{e}^Y = \mathrm{e}^{X+Y}, \forall X, Y \in \mathcal{L}$ and $\log(PQ) = \log P + \log Q, \forall P, Q \in G$. We are now in conditions to prove the following result, which generalizes to a torus the properties referred in remark 2.2.

Theorem 3.5. *If $Q_0, \ldots, Q_N$ belonging to a connected, compact and abelian subgroup G of $SO(n)$ and $0 = t_0 < t_1 < \ldots < t_N = 1$ are given, then a geodesic $t \mapsto \gamma(t) = \Theta \mathrm{e}^{tX}$ in G satisfies the equations (17) if and only if*

$$\left\{ \begin{array}{l} \displaystyle\sum_{i=0}^{N} \log\left(Q_i^\top \gamma\left(\overline{t}\right)\right) = 0 \\ \displaystyle\sum_{i=0}^{N} t_i \log\left(Q_i^\top \gamma\left(\overline{\overline{t}}\right)\right) = 0 \end{array} \right. , \tag{20}$$

where $\overline{t}$ and $\overline{\overline{t}}$ are defined as in remark 2.2.

Proof: Due to the commutativity property, the first equation of (17) is equivalent to the first equation of (20). Indeed,

$$\sum_{i=0}^{N} \log\left(Q_i^\top \gamma(\overline{t})\right) = \log\left(\left(\prod_{i=0}^{N} Q_i^\top\right) \Theta^{N+1} \mathrm{e}^{\sum_{i=0}^{N} t_i X}\right)$$

$$
\begin{aligned}
&= \log\left(\prod_{i=0}^{N}(Q_i^\top \Theta \mathrm{e}^{t_i X})\right)\\
&= \sum_{i=0}^{N}\log\left(Q_i^\top \Theta \mathrm{e}^{t_i X}\right)\\
&= \sum_{i=0}^{N}\log\left(\mathrm{e}^{t_i X} Q_i^\top \Theta\right).
\end{aligned}
$$

A similar calculation proves the equivalence between the second equations of (17) and (20). ■

3.2. Fitting geodesics to data on spheres

Analogously to the previous section, we will deduce the counterpart of the "normal equations" for the unit sphere $S^n = \left\{p \in \mathbb{R}^{n+1} \,:\, \|p\| = 1\right\}$, equipped with the Riemannian metric induced by the Euclidean metric in $\mathbb{R}^{n+1}$.

The tangent space to S^n at a point $p \in S^n$ is the vector subspace of $\mathbb{R}^{n+1}$ given by

$$
T_pS^n = \left\{v \in \mathbb{R}^{n+1} \,:\, \langle v, p\rangle = 0\right\},
$$

and the projection of a vector $w \in \mathbb{R}^{n+1}$ into the tangent space T_pS^n is

$$
w - \langle w, p\rangle p. \tag{21}
$$

A point $p \in S^n$ and a vector $v \in T_pS^n$ uniquely define a geodesic $t \mapsto \gamma(t)$ in S^n, which passes through p at $t = 0$, with velocity v and is given by:

$$
\gamma(t) = p\cos\left(\|v\|t\right) + \frac{v}{\|v\|}\sin\left(\|v\|t\right). \tag{22}
$$

Again, we are given a finite set of points $q_0, q_1, \ldots, q_N$ in S^n, a partition $0 = t_0 < t_1 < \cdots < t_N = 1$, of the time interval $[0, 1]$, and the objective is to find the geodesic on S^n that best fits this data. This least squares problem is thus formulated as:

$$
(\mathcal{P}_3) \qquad \min_{(p,v)\in TS^n} \frac{1}{2}\sum_{i=0}^{N} d^2(q_i, \gamma(t_i)), \tag{23}
$$

where $d(p, q)$ denotes the spherical distance between two points p and q.

In order to determine the first order necessary conditions for the optimization problem $(\mathcal{P}_3)$, we consider the following function defined in the tangent bundle TS^n,

$$\begin{array}{rccl} F: & TS^n & \longrightarrow & \mathbb{R} \\ & (p,v) & \longmapsto & F(p,v) = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N} d^2\left(q_i, p\cos(\|v\|t_i) + \tfrac{v}{\|v\|}\sin(\|v\|t_i)\right), \end{array}$$

and endow the manifold TS^n with a natural Riemannian metric (the Sasaki metric induced by the Riemannian metric on S^n) defined on each tangent space at $(p,v) \in TS^n$, by the inner product

$$\ll (w_1, z_1), (w_2, z_2) \gg = w_1^\top w_2 + z_1^\top z_2, \tag{24}$$

where $w_i,\ z_i \in T_pS^n$, for $i = 1, 2$.

In order to find the expression of the critical points for the function F, we recall the expression for the shortest geodesic $t \mapsto x(t)$ in S^n that joins 2 non antipodal points, p (at $t = 0$) to q (at $t = 1$):

$$x(t) = p\cos(\alpha t) + \frac{q - p\cos\alpha}{\sin\alpha}\sin(\alpha t),$$

where $\alpha = \cos^{-1}\langle p, q\rangle \in]0, \pi[$. The geodesic distance between the two points is precisely equal to α.

The geodesic distance between the points q_i and $\gamma(t_i)$ is, therefore,

$$d(q_i, \gamma(t_i)) = \cos^{-1}\langle q_i, \gamma(t_i)\rangle = \cos^{-1}\left\langle q_i, p\cos(\|v\|t_i) + \tfrac{v}{\|v\|}\sin(\|v\|t_i)\right\rangle,$$

which clearly depends on the point $(p,v) \in TS^n$. To emphasize this dependence, from now on we use the following notation

$$\alpha_i = \alpha_i(p,v) = \cos^{-1}\left\langle q_i, p\cos(\|v\|t_i) + \tfrac{v}{\|v\|}\sin(\|v\|t_i)\right\rangle. \tag{25}$$

Therefore, the function F above may be rewritten as

$$\begin{aligned} F(p,v) &= \frac{1}{2}\sum_{i=0}^{N}\alpha_i^2(p,v) \qquad (26)\\ &= \frac{1}{2}\sum_{i=0}^{N}\cos^{-2}\left\langle q_i, p\cos(\|v\|t_i) + \frac{v}{\|v\|}\sin(\|v\|t_i)\right\rangle. \end{aligned}$$

A point $(p,v) \in TS^n$ is a critical point of F if and only if

$$T_{(p,v)}F(w,z) = 0,\ \ \forall w,\ z \in T_pS^n,$$

where the tangent map of F at the point $(p,v) \in TS^n$, can be defined as

$$\begin{array}{rccc} T_{(p,v)}F: & T_{(p,v)}(TS^n) & \longrightarrow & \mathbb{R} \\ & (w,z) & \longmapsto & \left.\dfrac{d}{ds}\right|_{s=0} F(c(s)), \end{array}$$

for any smooth curve $s \mapsto c(s)$ in TS^n, defined in a small neighborhood of $s=0$ and satisfying $c(0)=(p,v)$ and $\dot{c}(0)=(w,z)$. We may take $c(s)=(p(s),v(s)) \in S^n \times T_pS^n$, with

$$p(s) = p\cos(\|w\|s) + \tfrac{w}{\|w\|}\sin(\|w\|s), \qquad c(s) = v + sz.$$

It is straightforward to check that c satisfies the required initial conditions. So, according to (26),

$$\left.\frac{d}{ds}\right|_{s=0} F(c(s)) = \sum_{i=0}^{N} \alpha_i(p,v) \left.\frac{d}{ds}\right|_{s=0} \alpha_i\left(p(s),v(s)\right). \tag{27}$$

In order to proceed with the computation of the tangent map of F, notice that,

$$\left.\frac{d}{ds}\right|_{s=0} \|v(s)\| = \frac{1}{\|v(0)\|}\langle v(0),\dot{v}(0)\rangle = \frac{1}{\|v\|}\langle v,z\rangle.$$

