A gradient method for geodesic data fitting on some symmetric Riemannian manifolds

Quentin Rentmeesters
Department of Mathematical Engineering
Université catholique de Louvain
Avenue Georges Lemaitre, 4
B-1348 Louvain-la-Neuve, Belgium

Abstract—In this paper, a method to compute the best geodesic approximation of a set of points that belong to a Riemannian manifold is proposed. This method is based on a gradient descent technique on the tangent bundle of the manifold. An expression for the gradient is derived using the theory of Jacobi fields and an efficient numerical technique is proposed to compute these Jacobi fields. The presented approach is valid on any locally symmetric space, and the sphere $S^2$, the set of symmetric positive definite matrices $P_n^+$, the special orthogonal group $SO(n)$, and the Grassmann manifold Grass$(n,p)$ are considered.

Index Terms—Geodesics, Riemannian manifolds, tangent bundle, Jacobi fields, curvature

I. INTRODUCTION

In many areas of signal and image processing, we have to deal with data that belong to nonlinear spaces or manifolds, see [1], [2] and the references therein. To extract some information out of these data, many authors have proposed different optimization schemes to compute the Karcher mean, i.e., a generalized notion of mean that is defined using the Riemannian distance, see [3]. Moreover, the problem of finding a smooth curve that interpolates a set of time labeled data points is discussed in [4], and a variational approach to find a smooth curve that approximates a set of time labeled data points is presented in [5] and [6]. In this paper, we propose a method to compute the generalization of the best linear regression model. More precisely, we want to find the best geodesic, i.e., a generalization of a parameterized straight line that approximates a given set of data on the manifold.

This problem can be formulated as an optimization problem. Let $q_1, ..., q_m$ be some points on a Riemannian manifold $(\mathcal{M}, g)$ and corresponding time instants $0 = t_1 < ... < t_m = 1$. In this paper, the $t_i$’s are assumed to be given. The goal is then to find a geodesic curve $t \mapsto \gamma^*(t)$ that minimizes the sum of the squared Riemannian distances $d$ between the data points $q_i$’s and their corresponding points on the geodesic curve $\gamma(t_i)$, see Fig. 1. This yields to the following optimization problem:

$$\gamma^* = \arg\min_{\gamma \text{ is a geodesic in } \mathcal{M}} \frac{1}{2} \sum_{i=1}^{m} d(q_i, \gamma(t_i))^2. \quad (1)$$

Such a best geodesic approximation can be used to get more insight on the relative position of the data on the manifold but also to identify the parameters of constant velocity dynamical models. In fact, many methods have been proposed to deal with data that possess a dynamic behavior. For instance, this happens in the subspace tracking problem, see [7] and the references therein, and in filtering problems on the special orthogonal group $SO(3)$, see [8]. In these situations a constant velocity model or geodesic model is often assumed (at least on some time window) to represent the dynamic behavior of the data on the manifold, and the $t_i$’s are the sampling instants. Our geodesic data fitting technique can be useful to assess the validity of such a model or to identify the parameters of a noise model as in [7].

A geodesic is a curve of zero acceleration, i.e., $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$, where $\nabla$ stands for the Riemannian connection. So, a geodesic is the solution of a second order differential equation. This equation has a unique solution on some time window, if $p = \gamma(0) \in \mathcal{M}$ and $v = \dot{\gamma}(0) \in T_p \mathcal{M}$ are specified. Thus, a geodesic can be uniquely represented by a point $(p,v)$ in the tangent bundle $\mathcal{T} \mathcal{M} = \{(p,v)|p \in \mathcal{M}, v \in T_p \mathcal{M}\}$. Notice that $\mathcal{T} \mathcal{M}$ is a manifold of dimension $2 \dim(\mathcal{M})$. The problem (1) with this parameterization of the geodesics in $\mathcal{M}$ is described in [5], where the authors give a first order necessary condition of optimality on the sphere $S^2$ and on the special orthogonal group $SO(n)$. This necessary condition is a nonlinear system in $p$ and $v$ and no algorithm is given to solve this system.
In this paper, a gradient descent technique on $\mathcal{T}M$ is proposed to find a stationary point of the objective function in (1). This gradient descent method is introduced in section II. Our geometric approach to derive the gradient is described in section III. Section IV explains how to diagonalize efficiently the Jacobi operator, which is one of the key computational steps in the computation of the gradient. Section V presents some numerical experimentations and section VI concludes.