Hence, according to the notation introduced in (25) and using the initial conditions satisfied by the curve c, we can write

$$\begin{aligned}
&\left.\frac{d}{ds}\right|_{s=0} \alpha_i\left(p(s),v(s)\right) \\
=\;& \left.\frac{d}{ds}\right|_{s=0} \cos^{-1}\left\langle q_i, p(s)\cos\left(\|v(s)\|t_i\right) + \tfrac{v(s)}{\|v(s)\|}\,\sin\left(\|v(s)\|t_i\right)\right\rangle \\
=\;& \tfrac{-1}{\sin\alpha_i}\left\langle q_i, w\cos\left(\|v\|t_i\right) - p\tfrac{\langle v,z\rangle}{\|v\|}t_i\sin\left(\|v\|t_i\right) + \tfrac{z\|v\|^2 - v\langle v,z\rangle}{\|v\|^3}\sin\left(\|v\|t_i\right)\right. \\
&\left. + \;\tfrac{v\langle v,z\rangle t_i}{\|v\|^2}\cos\left(\|v\|t_i\right)\right\rangle \\
=\;& \tfrac{-1}{\sin\alpha_i}\left[\left\langle \cos\left(\|v\|t_i\right)q_i, w\right\rangle - \right. \\
&\left.\left\langle \tfrac{\sin(\|v\|t_i)}{\|v\|}\left(t_i\langle q_i,p\rangle v - q_i + \langle q_i,v\rangle\tfrac{v}{\|v\|^2}\right) - t_i\cos\left(\|v\|t_i\right)\langle q_i,v\rangle\tfrac{v}{\|v\|^2}, z\right\rangle\right].
\end{aligned}$$

Replacing this last expression into (27), one obtains

$$\left.\frac{d}{ds}\right|_{s=0} F(c(s)) = \sum_{i=0}^{N} \tfrac{-\alpha_i}{\sin\alpha_i} \Big[\big\langle \cos\left(\|v\|t_i\right) q_i, w \big\rangle - \\ \Big\langle \tfrac{\sin(\|v\|t_i)}{\|v\|} \Big(t_i \langle q_i, p\rangle v - q_i + \langle q_i, v\rangle \tfrac{v}{\|v\|^2} \Big) - t_i \cos\left(\|v\|t_i\right) \langle q_i, v\rangle \tfrac{v}{\|v\|^2}, z \Big\rangle \Big] . \tag{28}$$

Consequently, $T_{(p,v)}F(w,z) = 0,\ \ \forall w,\ z \in T_pS^n$ if and only if the following conditions hold

$$\begin{cases} \displaystyle\sum_{i=0}^{N} \tfrac{-\alpha_i}{\sin\alpha_i} \Big\langle \cos\left(\|v\|t_i\right) q_i, w \Big\rangle = 0,\ \ \forall w \in T_pS^n \\ \displaystyle\sum_{i=0}^{N} \tfrac{-\alpha_i}{\sin\alpha_i} \Big\langle \tfrac{\sin(\|v\|t_i)}{\|v\|} \Big(-t_i \langle q_i, p\rangle v + q_i - \langle q_i, v\rangle \tfrac{v}{\|v\|^2} \Big) + \\ t_i \cos\left(\|v\|t_i\right) \langle q_i, v\rangle \tfrac{v}{\|v\|^2}, z \Big\rangle = 0,\ \ \forall z \in T_pS^n \end{cases} .$$

Now, taking into account that a vector in $\mathbb{R}^{n+1}$ belongs to $T_p^{\perp}S^n$ if and only if its orthogonal projection into T_pS^n (given by (21)) vanishes, we can conclude that the following statement is true.

Theorem 3.6. $(p,v) \in TS^n$ *is a critical point for the function* F *if and only if*

$$\begin{cases} \displaystyle\sum_{i=0}^{N} \tfrac{\alpha_i}{\sin\alpha_i} \cos(\|v\|t_i) \left(q_i - \langle q_i, p\rangle p \right) = 0 \\ \displaystyle\sum_{i=0}^{N} \tfrac{\alpha_i \sin(\|v\|t_i)}{\sin\alpha_i} \Big(q_i - \langle q_i, p\rangle (t_i v + p) - \tfrac{\langle q_i, v\rangle}{\|v\|^2} v \Big) = \sum_{i=0}^{N} \tfrac{-\alpha_i \cos(\|v\|t_i)}{\sin\alpha_i} \tfrac{\langle q_i, v\rangle}{\|v\|} t_i v \end{cases} , \tag{29}$$

where $\alpha_i = \cos^{-1} \Big\langle q_i, p\cos\left(\|v\|t_i\right) + \tfrac{v}{\|v\|} \sin\left(\|v\|t_i\right) \Big\rangle$, *for* $i = 0, \ldots, N$.

The equations (29) are called the "normal equations" for the unit sphere S^n. Attending to the nonlinearity of these equations, it seems a very difficult task to exhibit explicit solutions for problem $(\mathcal{P}_3)$.

4. Fitting smoothing splines to data on Riemannian manifolds

The linear regression problem was generalized to some particular Riemannian manifolds in the previous section. This was possible, due to the availability of explicit expressions for geodesics. However, for the more general least squares problem of fitting polynomial curves to data, the approach presented in the previous sections can't be generalized, the main difficulty being that no explicit solutions are known for the analogue to polynomial curves. However, polynomial curves on manifolds have been defined as solutions of the Euler-Lagrange equations associated to a variational problem. Interpolating polynomial splines on Riemannian manifolds can also be defined similarly, since they are composed of polynomial curves smoothly joined.

The most well studied are the cubic splines. Without being exhaustive, we refer to Jupp and Kent [11], where the equation for the cubic spline on S^2 has been deduced, Noakes, Heinzinger and Paden [17], Crouch and Silva Leite [6, 7] and Camarinha [2]. For general polynomial splines on Riemannian manifolds we refer the work of Camarinha, Silva Leite and Crouch [3]. The main motivation to study such problems arose from applications in engineering, and the number of works in this area has increased considerably in the last decade.

In this section we formulate a new least squares problem on a general Riemannian manifold M, that generates smoothing splines fitting a given data set of time labelled points in M.

The classical linear regression method developed in section 2 will arise as a limiting process of the method developed here.

In what follows, M denotes a complete and connected Riemannan manifold endowed with its Riemannian connection (Levi-Civita connection), that we denote by ∇, and $\frac{D}{dt}$ stands for the covariant derivative in M relative to its Riemannian connection.

For details concerning these standard notions in Riemannian manifolds, we refer to Boothby [1], Carmo [4], Helgason [8], Milnor [15], Nomizu [18], among others.

We will consider on M, a tensor of type (1,3), known as the curvature tensor, which is defined by

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, \tag{30}$$

where X, Y and Z are smooth vector fields on M.

We start with a collection of time labelled data points in M,

$$Q = \{(t_i, q_i) \ : i = 0, \ldots, N\}, \tag{31}$$

where the q_i's lie on M and $0 = t_0 < t_1 < \cdots < t_N = 1$ is a partition of the time interval $[0, 1]$. Our objective is to generate a smooth spline on M that best fits the data (in a sense to be made precise later). This process in now known as "smoothing by splines". We refer the work of Martin, Magnus and Sun ([21, 5]), for smoothing splines on Euclidean spaces and Jupp and Kent [11] for smoothing splines on the 2-dimensional sphere.

For the sake of simplicity, here we restrict our study to smoothing by cubic splines and start this section with some facts about cubic polynomials on Riemannian manifolds.

4.1. Cubic Polynomials on Riemannian Manifolds

Cubic polynomials in Euclidean spaces can be seen as curves along which changes in velocity are minimized.

In Riemannian manifolds, cubic polynomials arise as a generalization of this and result from the following variational principle

$$\min_{\gamma} \int_0^1 \langle \frac{D^2\gamma}{dt^2}, \frac{D^2\gamma}{dt^2} \rangle \, dt,$$

over the class of twice continuously differentiable paths $\gamma : [0, 1] \to M$ (typically satisfying some prescribed boundary conditions). (Here $\frac{D^2\gamma}{dt^2}$ denotes the covariant derivative of the velocity vector field $\frac{d\gamma}{dt}$, along γ).

The Euler-Lagrange equation associated to this problem is the fourth order differential equation

$$\frac{D^4\gamma}{dt^4} + R\left(\frac{D^2\gamma}{dt^2}, \frac{d\gamma}{dt}\right)\frac{d\gamma}{dt} = 0, \tag{32}$$

where R is the curvature tensor defined by (30). Although only twice continuous differentiability is required a priori, the solutions of this variational problem turn out to be C^∞ ([17, 6, 7, 2]).

Contrary to the situation in Euclidean spaces, there is no guarantee that the solutions of equation (32) minimize the energy functional above. Following Noakes et al. [17] we adopt the definition of a cubic polynomial as being any smooth curve, $\gamma : I \subset \mathbb{R} \longrightarrow M$, satisfying the fourth order differential equation (32). A geodesic on M, that is a smooth curve $\gamma : I \subset \mathbb{R} \longrightarrow M$, satisfying

$$\frac{D^2\gamma}{dt^2} = 0,$$

is just a particular cubic polynomial on M.

We now recall some useful results that were derived in Camarinha [2], the last two being concerned with the existence and uniqueness of cubic polynomials on M.

Remark 4.1. (Camarinha [2]) The quantity

$$\langle \frac{D^3\gamma}{dt^3}, \frac{d\gamma}{dt} \rangle - \frac{1}{2} \langle \frac{D^2\gamma}{dt^2}, \frac{D^2\gamma}{dt^2} \rangle,$$

is preserved along a smoothing path γ, satisfying (32).

Theorem 4.1. *For each point $(p, v, y, z) \in T^3M$, there exists a unique cubic polynomial $t \mapsto \gamma(t)$ on M, defined for $t \in (-\epsilon, \epsilon)$, $\epsilon > 0$, satisfying the initial conditions*

$$\gamma(0) = p, \; \frac{d\gamma}{dt}(0) = v, \; \frac{D^2\gamma}{dt^2}(0) = y, \; \frac{D^3\gamma}{dt^3}(0) = z. \tag{33}$$

This cubic polynomial depends differentiably on (p, v, y, z). Moreover, if there is another cubic polynomial satisfying the same initial conditions (33), then they coincide in an open interval containing $t = 0$.

The next theorem gives conditions that allow the extension indefinitely of the domain of a cubic polynomial.

Theorem 4.2. *Let (U, ϕ) be a system of local coordinates in M, $q \in U$ and τ a positive real number. Then, there exists a neighborhood D of q, $D \subset U$ and a real number $\delta > 0$ such that, if $p \in D$ and $v, y, z \in T_pM$ with $\|v\| < \delta$, $\|y\| < \delta$, $\|z\| < \delta$, there exists a unique cubic polynomial $\gamma : (-\tau, \tau) \longrightarrow U$ satisfying the initial conditions (33). Moreover, this cubic polynomial depends differentiably on (p, v, y, z).*

Cubic splines on a Riemannian manifold M are obtained by piecing together smoothly segments of cubic polynomials. The typical situation is that of an interpolating cubic spline, which is required to be C^2, to pass through some prescribed points q_i in M at prescribed instants of time t_i, and to be a cubic polynomial, when restricted to each subinterval. However, here we are mainly interested in cubic splines that best approximate some data.

4.2. Problem's Formulation

Before the correct formulation of our problem, we define the family of admissible paths. Let $0 = t_0 < \ldots < t_N = 1$ be a given partition of the time interval $[0, 1]$ and $q_0, \ldots, q_N$ be a distinct set of points in M.

Definition 4.1. By an *admissible path* will be meant a C^2 (twice continuously differentiable) path, $\gamma : [0, 1] \to M$, satisfying

- $\gamma\Big|_{[t_i,t_{i+1}]}$ is smooth, for $i = 0, \ldots, N-1$;
- the left and right covariant derivatives at the knot point t_i,

$$\lim_{t\to t_i^-} \frac{D^k\gamma}{dt^k}(t) = \frac{D^k\gamma}{dt^k}(t_i^-), \qquad \lim_{t\to t_i^+} \frac{D^k\gamma}{dt^k}(t) = \frac{D^k\gamma}{dt^k}(t_i^+),$$

  exist, for every integer $k \geq 3$.

The set of all admissible paths will be hereafter denoted by Ω.

Our main objective is to find an admissible path on M that best fits the given points at the given instants of time, in the sense that the cost functional

$$J(\gamma) = \frac{1}{2}\sum_{i=0}^{N} \frac{1}{w_i} d^2\left(q_i, \gamma(t_i)\right) + \frac{\lambda}{2}\int_0^1 \langle \frac{D^2\gamma}{dt^2}, \frac{D^2\gamma}{dt^2}\rangle\, dt, \tag{34}$$

defined over Ω, should be as small as possible. Here $d(p, q)$ denotes the Riemannian distance between p and q, $\lambda > 0$ is a real parameter and the w_i are weights (positive real numbers) associated with the given data points. For our purposes here, we will assume $w_i = 1$, for all $i = 0, \ldots, N$. λ plays the role of a smoothing parameter.

If p and q are points in M sufficiently close, the shortest geodesic arc joining them may be parameterized explicitly by (Karcher [12])

$$c(s) = \exp_p\left(s \exp_p^{-1}(q)\right), \;\; s \in [0, 1], \tag{35}$$

and the geodesic distance $d(p, q)$, is given by

$$d(p, q) = \langle c'(s), c'(s)\rangle^{\frac{1}{2}} = \int_0^1 \langle c'(s), c'(s)\rangle^{\frac{1}{2}}\, ds,$$

where $c'(s) = dc(s)/ds$.

If $s \mapsto c_i(s)$ denotes the geodesic curve joining q_i to $\gamma(t_i)$, for $i = 0, \ldots, N$, we can rewrite the functional J as

$$J(\gamma) = \frac{1}{2}\sum_{i=0}^{N}\int_0^1 \langle c_i'(s), c_i'(s)\rangle \, ds + \frac{\lambda}{2}\int_0^1 \langle \frac{D^2\gamma}{dt^2}, \frac{D^2\gamma}{dt^2}\rangle \, dt, \tag{36}$$

Consequently, our optimization problem can be formulated as follows:

$$(\mathcal{P}_4) \qquad \min_{\gamma\in\Omega} \; J(\gamma) = \frac{1}{2}\int_0^1 \left(\sum_{i=0}^{N}\langle c_i'(s), c_i'(s)\rangle\right) ds + \frac{\lambda}{2}\int_0^1 \langle \frac{D^2\gamma}{dt^2}, \frac{D^2\gamma}{dt^2}\rangle \, dt.$$

4.3. Main results

Our objective consists on finding the curves $\gamma \in \Omega$ such that

$$J(\gamma) \leq J(\omega),$$

for all admissible paths ω in a neighborhood of γ. These curves γ are called *local minimizers* for the functional J and the values $J(\gamma)$ are called *local minima* of J. Those curves $\gamma \in \Omega$, which satisfy

$$J(\gamma) \leq J(\omega),$$

for all admissible paths ω are called *global minimizers* of J and the values $J(\gamma)$ are *global minima* of J.

In order to find the critical paths for J, one needs to define an *admissible variation* of $\gamma \in \Omega$. Apart from adaptations to the present situation, this follows closely what has been done in the literature, to derive first order conditions for Riemannian cubic splines.

Definition 4.2. Let $\gamma : [0,1] \to M$ be an admissible path in M, in the sense of definition 4.1. By a *one-parameter variation* of γ will be meant a C^2 function

$$\alpha : (-\epsilon, \epsilon) \times [0,1] \to M,$$

for some $\epsilon > 0$, such that:

- $\alpha(0,t) = \gamma(t)$,
- α is smooth on each strip $(-\epsilon, \epsilon) \times [t_i, t_{i+1}]$, $i = 0, \ldots, N-1$.

Moreover, if $\alpha(u,0) = \gamma(0)$ and $\alpha(u,1) = \gamma(1)$, for all $u \in (-\epsilon, \epsilon)$, then α is called a *proper variation* of γ.

We may use the notation $\alpha_t : (-\epsilon, \epsilon) \to \Omega$, defined by $\alpha_t(u) = \alpha(u, t)$, to denote the variation α. α_t can be seen as a "smooth path" in Ω and its velocity vector $\dfrac{d\alpha_t}{du}(0) \in T_\gamma\Omega$ is defined as the vector field W along γ given by

$$W(t) = \frac{d\alpha_t}{du}(0) = \frac{\partial\alpha}{\partial u}(0, t).$$

Clearly $W \in T_\gamma\Omega$ and we will refer to this vector field as the *variational vector field* associated with the variation α.

We can thought of Ω as an infinite dimensional manifold and introduce the tangent space of Ω at a path γ, $T_\gamma\Omega$, as the set of all C^2 variational vector fields $t \mapsto W(t)$ along γ, satisfying

- $t \longmapsto W(t)$ is smooth on the domains $[t_i, t_{i+1}]$, for $i = 0, \ldots, N$.
- $t \longmapsto \dfrac{D^2W}{dt^2}(t)$ is continuous in $[0, 1]$.

Hence, by exponentiating a vector field $W \in T_\gamma\Omega$, we obtain a one-parameter variation of γ, $\alpha : (-\epsilon, \epsilon) \times [0, 1] \to M$, defined by

$$\alpha(u, t) = \exp_{\gamma(t)}\left(uW(t)\right), \tag{37}$$

for some $\epsilon > 0$.