II. STEEPEST DESCENT ON THE TANGENT BUNDLE

In this section, some notations and concepts about the tangent bundle of $\mathcal{M}$ are introduced and our gradient descent technique on $\mathcal{T}M$ is described. The tangent space of $\mathcal{T}M$ at $(p,v)$ denoted by $T_{(p,v)} \mathcal{T}M$ is a vector space of dimension $2 \dim(M)$. This tangent space can be expressed as the direct sum of two sub-vector spaces of dimension $\dim(M)$ called the vertical space and the horizontal space. A tangent vector field on $\mathcal{T}M$ is a vector space of dimension $\dim(M)$.

Theorem 2. The differential of $h$, $Dh : T_{(p,v)} \mathcal{T}M \to T_{r(p,v,v_t)}M$ is given by

$$Dh((p,v))[(\Delta p, \Delta v)] = J(t_i),$$

where $J(t_i)$ is the Jacobi operator, which is one of the key computational steps in the computation of the gradient. In order to compute the differential of $h$, we need to define two functions $g$ and $h$ such that $f = g \circ h$ as follows

$$g : \mathcal{M} \to \mathbb{R}, \; p \mapsto \frac{1}{2} d(q_i, \gamma(p,v,t_i))^2, \quad (4)$$

where $\gamma(p,v,t)$ is a geodesic curve such that $\gamma(p,v,0) = p$ and $\dot{\gamma}(p,v,0) = v$. In order to compute the differential of $f$, we need to define two functions $g$ and $h$ such that $f = g \circ h$ as follows

$$h : \mathcal{T}M \to \mathcal{M}, \; (p,v) \mapsto \gamma(p,v,t_i). \quad (5)$$

The following two theorems give the differential of $h$ and $g$.

**Theorem 1.** The differential of $g$ at $c \in \mathcal{M}$, $Dg : T_c \mathcal{M} \to T_c \mathbb{R} \cong \mathbb{R}$ is given by

$$Dg(c)[\Delta c] = g_c(\Delta c, \beta(1)), \quad (6)$$

where $\beta(t)$ is the geodesic curve such that $\beta(0) = q_i$ and $\beta(1) = c$. For a proof, see [3].

**Theorem 2.** The differential of $h$, $Dh : T_{(p,v)} \mathcal{T}M \to T_{\gamma(p,v,t_i)}M$ is given by

$$Dh((p,v))[(\Delta p, \Delta v)] = J(t_i), \quad (7)$$

### Algorithm 1 Gradient descent

1. Given a required precision $\epsilon$ and an initial iterate $(p_0, v_0) \in \mathcal{T}M$;
2. set $k = 0$, $\alpha = 0.5$ and set $(p_k, v_k) = (p_0, v_0)$;
3. until $G(\text{grad} f((p_k, v_k)), \text{grad} f((p_k, v_k))) < \epsilon$ do
4. compute $\text{grad} f((p_k, v_k))$; the gradient of $f$ with respect to the Sasaki metric;
5. set $\alpha = 2\alpha$ and $(p_{\text{new}}, v_{\text{new}}) = r(c(p_k, v_k))(-\alpha \text{grad} f)$;
6. set $(p_{\text{new}}, v_{\text{new}}) < f((p_k, v_k))$ do
7. set $\alpha = \frac{\alpha}{2}$;
8. set $(p_{\text{new}}, v_{\text{new}}) = r(c(p_k, v_k))(-\alpha \text{grad} f)$;
9. end
10. set $k = k + 1$;
11. set $(p_k, v_k) = (p_{\text{new}}, v_{\text{new}})$;
12. end
13. return $(p_k, v_k)$.