Theorem 4.3. *If α is a one-parameter variation of $\gamma \in \Omega$ and $W \in T_\gamma\Omega$ is the variational vector field associated to α, then*

$$\begin{aligned}
&\left.\frac{d}{du}\right|_{u=0} J(\alpha_t(u)) = \sum_{i=1}^{N-1} \langle W(t_i), \lambda\frac{D^3\gamma}{dt^3}(t_i^+) - \lambda\frac{D^3\gamma}{dt^3}(t_i^-) - \exp^{-1}_{\gamma(t_i)}(q_i)\rangle + \\
&\langle W(0), \lambda\frac{D^3\gamma}{dt^3}(0) - \exp^{-1}_{\gamma(0)}(q_0)\rangle - \langle W(1), \lambda\frac{D^3\gamma}{dt^3}(1) + \exp^{-1}_{\gamma(1)}(q_N)\rangle + \\
&\lambda\langle\frac{DW}{dt}(1), \frac{D^2\gamma}{dt^2}(1)\rangle - \lambda\langle\frac{DW}{dt}(0), \frac{D^2\gamma}{dt^2}(0)\rangle + \\
&\lambda\int_0^1 \langle\frac{D^4\gamma}{dt^4} + R\left(\frac{D^2\gamma}{dt^2}, \frac{d\gamma}{dt}\right)\frac{d\gamma}{dt}, W\rangle\, dt.
\end{aligned}$$

Proof: Before we compute the value of $\frac{d}{du}J(\alpha_t(u))$, we make some considerations.

Since $d(q_i, \gamma(t_i))$ denotes the Riemannian distance between points q_i and $\gamma(t_i)$, and since, according to (35),

$$c_i(s) = \exp_{q_i}\left(s \exp_{q_i}^{-1}(\gamma(t_i))\right), \;\; s \in [0,1],$$

by considering the variation α of γ, defined by (37), one obtains the parameterized surface in M (Krakowski [13]), given by

$$c_i(s,u) = \exp_{q_i}\left(s \exp_{q_i}^{-1}(\alpha_{t_i}(u))\right), \;\; s \in [0,1], \, u \in (-\epsilon, \epsilon). \tag{38}$$

Let us introduce the following vector fields associated to (38),

$$c_i'(s,u) = \frac{\partial}{\partial s} c_i(s,u), \qquad \dot{c}_i(s,u) = \frac{\partial}{\partial u} c_i(s,u). \tag{39}$$

Since for fixed u, $s \longmapsto c_i(s,u)$ is a family of geodesics, $s \longmapsto \dot{c}_i(s,u)$ is a family of Jacobi vector fields along the family of geodesics $s \longmapsto c_i(s,u)$. This can now be used in the following .

$$\begin{aligned}
\frac{d}{du} J(\alpha_t(u)) &= \sum_{i=0}^{N} \int_0^1 \langle \frac{D}{\partial u} c_i'(s,u), c_i'(s,u) \rangle \, ds + \lambda \int_0^1 \langle \frac{D}{\partial u} \frac{D}{\partial t} \frac{\partial \alpha}{\partial t}, \frac{D^2 \alpha}{\partial t^2} \rangle \, dt \\
&= \sum_{i=0}^{N} \int_0^1 \langle \frac{D}{\partial s} \dot{c}_i(s,u), c_i'(s,u) \rangle \, ds \\
&\quad + \lambda \int_0^1 \langle \frac{D}{\partial t} \frac{D}{\partial u} \frac{\partial \alpha}{\partial t} + R\left(\frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial t}\right) \frac{\partial \alpha}{\partial t}, \frac{D^2 \alpha}{\partial t^2} \rangle \, dt \\
&= \sum_{i=0}^{N} \left(\langle \dot{c}_i(1,u), c_i'(1,u) \rangle - \langle \dot{c}_i(0,u), c_i'(0,u) \rangle \right) \\
&\quad + \lambda \int_0^1 \langle \frac{D}{\partial t} \frac{D}{\partial t} \frac{\partial \alpha}{\partial u}, \frac{D^2 \alpha}{\partial t^2} \rangle \, dt \\
&\quad + \lambda \int_0^1 \langle R\left(\frac{D^2 \alpha}{\partial t^2}, \frac{\partial \alpha}{\partial u},\right) \frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial u} \rangle \, dt.
\end{aligned} \tag{40}$$

Since for each $i = 0, \ldots, N$, $c_i(0,u) = q_i$, we have $\dot{c}_i(0,u) = 0$.

Moreover, since the C^2-path α is smooth on each strip $(-\epsilon, \epsilon) \times [t_i, t_{i+1}]$, we can integrate by parts twice on $[t_i, t_{i+1}]$ the first integral on (40), (as in

[6], [2])

$$\begin{aligned}
\int_0^1 \langle \frac{D}{\partial t}\frac{D}{\partial t}\frac{\partial \alpha}{\partial u}, \frac{D^2\alpha}{\partial t^2}\rangle\, dt &= \int_0^1 \frac{d}{dt}\langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial u}, \frac{D^2\alpha}{\partial t^2}\rangle\, dt - \int_0^1 \langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial u}, \frac{D^3\alpha}{\partial t^3}\rangle\, dt \\
&= \langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial u}(1), \frac{D^2\alpha}{\partial t^2}(1)\rangle - \langle \frac{D}{dt}\frac{\partial \alpha}{\partial u}(0), \frac{D^2\alpha}{\partial t^2}(0)\rangle \\
&\quad - \int_0^1 \frac{d}{dt}\langle \frac{\partial \alpha}{\partial u}, \frac{D^3\alpha}{\partial t^3}\rangle\, dt + \int_0^1 \langle \frac{\partial \alpha}{\partial u}, \frac{D^4\alpha}{\partial t^4}\rangle\, dt \\
&= \langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial u}(1), \frac{D^2\alpha}{\partial t^2}(1)\rangle - \langle \frac{D}{\partial t}\frac{\partial \alpha}{\partial u}(0), \frac{D^2\alpha}{\partial t^2}(0)\rangle \\
&\quad - \sum_{i=0}^{N-1}\langle \frac{\partial \alpha}{\partial u}, \frac{D^3\alpha}{\partial t^3}\rangle\Big|_{t_i^+}^{t_{i+1}^-} + \int_0^1 \langle \frac{\partial \alpha}{\partial u}, \frac{D^4\alpha}{\partial t^4}\rangle\, dt. \qquad (41)
\end{aligned}$$

Plugging expression (41) into (40) and considering $u = 0$, we obtain,

$$\begin{aligned}
&\frac{d}{du}\Big|_{u=0} J(\alpha_t(u)) \\
&= -\sum_{i=0}^{N}\langle W(t_i), \exp^{-1}_{\gamma(t_i)}(q_i)\rangle + \lambda\langle \frac{DW}{dt}(1), \frac{D^2\gamma}{dt^2}(1)\rangle - \lambda\langle \frac{DW}{dt}(0), \frac{D^2\gamma}{dt^2}(0)\rangle \\
&+\lambda\langle W(0), \frac{D^3\gamma}{dt^3}(0)\rangle - \lambda\langle W(1), \frac{D^3\gamma}{dt^3}(1)\rangle \\
&+\sum_{i=1}^{N-1}\langle W(t_i), \lambda\frac{D^3\gamma}{dt^3}(t_i^+) - \lambda\frac{D^3\gamma}{dt^3}(t_i^-)\rangle \\
&+\lambda\int_0^1 \langle \frac{D^4\gamma}{dt^4} + R\left(\frac{D^2\gamma}{dt^2}, \frac{d\gamma}{dt}\right)\frac{d\gamma}{dt}, W\rangle\, dt \\
= \;& \sum_{i=1}^{N-1}\langle W(t_i), \lambda\frac{D^3\gamma}{dt^3}(t_i^+) - \lambda\frac{D^3\gamma}{dt^3}(t_i^-) - \exp^{-1}_{\gamma(t_i)}(q_i)\rangle \\
&+\langle W(0), \lambda\frac{D^3\gamma}{dt^3}(0) - \exp^{-1}_{\gamma(0)}(q_0)\rangle - \langle W(1), \lambda\frac{D^3\gamma}{dt^3}(1) + \exp^{-1}_{\gamma(1)}(q_N)\rangle \\
&+\lambda\langle \frac{DW}{dt}(1), \frac{D^2\gamma}{dt^2}(1)\rangle - \lambda\langle \frac{DW}{dt}(0), \frac{D^2\gamma}{dt^2}(0)\rangle
\end{aligned}$$

$$+\lambda \int_0^1 \langle \frac{D^4\gamma}{dt^4} + R\left(\frac{D^2\gamma}{dt^2}, \frac{d\gamma}{dt}\right)\frac{d\gamma}{dt}, W\rangle \, dt.$$

■

As a consequence of the previous theorem, we can define the linear form $T_\gamma J$, in $T_\gamma\Omega$, by setting

$$T_\gamma J(W) = \left.\frac{d}{du}\right|_{u=0} J(\alpha_t(u)).$$

$T_\gamma J(W)$ is known as the *first variation* of J in γ and plays a crucial role to establish necessary conditions for γ to be a solution of the optimization problem $(\mathcal{P}_4)$. Indeed, if $\gamma \in \Omega$ is a minimizer for J then γ is a critical path of J, that is $T_\gamma J(W) = 0$, for all $W \in T_\gamma\Omega$.

The next result gives necessary optimality conditions for problem $(\mathcal{P}_4)$. It generalizes for general Riemannian manifolds the results of Jupp and Kent [11] for the 2-sphere.

Theorem 4.4. *If $\gamma \in \Omega$ is a minimizer for J, then on each subinterval $[t_i, t_{i+1}]$, $i = 0, \dots, N-1$, γ is smooth and satisfies the following condition*

$$\frac{D^4\gamma}{dt^4} + R\left(\frac{D^2\gamma}{dt^2}, \frac{d\gamma}{dt}\right)\frac{d\gamma}{dt} = 0. \tag{42}$$

Moreover, at the knot points t_i, γ satisfies the following differentiability conditions

$$\frac{D^k\gamma}{dt^k}(t_i^+) - \frac{D^k\gamma}{dt^k}(t_i^-) = \begin{cases} 0, & k = 0, 1, \quad (i = 1, \dots, N-1) \\ 0, & k = 2, \quad (i = 0, \dots, N) \\ \dfrac{1}{\lambda}\exp^{-1}_{\gamma(t_i)}(q_i), & k = 3, \quad (i = 0, \dots, N) \end{cases}, \tag{43}$$

where we assume for shorten of notation that

$$\frac{D^2\gamma}{dt^2}(t_0^-) = \frac{D^3\gamma}{dt^3}(t_0^-) = \frac{D^2\gamma}{dt^2}(t_N^+) = \frac{D^3\gamma}{dt^3}(t_N^+) = 0. \tag{44}$$

Proof: Assume that $\gamma \in \Omega$ is a minimizer for J. Then $T_\gamma J(W) = 0$, $\forall W \in T_\gamma\Omega$. Taking into account conditions (44), we can write $T_\gamma J$ in the simplified form

$$T_\gamma J(W) = \sum_{i=0}^{N} \langle W(t_i), \lambda\frac{D^3\gamma}{dt^3}(t_i^+) - \lambda\frac{D^3\gamma}{dt^3}(t_i^-) - \exp^{-1}_{\gamma(t_i)}(q_i)\rangle +$$

$$\lambda\langle\frac{DW}{dt}(1),\frac{D^2\gamma}{dt^2}(1)\rangle-\lambda\langle\frac{DW}{dt}(0),\frac{D^2\gamma}{dt^2}(0)\rangle+$$
$$\lambda\int_0^1\langle\frac{D^4\gamma}{dt^4}+R\left(\frac{D^2\gamma}{dt^2},\frac{d\gamma}{dt}\right)\frac{d\gamma}{dt},W\rangle\,dt \tag{45}$$

Now, take $W\in T_\gamma\Omega$ to be the vector field defined by

$$W(t)=F(t)\left[\frac{D^4\gamma}{dt^4}+R\left(\frac{D^2\gamma}{dt^2},\frac{d\gamma}{dt}\right)\frac{d\gamma}{dt}\right],$$

where $F:[0,1]\to\mathbb{R}$ is a smooth real function defined in the interval $[0,1]$, such that $F(t_i)=\frac{dF}{dt}(t_i)=0$, and $F(t)>0$, for $t\neq t_i$, for $i=0,\ldots,N$.

Therefore, the equality

$$T_\gamma J(W)=\int_0^1\lambda F(t)\left\|\frac{D^4\gamma}{dt^4}+R\left(\frac{D^2\gamma}{dt^2},\frac{d\gamma}{dt}\right)\frac{d\gamma}{dt}\right\|^2 dt,$$

holds. But, since $\lambda>0$ and $F(t)>0$ for all $t\in[0,1]$, except for a finite number, we conclude that for each domain $[t_i,t_{i+1}]$

$$\left\|\frac{D^4\gamma}{dt^4}+R\left(\frac{D^2\gamma}{dt^2},\frac{d\gamma}{dt}\right)\frac{d\gamma}{dt}\right\|^2=0,$$

which implies (42). This means that γ should be a smooth cubic polynomial in each interval $[t_i,t_{i+1}]$.

Now consider a vector field $W\in T_\gamma\Omega$ in such a way that

$$\frac{DW}{dt}(0)=-\frac{D^2\gamma}{dt^2}(0),\qquad \frac{DW}{dt}(1)=\frac{D^2\gamma}{dt^2}(1),$$

and

$$W(t_i)=\lambda\frac{D^3\gamma}{dt^3}(t_i^+)-\lambda\frac{D^3\gamma}{dt^3}(t_i^-)-\exp^{-1}_{\gamma(t_i)}(q_i),\ \ i=0,\ldots,N.$$

Then,

$$\begin{aligned}T_\gamma J(W) &= \lambda\left[\left\|\frac{D^2\gamma}{dt^2}(0)\right\|^2+\left\|\frac{D^2\gamma}{dt^2}(1)\right\|^2\right]+\\ &\quad \sum_{i=0}^N\left\|\lambda\frac{D^3\gamma}{dt^3}(t_i^+)-\lambda\frac{D^3\gamma}{dt^3}(t_i^-)-\exp^{-1}_{\gamma(t_i)}(q_i)\right\|^2,\end{aligned}$$

hence,

$$\frac{D^2\gamma}{dt^2}(0) = \frac{D^2\gamma}{dt^2}(1) = 0,$$

and

$$\lambda\frac{D^3\gamma}{dt^3}(t_i^+) - \lambda\frac{D^3\gamma}{dt^3}(t_i^-) - \exp^{-1}_{\gamma(t_i)}(q_i) = 0, \;\; i = 0,\dots,N.$$

Putting together all the conditions above, we see that if γ is a minimizer for J then γ is a cubic polynomial in each domain $[t_i, t_{i+1}]$, $i = 0,\dots,N-1$, and at the knot points t_i, it satisfies the system of equations (43). ■

Proposition 4.5. *If in conditions (42)-(43) of theorem 4.4, one considers the smoothing parameter λ converging to $+\infty$, then the cubic spline defined in $[0,1]$ converges to a geodesic curve on M that best fits the given data set of points at the given instants of time.*

Proof: If λ tends to $+\infty$, then according to the last condition given in (43), we conclude that for each $i = 0,\dots,N$

$$\frac{D^3\gamma}{dt^3}(t_i^-) = \frac{D^3\gamma}{dt^3}(t_i^+), \tag{46}$$

which means that γ is of class C^3 in the whole interval $[0,1]$. In particular, attending to equations (44)-(46), we have

$$\frac{D^3\gamma}{dt^3}(t_0^+) = \frac{D^3\gamma}{dt^3}(t_N^-) = 0. \tag{47}$$

We will denote the unique value given in (46) by $\frac{D^3\gamma}{dt^3}(t_i)$ and the cubic spline defined in the interval $[t_i, t_{i+1}]$ by γ_i.