$\alpha(0) = (\Delta p, \Delta v)$. The following curve in $\mathcal{T}M$ has been chosen:

$$\alpha(t) = (\exp_p(t \Delta p), \Gamma_p \circ \exp_p(t \Delta p)(v + t \Delta v)), \quad (3)$$

where $\exp_p : T_p \mathcal{M} \to \mathcal{M}$ is the exponential map and $\Gamma_p \circ q : T_p \mathcal{M} \to T_p \mathcal{M}$ is the parallel transport along the geodesic joining $p$ to $q$. As a retraction, we take $r(c(p_k, v_k))((\Delta p, \Delta v)) = \alpha(1)$. This means that we move along the geodesic curve (in $\mathcal{M}$) starting at $p$ in the direction $\Delta p$ and parallel translate the vector $v + \Delta v$ along this geodesic curve.

III. COMPUTATION OF THE GRADIENT

In this section, the gradient of the $i$-th term of the objective function (1) is computed. More precisely, let us consider the following function:

$$f : T_{(p,v)} \mathcal{T}M \to \mathbb{R}, \; (p,v) \mapsto \frac{1}{2} d(q_i, \gamma(p,v,t_i))^2, \quad (4)$$

The following two theorems give the differential of $h$ and $g$.

**Theorem 1.** The differential of $g$ at $c \in \mathcal{M}$, $Dg : T_c \mathcal{M} \to T_c \mathbb{R} \cong \mathbb{R}$ is given by

$$Dg(c)[\Delta c] = g_c(\Delta c, \beta(1)), \quad (6)$$

where $\beta(t)$ is the geodesic curve such that $\beta(0) = q_i$ and $\beta(1) = c$. For a proof, see [3].

**Theorem 2.** The differential of $h$, $Dh : T_{(p,v)} \mathcal{T}M \to T_{\gamma(p,v,t_i)}M$ is given by

$$Dh((p,v))[(\Delta p, \Delta v)] = J(t_i), \quad (7)$$

$\alpha(0) = (\Delta p, \Delta v)$. The following curve in $\mathcal{T}M$ has been chosen:

$$\alpha(t) = (\exp_p(t \Delta p), \Gamma_p \circ \exp_p(t \Delta p)(v + t \Delta v)), \quad (3)$$

where $\exp_p : T_p \mathcal{M} \to \mathcal{M}$ is the exponential map and $\Gamma_p \circ q : T_p \mathcal{M} \to T_p \mathcal{M}$ is the parallel transport along the geodesic joining $p$ to $q$. As a retraction, we take $r(c(p_k, v_k))((\Delta p, \Delta v)) = \alpha(1)$. This means that we move along the geodesic curve (in $\mathcal{M}$) starting at $p$ in the direction $\Delta p$ and parallel translate the vector $v + \Delta v$ along this geodesic curve.

III. COMPUTATION OF THE GRADIENT

In this section, the gradient of the $i$-th term of the objective function (1) is computed. More precisely, let us consider the following function:

$$f : T_{(p,v)} \mathcal{T}M \to \mathbb{R}, \; (p,v) \mapsto \frac{1}{2} d(q_i, \gamma(p,v,t_i))^2, \quad (4)$$

where $\gamma(p,v,t)$ is a geodesic curve such that $\gamma(p,v,0) = p$ and $\dot{\gamma}(p,v,0) = v$. In order to compute the differential of $f$, let us define two functions $g$ and $h$ such that $f = g \circ h$ as follows

$$g : \mathcal{M} \to \mathbb{R}, \; c \mapsto \frac{1}{2} d(q_i, c)^2, \quad (5)$$

$$h : \mathcal{T}M \to \mathcal{M}, \; (p,v) \mapsto \gamma(p,v,t_i). \quad (6)$$

The following two theorems give the differential of $h$ and $g$.