Then, attending to the existence and uniqueness theorem for ordinary differential equations and in particular to the existence and uniqueness theorem for cubic polynomials (theorem 4.1), given

$$\left(\gamma_i(t_i), \frac{d\gamma_i}{dt}(t_i), \frac{D^2\gamma_i}{dt^2}(t_i), \frac{D^3\gamma_i}{dt^3}(t_i)\right) \in T^3M, \tag{48}$$

there exists a unique cubic polynomial on M,

$$c : (t_i - \epsilon, t_i + \epsilon) \longrightarrow M, \;\; (\epsilon > 0)$$

satisfying

$$c(t_i) = \gamma_i(t_i),\; \frac{dc}{dt}(t_i) = \frac{d\gamma_i}{dt}(t_i),\; \frac{D^2c}{dt^2}(t_i) = \frac{D^2\gamma_i}{dt^2}(t_i),\; \frac{D^3c}{dt^3}(t_i) = \frac{D^3\gamma_i}{dt^3}(t_i).$$

Now, attending to theorem 4.2 and to the fact that two polynomials satisfying the same initial conditions (48) must coincide in an open interval containing t_i, we conclude that

$$\gamma_i \equiv \gamma_{i-1}, \ \ \forall i = 1, \dots, N-1,$$

and it is an immediate consequence that all the piecewise cubic splines are in fact a unique (smooth) cubic polynomial in the whole interval $[0,1]$. That is, we have

$$\frac{D^4\gamma}{dt^4} + R\left(\frac{D^2\gamma}{dt^2}, \frac{d\gamma}{dt}\right)\frac{d\gamma}{dt} = 0, \ \ \forall t \in [0,1].$$

Let us denote by c the unique cubic polynomial on M satisfying the initial conditions

$$c(0) = \gamma(0), \ \ \frac{dc}{dt}(0) = \frac{d\gamma}{dt}(0), \ \ \frac{D^2c}{dt^2}(0) = 0, \ \ \frac{D^3c}{dt^3}(0) = 0. \tag{49}$$

Now, since a geodesic is a particular case of a cubic polynomial, according to the existence and uniqueness of cubic polynomials given by theorem 4.1, the cubic polynomial c characterized by (49) is in fact a geodesic in M. ■

We notice that an alternative proof of proposition 4.5 can be made by integrating the invariant along a cubic polynomial given in remark 4.1.

Proposition 4.6. *If in conditions (42)-(43) of theorem 4.4, one considers the smoothing parameter λ converging to* 0*, then the cubic spline defined in* $[0,1]$ *converges to an interpolating smoothing spline on M that passes through the given data set of points at the given instants of time.*

Proof: In fact, the third condition of the system of equations (43) reduces to

$$\gamma(t_i) = q_i, \ \ i = 0, \dots, N,$$

and the solution of (42)-(43) is a cubic spline that interpolates the given points at the given instants of time. ■

4.4. The Particular Case When $M = \mathbb{R}^n$

As a title of example, we will see that the method developed in the previous section is a generalization of the classical variational method in Euclidean spaces for generating smoothing spline functions and presented in Reinsch [19].

The counterpart of theorem 4.4 for the Euclidean space $\mathbb{R}^n$, is given by the following.

Theorem 4.7. *If $\gamma \in \Omega$ is a minimizer for J, then on each domain $[t_i, t_{i+1}]$, γ is smooth and satisfies the following condition*

$$\frac{d^4\gamma}{dt^4} = 0. \tag{50}$$

Moreover, at the knot points t_i, γ satisfies the following system of equations

$$\frac{d^k\gamma}{dt^k}(t_i^+) - \frac{d^k\gamma}{dt^k}(t_i^-) = \begin{cases} 0, & k = 0, 1, \quad (i = 1, \dots, N-1) \\ 0, & k = 2, \quad (i = 0, \dots, N) \\ \dfrac{1}{\lambda}\left(q_i - \gamma(t_i)\right), & k = 3, \quad (i = 0, \dots, N) \end{cases}, \tag{51}$$

where we have assumed for convenience

$$\frac{d^2\gamma}{dt^2}(t_0^-) = \frac{d^3\gamma}{dt^3}(t_0^-) = \frac{d^2\gamma}{dt^2}(t_N^+) = \frac{d^3\gamma}{dt^3}(t_N^+) = 0.$$

As an immediate consequence of propositions 4.5 and 4.6 of the previous section, we conclude, that if in the system of equations (50)-(51), we consider the smoothing parameter λ converging to $+\infty$, we will obtain a smooth line that best fits the given data set of points at the given instants of time, respectively, if we consider λ converging to zero, we will obtain a smoothing spline that interpolates the given data set of points.

As in Reinsch [19], we can integrate explicitly equation (50) in each interval $[t_i, t_{i+1}[$, $i = 0, \dots, N-1$, and deduce a method for computing the solution of (50)-(51).

We can write γ explicitly in the form

$$\gamma(t) = \begin{cases} a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3, & t_i \le t < t_{i+1} \\ a_{N-1} + b_{N-1}(t - t_{N-1}) + & \\ c_{N-1}(t - t_{N-1})^2 + d_{N-1}(t - t_{N-1})^3, & t_{N-1} \le t \le t_N \end{cases}, \tag{52}$$

where $a_i, b_i, c_i, d_i \in \mathbb{R}^n$, for $i = 0, \dots, N-1$.

Inserting (52) in (51) we obtain relations in the spline coefficients.

By differentiating γ in order to t, we obtain, successively,

$$\begin{aligned} \frac{d\gamma}{dt}(t) &= b_i + 2c_i(t - t_i) + 3d_i(t - t_i)^2, & t_i \le t < t_{i+1} \\ \frac{d^2\gamma}{dt^2}(t) &= 2c_i + 6d_i(t - t_i), & t_i \le t < t_{i+1} \\ \frac{d^3\gamma}{dt^3}(t) &= 6d_i, & t_i \le t < t_{i+1}, \end{aligned}$$

for $i = 0, \ldots, N-2$. The expressions for $t \in [t_{N-1}, t_N]$ are equivalent.

Now it is convenient to introduce the following positive quantities

$$h_i = t_{i+1} - t_i, \;\; i = 0, \ldots, N-1.$$

For $k = 2$ in the system of equations (51), we deduce

$$\begin{aligned}
&\frac{d^2\gamma}{dt^2}(t_0) = 2c_0 \Longrightarrow c_0 = 0,\\
&\frac{d^2\gamma}{dt^2}(t_N) = 2c_{N-1} + 6d_{N-1}h_{N-1} \Longrightarrow d_{N-1} = -\frac{c_{N-1}}{3h_{N-1}}\\
&\frac{d^2\gamma}{dt^2}(t_{i+1}^+) - \frac{d^2\gamma}{dt^2}(t_{i+1}^-) = 2c_{i+1} - 2c_i - 6d_ih_i \Longrightarrow d_i = \frac{c_{i+1} - c_i}{3h_i},\\
&\hspace{22em} i = 0, \ldots, N-2.
\end{aligned}$$

Introducing the new variable c_N, by considering $c_N = 0$, the equality

$$d_i = \frac{c_{i+1} - c_i}{3h_i}, \tag{53}$$

is valid for $i = 0, \ldots, N-1$.

Now, for $k = 0$, in the system of equations (51), we obtain

$$\gamma(t_{i+1}^+) - \gamma(t_{i+1}^-) = 0 \iff b_i = \frac{a_{i+1} - a_i}{h_i} - c_ih_i - d_ih_i^2, \tag{54}$$

for $i = 0, \ldots, N-2$.

For $k = 1$, in the system of equations (51),

$$\frac{d\gamma}{dt}(t_i^+) - \frac{d\gamma}{dt}(t_i^-) = 0 \iff b_i = b_{i-1} + 2c_{i-1}h_{i-1} + 3d_{i-1}h_{i-1}^2, \tag{55}$$

for $i = 1, \ldots, N-1$.

Using equation (55) for $i = N-1$, we can obtain b_{N-1} and introducing the new variable a_N given by

$$a_N = \gamma(t_N) = a_{N-1} + b_{N-1}h_{N-1} + c_{N-1}h_{N-1}^2 + d_{N-1}h_{N-1}^3,$$

the formula

$$b_i = \frac{a_{i+1} - a_i}{h_i} - c_ih_i - d_ih_i^2, \tag{56}$$

is valid for $i = 0, \ldots, N-1$.