**Theorem 1.** The differential of $g$ at $c \in \mathcal{M}$, $Dg : T_c \mathcal{M} \to T_c \mathbb{R} \cong \mathbb{R}$ is given by

$$Dg(c)[\Delta c] = g_c(\Delta c, \beta(1)), \quad (6)$$

where $\beta(t)$ is the geodesic curve such that $\beta(0) = q_i$ and $\beta(1) = c$. For a proof, see [3].

**Theorem 2.** The differential of $h$, $Dh : T_{(p,v)} \mathcal{T}M \to T_{\gamma(p,v,t_i)}M$ is given by

$$Dh((p,v))[(\Delta p, \Delta v)] = J(t_i), \quad (7)$$
where $J(t)$ is the Jacobi field along the geodesic curve $t \mapsto \gamma(p, v, t)$ such that
\[ D^2_t J(t) + R(J(t), \dot{\gamma}(p, v, t), \dot{\gamma}(p, v, t)) = 0, \]
where $D_t J(t)$ is the covariant derivative of the vector field $J(t)$ along the geodesic curve $t \mapsto \gamma(p, v, t)$ and $R$ stands for the curvature endomorphism defined by:
\[ R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z. \]


Using these two theorems, the differential of $f$ at $(p, v) \in TM$ in the direction $(p, v)$ is then given by
\[ Df((p, v))[(\Delta p, \Delta v)] = g_p(\nabla_{p,v} (\Delta p), \nabla_{p,v} (\Delta v)) = g_p(\Gamma_{\gamma(p,v,t)\rightarrow p}(J(t)), \Gamma_{\gamma(p,v,t)\rightarrow p}(\dot{\beta}(1))) \]
\[ = g_p(\Gamma_{\gamma(p,v,t)\rightarrow p}(J(1)), \dot{\beta}(1)) \]
\[ (8) \]
since the parallel transport is an isometry. Notice that $\dot{\beta}(1) = -\exp^{-1}_{\gamma(p,v,t)}(q_i)$ where $\exp^{-1}(q) : TM \rightarrow T_p M$ stands for the log-mapping, i.e., the inverse of the exponential map. This mapping can be computed efficiently on the spaces we have considered, see Section IV. Consequently, to compute the gradient, it remains to solve the Cauchy problem (7).

Since the curvature is a linear operator, (7) is a second order time-varying linear differential equation in $J(t)$. But if we assume that $\dot{\beta}(1)$ is symmetric, it admits an orthogonal basis of eigenvectors $E_k$. Let $E_{\dim(M)}$ be a basis of eigenvectors and $\lambda_1 = 0, ..., \lambda_{\dim(M)}$ be the corresponding eigenvalues, i.e.,
\[ L_{\dot{\gamma}}(E_k) = \lambda_k E_k \]
for $k = 1, ..., \dim(M)$.

Then, the Jacobi equation (7) can be expressed in this basis of eigenvectors. Let $J(t) = \sum_{k=1}^{\dim(M)} u_k(t) E_k(t)$ where $E_k(t)$ is the extension of $E_k = E_k(0)$ along the curve $\gamma$ by parallel transport. By replacing $J(t)$ in (7), one obtains
\[ \sum_{k=1}^{\dim(M)} (\ddot{u}_k(t) + u_k(t)\lambda_k) E_k(t) = 0, \]
since $D_t E_k(t) = 0$. So, we end up with $\dim(M)$ decoupled second order time-invariant linear ode’s whose solutions are of the form
\[ u_k(t) = c_k(t) \Delta p_k + s_k(t) \Delta v_k, \]
where
\[ \cdot c_k(t) = \cos(\sqrt{-\lambda_k} t) \quad \text{and} \quad s_k(t) = \frac{\sinh(\sqrt{-\lambda_k} t)}{\sqrt{-\lambda_k}} \quad \text{if} \quad \lambda_k < 0, \]
and $\Delta p_k = g_p(\Delta p, E_k)$ and $\Delta v_k = g_p(\Delta v, E_k)$. With this notation, we can write
\[ J(t) = \sum_{k=1}^{\dim(M)} (c_k(t) \Delta p_k + s_k(t) \Delta v_k) E_k(t), \]
and the differential (8) becomes
\[ g_p(\sum_{k=1}^{\dim(M)} (c_k(t) \Delta p_k + s_k(t) \Delta v_k) E_k(0)), \Gamma_{\gamma(p,v,t)\rightarrow p}(\dot{\beta}(1)) \]
\[ = g_p(\Delta p, \sum_{k=1}^{\dim(M)} c_k(t) g_p(\Gamma_{\gamma(p,v,t)\rightarrow p}(\dot{\beta}(1)), E_k) E_k) \]
\[ + g_p(\Delta v, \sum_{k=1}^{\dim(M)} s_k(t) g_p(\Gamma_{\gamma(p,v,t)\rightarrow p}(\dot{\beta}(1)), E_k) E_k). \]