Inserting (56) and (53) into (55), we get

$$a_{i-1}\left(\frac{1}{h_{i-1}}\right)+a_i\left(-\frac{1}{h_i}-\frac{1}{h_{i-1}}\right)+a_{i+1}\left(\frac{1}{h_i}\right)=c_{i-1}\left(\frac{h_{i-1}}{3}\right)+$$
$$c_i\left(\frac{2}{3}h_i+\frac{2}{3}h_{i-1}\right)+c_{i+1}\left(\frac{h_i}{3}\right),$$

for $i=1,\ldots,N-1$, which is equivalent to the system of equations

$$\left\{\begin{array}{l} a_0\left(\frac{1}{h_0}\right)+a_1\left(-\frac{1}{h_1}-\frac{1}{h_0}\right)+a_2\left(\frac{1}{h_1}\right)=c_1\left(\frac{2}{3}h_0+\frac{2}{3}h_1\right)+c_2\left(\frac{h_1}{3}\right)\\ a_1\left(\frac{1}{h_1}\right)+a_2\left(-\frac{1}{h_2}-\frac{1}{h_1}\right)+a_3\left(\frac{1}{h_2}\right)=c_1\left(\frac{h_1}{3}\right)+c_2\left(\frac{2}{3}h_1+\frac{2}{3}h_2\right)+c_3\left(\frac{h_2}{3}\right)\\ \vdots\\ a_{N-2}\left(\frac{1}{h_{N-2}}\right)+a_{N-1}\left(-\frac{1}{h_{N-2}}-\frac{1}{h_{N-1}}\right)+a_N\left(\frac{1}{h_{N-1}}\right)=c_{N-2}\left(\frac{h_{N-2}}{3}\right)+\\ \qquad c_{N-1}\left(\frac{2}{3}h_{N-2}+\frac{2}{3}h_{N-1}\right)\end{array}\right. . \tag{57}$$

The system of equations (57) can be written as the matrix equation

$$Q^\top A=TC, \tag{58}$$

where $Q^\top$ is the tridiagonal matrix with $N-1$ rows and $N+1$ columns,

$$Q^\top=\begin{bmatrix} \frac{1}{h_0} & -\frac{1}{h_0}-\frac{1}{h_1} & \frac{1}{h_1} & & \cdots & & 0\\ 0 & \frac{1}{h_1} & -\frac{1}{h_1}-\frac{1}{h_2} & \frac{1}{h_2} & \cdots & & 0\\ \vdots & & & & \vdots & & \vdots\\ 0 & & \cdots & & \frac{1}{h_{N-2}} & -\frac{1}{h_{N-2}}-\frac{1}{h_{N-1}} & \frac{1}{h_{N-1}} \end{bmatrix}, \tag{59}$$

$$T=\begin{bmatrix} \frac{2}{3}(h_0+h_1) & \frac{h_1}{3} & \cdots & 0\\ \frac{h_1}{3} & \frac{2}{3}(h_1+h_2) & \frac{h_2}{3} & 0\\ \vdots & & \ddots & \vdots\\ 0 & \cdots & \frac{h_{N-2}}{3} & \frac{2}{3}(h_{N-2}+h_{N-1}) \end{bmatrix}\in\mathbb{R}^{(N-1)\times(N-1)},$$

$$A=\begin{bmatrix} a_0^\top\\ a_1^\top\\ \vdots\\ a_N^\top \end{bmatrix}=\begin{bmatrix} a_0^1 & a_0^2 & \cdots & a_0^n\\ a_1^1 & a_1^2 & \cdots & a_1^n\\ \vdots & \vdots & \vdots & \vdots\\ a_N^1 & a_N^2 & \cdots & a_N^n \end{bmatrix}\in\mathbb{R}^{(N+1)\times n}, \tag{60}$$

and finally

$$C=\begin{bmatrix} c_1^\top \\ c_2^\top \\ \vdots \\ c_{N-1}^\top \end{bmatrix}=\begin{bmatrix} c_1^1 & c_1^2 & \cdots & c_1^n \\ c_2^1 & c_2^2 & \cdots & c_2^n \\ \vdots & \vdots & \vdots & \vdots \\ c_{N-1}^1 & c_{N-1}^2 & \cdots & c_{N-1}^n \end{bmatrix}\in \mathbb{R}^{(N-1)\times n}. \tag{61}$$

Now, for $k=3$, in the system of equations (51), we have

$$\frac{d^3\gamma}{dt^3}(t_i^+)-\frac{d^3\gamma}{dt^3}(t_i^-)=\frac{1}{\lambda}\left(q_i-\gamma(t_i)\right) \iff (d_i-d_{i-1})=\frac{1}{\lambda}\left(q_i-\gamma(t_i)\right),$$

and attending to (53), we obtain

$$c_{i-1}\left(\frac{1}{h_{i-1}}\right)+c_i\left(-\frac{1}{h_i}-\frac{1}{h_{i-1}}\right)+c_{i+1}\left(\frac{1}{h_{i+1}}\right)=\frac{1}{2\lambda}\left(q_i-a_i\right), \tag{62}$$

for $i=1,\dots,N-1$.

Since $\dfrac{d^3\gamma}{dt^3}(t_N)=6d_{N-1}$, we have

$$c_{N-1}\left(\frac{1}{h_{N-1}}\right)=\frac{1}{2\lambda}\left(q_N-a_N\right). \tag{63}$$

Therefore, we can write the system of equations (62)-(63) as the matrix equation

$$QC=\frac{1}{2\lambda}\left(P-A\right), \tag{64}$$

where

$$P=\begin{bmatrix} q_0^\top \\ q_1^\top \\ \vdots \\ q_N^\top \end{bmatrix}=\begin{bmatrix} q_0^1 & q_0^2 & \cdots & q_0^n \\ q_1^1 & q_1^2 & \cdots & q_1^n \\ \vdots & \vdots & \vdots & \vdots \\ q_N^1 & q_N^2 & \cdots & q_N^n \end{bmatrix}\in \mathbb{R}^{(N+1)\times n}, \tag{65}$$

and A and C are given respectively by (60) and (61).

Then, according to (58)-(64), conditions (50)-(51) are equivalent to the following system of matrix equations

$$\begin{cases} Q^\top A = TC \\ QC=\dfrac{1}{2\lambda}\left(P-A\right) \end{cases}. \tag{66}$$

From this, we can get explicitly the matrices A and C, in terms of the known matrices T, Q and P. Indeed, since the matrix $T + 2\lambda Q^\top Q$ is symmetric and positive definite, the system of equations (66) is equivalent to the following:

$$\begin{cases} Q^\top (P - 2\lambda QC) = TC \\ P - 2\lambda QC = A \end{cases} \iff \begin{cases} (T + 2\lambda Q^\top Q)\, C = Q^\top P \\ P - 2\lambda QC = A \end{cases}$$
$$\iff \begin{cases} C = (T + 2\lambda Q^\top Q)^{-1} Q^\top P \\ A = P - Q\left(\frac{1}{2\lambda}T + Q^\top Q\right)^{-1} Q^\top P \end{cases}.$$

The remaining coefficients of the curve γ defined in (52), can now be computed from these using the identities (53) and (56). So, we conclude that, for each $\lambda > 0$, there is a unique Euclidean cubic spline that minimizes the functional J.

Clearly, when λ tends to $+\infty$, the matrix A tends to

$$A = P - Q\left(Q^\top Q\right)^{-1} Q^\top P. \tag{67}$$

In this case $Q^\top A = 0$, and since the matrix T is also invertible, it follows from the first equation in (66) that $C = 0$. This, together with (53) and (56), imply that γ is a straight line in each subinterval. The required differentiability implies that γ is a straight line. At this point it is not obvious that this straight line is the solution of the linear regression problem formulated at the end of section 2.

Theorem 4.8. *The straight line in the Euclidean space $\mathbb{R}^n$ obtained from the solution of equations (50)-(51) when the parameter λ converges to $+\infty$ is the linear regression line given by the classical least squares problem presented in section 2.*

Proof: According to theorem 2.2, given in section 2, there exists a unique straight line $\gamma : [0,1] \longrightarrow \mathbb{R}^n$ that minimizes the functional

$$E(\gamma) = \frac{1}{2}\sum_{i=0}^{N} \|\gamma(t_i) - q_i\|^2. \tag{68}$$

In order to show that this is precisely the straight line obtained as the limiting process (when $\lambda \to +\infty$) of the best fitting cubic spline, we have to reparameterize γ so that $\gamma(t_i) = a_i$. So, we are now looking for vectors

$a_0, \ldots, a_N$ and b_0 in $\mathbb{R}^n$, so that the straight line given by

$$\gamma(t) = \begin{cases} a_i + b_0(t - t_i), & t_i \leq t < t_{i+1}, \quad i = 0, \ldots, N-1 \\ a_N, & t = t_N \end{cases}, \tag{69}$$

minimizes the functional (68).

Since γ is continuous in the whole interval $[t_0, t_N]$, the following restrictions have to be fulfilled:

$$a_{i+1} = a_i + b_0(t_{i+1} - t_i), \quad i = 0, \ldots, N-1. \tag{70}$$

Using the notation $h_i = t_{i+1} - t_i$, we obtain the following expression for b_0 in terms of the a_i's and some restrictions for these coefficients:

$$b_0 = \frac{a_{i+1} - a_i}{h_i}, \tag{71}$$

and

$$\frac{a_{i+1} - a_i}{h_i} = \frac{a_i - a_{i-1}}{h_{i-1}}, \quad i = 1, \ldots N-1, \tag{72}$$

which is equivalent to

$$\frac{1}{h_{i-1}} a_{i-1} + \left(-\frac{1}{h_{i-1}} - \frac{1}{h_i} \right) a_i + \frac{1}{h_i} a_{i+1} = 0, \quad i = 1, \ldots N-1. \tag{73}$$

Attending to the matrices Q and A given respectively by (59) and (60), equation (73) is equivalent to the matrix equation

$$Q^\top A = 0,$$

that is, the matrix A, of the coefficients a_i, lies in the vector subspace of $\mathbb{R}^{(N+1)\times n}$,

$$\mathcal{M} = \left\{ A \in \mathbb{R}^{(N+1)\times n} \ : \ Q^\top A = 0 \right\}.$$

With this new parametrization, the functional E reduces to

$$E(a_0, \ldots, a_N) = \frac{1}{2} \sum_{i=0}^{N} d^2(a_i, q_i) = \frac{1}{2} \sum_{i=0}^{N} \left(a_i^\top a_i - 2a_i^\top q_i + q_i^\top q_i \right),$$

or in matrix form

$$E(A) \ = \ \frac{1}{2} \operatorname{tr}\left(AA^\top - 2AP^\top + PP^\top \right) = \frac{1}{2} d^2(A, P),$$

where the last d denotes the distance induced by the Frobenius norm on $\mathbb{R}^{(N+1)\times n}$. Since P is a given matrix, the linear regression line in $\mathbb{R}^n$ can now be formulated as the following optimization problem in the Riemannian manifold $\mathcal{M}$,

$$(\mathcal{P}_5) \qquad \min_{A\in\mathcal{M}} \frac{1}{2}\operatorname{tr}\left(AA^\top - 2AP^\top\right).$$

We will show that A is a solution of this optimization if and only if $A = P - Q\left(Q^\top Q\right)^{-1}Q^\top P$. This will be enough to prove the theorem.

Since $\mathcal{M}$ is a vector subspace of $\mathbb{R}^{(N+1)\times n}$, the tangent map of $\mathcal{M}$ at any point $A \in \mathcal{M}$ coincides with $\mathcal{M}$. So, $A \in \mathcal{M}$ is a critical point for the function

$$\begin{array}{rccl} F: & \mathcal{M} & \longrightarrow & \mathbb{R} \\ & A & \longmapsto & F(A) = \frac{1}{2}\operatorname{tr}\left(AA^\top - 2AP^\top\right), \end{array}$$

if and only if

$$T_AF(Y) = 0, \ \ \forall\, Y \in \mathcal{M}.$$

Using standard techniques from optimization on manifolds, already used previously, we can write

$$\begin{aligned} T_AF(Y) &= \frac{1}{2}\operatorname{tr}\left(YA^\top + AY^\top - 2YP^\top\right) \\ &= \operatorname{tr}\left((A-P)Y^\top\right), \end{aligned}$$

and therefore, $A \in \mathcal{M}$ is a critical point for the function F if and only if

$$\operatorname{tr}\left((A-P)Y^\top\right) = 0, \ \ \forall\, Y \in \mathcal{M}.$$

Now, since A and Y belong to $\mathcal{M}$ (that is, $Q^\top A = 0$ and $Y^\top Q = 0$), we can conclude that

$$\begin{aligned} \operatorname{tr}\left(Q\left(Q^\top Q\right)^{-1}Q^\top(A-P)Y^\top\right) &= \operatorname{tr}\left(Q\left(Q^\top Q\right)^{-1}Q^\top AY^\top\right) - \\ &\quad\ \operatorname{tr}\left(Q\left(Q^\top Q\right)^{-1}Q^\top PY^\top\right) \\ &= -\operatorname{tr}\left(Y^\top Q\left(Q^\top Q\right)^{-1}Q^\top P\right) \\ &= 0, \end{aligned}$$

and therefore,

$$\begin{aligned} \operatorname{tr}\left((A-P)Y^\top\right) &= \operatorname{tr}\left((A-P)Y^\top - Q\left(Q^\top Q\right)^{-1}Q^\top(A-P)Y^\top\right) \\ &= \operatorname{tr}\left(\left(A - P + Q\left(Q^\top Q\right)^{-1}Q^\top P\right)Y^\top\right). \end{aligned}$$

In conclusion, $\operatorname{tr}((A-P)Y^\top) = 0, \forall Y \in \mathcal{M}$ if and only if

$$\operatorname{tr}\left(\left(A - P + Q\left(Q^\top Q\right)^{-1} Q^\top P\right) Y^\top\right) = 0, \forall Y \in \mathcal{M}.$$

Now, just notice that $A - P + Q\left(Q^\top Q\right)^{-1} Q^\top P \in \mathcal{M}$, to conclude that $A \in \mathcal{M}$ is a critical point for the function F if and only if

$$A - P + Q\left(Q^\top Q\right)^{-1} Q^\top P = 0.$$

∎

We would like to thank M. Camarinha and K. Hüeper for helpful discussions while preparing this manuscript.

References

[1] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, New York, 1975.

[2] M. Camarinha. *The Geometry of Cubic Polynomials on Riemannian Manifolds.* PhD Thesis, Department of Mathematics, University of Coimbra, Portugal, 1996.

[3] M. Camarinha, F. Silva Leite, and P. Crouch. Splines of class C^k on Non-Euclidean Spaces. *IMA Journal of Mathematical Control and Information*, 12:399–410, 1995.

[4] M. P. Carmo. *Riemannian Geometry.* Mathematics: Theory and Applications, Birkhäuser, Boston, 1992.

[5] Martin F. Clyde, Shan Sun, and Magnus Egerstedt. Optimal Control, Statistics and Path Planning. *Mathematical and Computer Modeling*, (33):237–253, 2001.

[6] P. Crouch and F. Silva Leite. Geometry and the Dynamic Interpolation Problem. *Proc. American Control Conference, Boston*, pages 1131–1137, 1991.

[7] P. Crouch and F. Silva Leite. The Dynamic Interpolation Problem: on Riemannian Manifolds, Lie Groups and Symmetric Spaces. *Journal of Dynamical and Control Systems*, 1(2):177–202, 1995.

[8] S. Helgason. *Differential Geometry, Lie Groups and Symmetric Spaces.* Academis Press, London, 1978.

[9] U. Helmke and J. Moore. *Optimization and Dynamical Systems.* CCES, Springer-Verlag, London, 1994.

[10] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, New York, 1991.

[11] P. E. Jupp and J. T. Kent. Fitting Smooth Paths to Spherical Data. *Appl. Statist.*, 36(1):34–46, 1987.

[12] H. Karcher. Riemannian Center of Mass and Mollifier Smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

[13] A. K. Krakowski. *Geometrical Methods of Inference.* PhD Thesis, Department of Mathematics and Statistics, University of Western, Australia, 2002.

[14] P. Lancaster and K. Salkauskas. *Curve and Surface Fitting.* Academic Press, 1990.

[15] J. W. Milnor. *Morse Theory.* Princeton University Press, Princeton, New Jersey, 1963.

[16] Maher Moakher. A Differential Geometric Approach to the Arithmetic and Geometric Means of Operators in some Symmetric Spaces. *SIAM. J. Matrix Anal. Appl.*, page (to appear).

[17] L. Noakes, G. Heinzinger, and B. Paden. Cubic Splines on Curved Spaces. *IMA Journal of Mathematics Control and Information*, 6:465–473, 1989.
[18] K. Nomizu. Invariant affine Connections on Homogeneous Spaces. *Amer. J. Math.*, 76:33–65, 1954.
[19] C. H. Reinsch. Smoothing by Spline Functions. *Numerische Mathematik*, 10:177–183, 1967.
[20] D. H. Sattinger and O. L. Weaver. *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics*. Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1980.
[21] Shan Sun, Magnus Egerstedt, and Martin F. Clyde. Control Theoretic Smoothing Splines. *IEEE Trans. Aut. Control*, 45(12):2271–2279, 2000.
[22] C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds*, volume 297. Mathematics and Its Applications, Kluwer Academic, Dordrecht, 1994.

Luís Miguel Machado
Departamento de Matemática, Universidade de Trás-os-Montes e Alto Douro, 5000-911 Vila Real, Portugal
E-mail address: lmiguel@utad.pt
URL: http://home.utad.pt/∼lmiguel

F. Silva Leite
Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal
E-mail address: fleite@mat.uc.pt
URL: http://www.mat.uc.pt/∼fleite