Consequently, the gradient $(gp, gv)$ in the Sasaki metric is given by
\[ (gp, gv) = (\sum_{k=1}^{\dim(M)} c_k(t) g_p(T, E_k) E_k, \sum_{k=1}^{\dim(M)} s_k(t) g_p(T, E_k) E_k), \]
\[ (10) \]
where $T = \Gamma_{\gamma(p,v,t)\rightarrow p}(\dot{\beta}(1))$. The procedure to compute the gradient of $f$ defined in (4) is then the following:

1) computation of $\dot{\beta}(1)$ using the log-mapping,
2) computation of $\Gamma_{\gamma(p,v,t)\rightarrow p}(\dot{\beta}(1))$, the parallel transport of $\dot{\beta}(1)$ at $T_p M$ along the geodesic curve $t \mapsto \gamma(p, v, t)$,
3) diagonalization of the operator $L_{\dot{\gamma}}$,
4) computation of $(gp, gv) \in T(p, v) TM$ using (10).

To compute the gradient using this procedure and so, to implement Algorithm 1, it remains to find an efficient way to diagonalize $L_{\dot{\gamma}}$ and to compute $(gp, gv)$ in (10). This is the subject of the next section.

**IV. DIAGONALIZATION OF THE CURVATURE**

The goal of this section is to describe an efficient way to diagonalize the Jacobi operator $L_{\dot{\gamma}}$ in (9). This is done here by exploiting the specific structure of the algebraic representation of the curvature endomorphism $R$ on symmetric spaces, see [13]. Mainly, it is possible to find an explicit solution for the eigenvalues and eigenvectors of the Jacobi operator $L_{\dot{\gamma}}$ in terms of some factorization of a matrix $A$ representing the tangent vector $\dot{\gamma}(p, v, 0)$. A way to represent points and formulas for the geodesics, the parallel transports and the log-mappings are first recalled for the sphere $S^2$, the set of symmetric positive definite matrices $P^+_n$, the special orthogonal group $SO(3)$, and the Grassmann manifold Grass$(n, p)$. Then, explicit expressions for the eigenpairs of the Jacobi operator are given.
A. The sphere $S^2$

Let $S^2 = \{p \in \mathbb{R}^3| p^\top p = 1\}$ be the unit sphere in $\mathbb{R}^3$ endowed with the ambient metric: $g_p(X,Y) = X^\top Y$. The tangent space at a point $p \in S^2$ is

$$T_pS^2 = \{v \in \mathbb{R}^3| p^\top v = 0\}.$$ 

Using these representations of points and tangent vectors, one has the following formulas:

$$\exp_p(v) = p \cos(|v|) + \frac{v}{|v|} \sin(|v|),$$

$$\exp_p^{-1}(q) = \frac{(I_3 - pp^\top)q}{\sqrt{1 - (p^\top q)^2}} \sin(p^\top q),$$

$$\Gamma_{p \rightarrow q}(w) = (-p \sin(|v|) + \frac{v}{|v|} \cos(|v|))v^\top w + (I_3 - vv^\top)w \text{ where } v = \exp_p^{-1}(q).$$

The curvature endomorphism is given by, see [12],

$$R(X,Y)Z = (Y^\top Z)X - (X^\top Z)Y.$$ 

Thus, the Jacobi operator (9) on the sphere $S^2$ becomes:

$$L_{A,t}(X) = R(X,A)A = A^\top AX - AA^\top X = \|A\|^2(I - \frac{A^\top A}{\|A\|^2})X.$$ 

So, the eigenpairs $(\lambda, X)$ of $L_A$ are $(0, 1_A)$, and $(\|A\|^2, V)$, where $V$ is a unit tangent vector orthogonal to $A$.

B. The set of symmetric positive definite matrices $P_n^+$

Let $P_n^+$ be the set of symmetric positive definite matrices endowed with the affine invariant metric

$$g_p(X,Y) = \text{trace}(Xp^{-1}Yp^{-1}),$$

defined on the tangent space at $p$:

$$T_pP_n^+ = \{v \in \mathbb{R}^{n \times n}| v = v^\top\}.$$ 

The exponential map, the parallel transport and the logarithm are given by, see [2], [14],

$$\exp_p(v) = p^{1/2} \exp((p^{-1/2}v)^{(p^{-1/2})})p^{1/2},$$

$$\exp_p^{-1}(q) = p^{1/2} \log((p^{-1/2}q)^{(p^{-1/2})})p^{1/2},$$

$$\Gamma_{p \rightarrow q}(w) = p^{1/2}w^{p^{-1/2}}p^{1/2}w^{p^{-1/2}}p^{1/2}w \text{ where } r = \exp((p^{-1/2}v)^{(p^{-1/2})}) \text{ and } v = \exp_p^{-1}(q),$$

where Exp and Log stand for the matrix exponential and the logarithm.

Notice that the action $x \mapsto p^{-1/2}xp^{-1/2}$ maps $p$ to the identity $I$ and it is an isometry. So, one can “translate” the problem to the identity using this isometry at each iteration. Thus, we can assume that $p = I$ and the diagonalization of the Jacobi operator (9) will be only required at the identity. The curvature endomorphism at the identity $I$ is given by, see [13],

$$R_I(X_t, Y_t)Z_I = \frac{1}{4}[Z_I, [X_t, Y_t]].$$

where $[A, B] = AB - BA$ stands for the matrix Lie bracket and $X_t, Y_t, Z_t$ are tangent vectors at the identity, i.e., symmetric matrices. So, one obtains the following expression for the Jacobi operator (9) on $P_n^+$:

$$L_{A,t}(X_I) = \frac{1}{4}(-A^2X_I + 2A_IX_I - X_IA_I^2).$$

Since $A_I$ is symmetric, it can be diagonalized by an orthogonal transformation denoted by $A_I = Ud Ud^\top$ where $D = \text{diag}(d_1, \ldots, d_n)$. Using this, it can be shown that the eigenvalues $\lambda$ and the corresponding eigenvectors $X$ of the Jacobi operator on $P_n^+$ are

- $\lambda = 0$, $X = Ue_i e_i^\top U^\top$ for $1 \leq i \leq n$;
- $\lambda = -\frac{1}{4}(d_i - d_j)^2$, $X = U\frac{1}{\sqrt{2}}(e_i e_j^\top + e_j e_i^\top)U^\top$ for $1 \leq i < j \leq n$;

where $e_i$ stands for the $i$-th identity vector.

C. The special orthogonal group $SO(3)$

Let $SO(3) = \{p \in \mathbb{R}^{3 \times 3}|p^\top p = I_3, \ det(p) = 1\}$ be the set of rotation matrices with the bi-invariant metric

$$g(X,Y) = \frac{1}{2}\text{trace}(X^\top Y).$$

The Lie algebra or the tangent space at the identity is given by

$$so(3) = \{v \in \mathbb{R}^{3 \times 3}| v^\top = -v\}.$$ 

Notice that a tangent vector at any point can be represented by an element of the Lie algebra, see [12]. Using this representation, we have the following formulas:

$$\exp_p(v) = p \exp(v),$$

$$\exp_p^{-1}(q) = \log(p^\top q),$$

$$\Gamma_{p \rightarrow q}(w) = q^\top p \exp(v/2)w \exp(v/2) \text{ where } v = \exp_p^{-1}(q).$$

The curvature endomorphism on a Lie group with a bi-invariant metric is expressed by (see [15])

$$R(X,Y)Z = \frac{1}{4}[Z, [X,Y]].$$

Thus, the Jacobi operator in (9) is $L_{A,t}(X) = \frac{1}{4}(-A^2X + 2AXA - XA^2)$. Since $A \in so(3)$ is skew-symmetric, we can consider its Schur decomposition $A = U(\beta e_3 e_3^\top - e_2 e_2^\top)U^\top$ and the eigenpairs of the Jacobi operator are described by

- $\lambda = 0, U(e_3 e_3^\top - e_2 e_2^\top)U^\top$;
- $\lambda = \frac{\beta^2}{4}, U(e_2 e_2^\top - e_1 e_1^\top)U^\top, U(e_3 e_3^\top - e_1 e_1^\top)U^\top$.

Notice that all these eigenvectors are orthogonal and of unit norm with respect to the bi-invariant metric $g(X,Y)$ in (12).

D. The Grassmann manifold $\text{Grass}(n, p)$

A point on the Grassmann manifold, i.e., a subspace, can be represented by the column space of an $n \times p$ orthogonal matrix $P$ and a tangent vector at $P$ by an $n \times p$ matrix $V$ such that $V^\top P = 0$, see [16]. Notice that the dimension of the tangent space, i.e., the dimension of the manifold, is $p(n - p)$. The following Riemannian metric:

$$g_P(X,Y) = \text{trace}(X^\top Y),$$

7144
turns Grass\((n,p)\) into a symmetric space, see [13]. Using these representations for the subspaces and for the tangent vectors, we have the following formulas (see [16], [7]):
\[
\exp_P(V) = (PW \cos(\Sigma) + U \sin(\Sigma)) W^T
\]
where \(V = U\Sigma W^T\) (compact SVD),
\[
\exp_P^{-1}(Q) = W_2 \Sigma W_1^T
\]
with
\[
\begin{bmatrix}
X^T Q \\
(I_n - XX^T) Q
\end{bmatrix} = \begin{bmatrix}
W_1 \cos(\Sigma) Z^T \\
W_2 \sin(\Sigma) Z^T
\end{bmatrix}
\]
(CS decomposition),
\[
\Gamma_{p \rightarrow Q}(T) = (-XW \sin(\Sigma) + U \cos(\Sigma)) U^T T \\
+ (I_n - UU^T) T
\]
where \(\exp_P^{-1}(Q) = U\Sigma W^T\) (compact SVD).

Let \(X, Y\) and \(Z\) be \(n \times p\) matrices representing tangent vectors at \(P \in \text{Grass}(n,p)\). According to [17], the curvature endomorphism is given by
\[
R(X,Y)Z = (XY^T - YX^T)Z + Z(Y^T(X - XX^T)Y)
\]
and thus, the Jacobi operator is
\[
L_A(X) = XA^T A - 2AX^T A + AA^T X,
\]
where \(A\) is an \(n \times p\) matrices representing \(\dot{\gamma}(p,v,0)\). Let \(A = U\Sigma W^T\) where \(\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)\) be the compact SVD of \(A\) and define \(U_\perp \in \mathbb{R}^{n \times (n-2p)}\) such that \([P][U_\perp]\) is an \(n \times n\) orthogonal matrix. It can be shown that the eigenpairs \((\lambda, X)\) of the Jacobi operator (9) are given by
- \(\lambda = 0, X = UD_1 W^T\) with \(D_i = e_i e_i^T\) for \(1 \leq i \leq p\);
- \(\lambda = (\sigma_i - \sigma_j)^2, X = US_i>j W^T\) with \(S_{ij} = \frac{1}{\sqrt{2}}(e_i e_j + e_j e_i)\) for \(1 \leq i < j \leq p\);
- \(\lambda = (\sigma_i + \sigma_j)^2, X = UR_i>j W^T\) with \(R_{ij} = \frac{1}{\sqrt{2}}(e_i e_j - e_j e_i)\) for \(1 \leq i < j \leq p\);
- \(\lambda = \sigma_i^2, X = U_\perp e_i e_i W^T\) for \(1 \leq i \leq p\) and \(1 \leq j \leq (n-2p)\);
where \(e_i\) is the \(i\)-th identity vector in \(\mathbb{R}^p\) and \(\tilde{e}_j\) the \(j\)-th identity vector in \(\mathbb{R}^{n-2p}\). To get an acceptable numerical complexity, one of the key points is to avoid the computation of \(U_\perp\). This is particularly important in applications related to image processing that use high dimensional Grassmann manifolds (typically \(n \approx 1000, p \approx 10\)). This can be done, since only an orthogonal projection on \(U_\perp\) is required to compute (10). In fact, for \(\lambda = \sigma_i^2\), the vectorized form of \(X = U_\perp \tilde{e}_j e_i W^T\) is \((W \otimes U_\perp) \text{vec}(\tilde{e}_j e_i^\top)\), and the components of the gradient in (10), in \(\text{col}(U_\perp)\), are of the form:
\[
(W \otimes U_\perp) \begin{bmatrix}
\sigma_i^2 \\
\vdots \\
\sigma_p^2
\end{bmatrix} \otimes \text{vec}(T)
\]
where \(T = \Gamma_{\gamma(p,v,t_i) \rightarrow p}(\beta(1))\). Using \((A \otimes B)(C \otimes D) = (AC \otimes BD)\) and \((W \otimes U_\perp)^\top = (W^T \otimes U_\perp)^\top\), (13) becomes
\[
(WCW^T \otimes U_\perp U_\perp^T) \text{vec}(T) = (WCW^T \otimes (I_n - [P][U][U]^\top))^\top \text{vec}(T),
\]
whose matrix form is
\[
(I_n - [P][U][P]^\top)T(WCW^T).
\]
This last formula has a computational complexity of \(O(np^2)\).

V. NUMERICAL EXAMPLES

To illustrate the method, we have first picked 3 points on the sphere \(S^2\). The best geodesic approximation of these 3 points is presented in Fig. 2. In \(\mathbb{R}^n\), the best geodesic approximation passes through the two following points:
\[
\bar{t} = \frac{1}{m} \sum_{i=1}^m t_i \quad \gamma(\bar{t}) = \frac{1}{m} \sum_{i=1}^m q_i, \quad (14)
\]
\[
\bar{\gamma}(\bar{t}) = \frac{1}{\sum_{i=1}^m t_i} \sum_{i=1}^m t_i \gamma_i, \quad (15)
\]
Notice that \(\bar{\gamma}(\bar{t})\) is the center of mass of the data points, and \(\gamma(\bar{t})\) is a weighted mean where the \(t_i\)'s can be seen as the mass attached to each \(q_i\). This mean (14) and this weighted mean (15) defined in \(\mathbb{R}^n\) can be generalized to our setting using the Karcher mean. In [5], the authors have proved that the best geodesic passes through the Karcher mean of the data points if the data belong to an abelian subgroup of \(SO(n)\) and they believe that this property does not hold in general. As shown in Fig. 2, the best geodesic does neither pass through the Karcher mean \(\kappa_1\) nor to the weighted Karcher mean \(\kappa_2\). But this is still interesting because the geodesic passing through these two points \(\kappa_1\) and \(\kappa_2\) can be a good initial condition for our gradient descent technique. In fact, our numerical experiments on the sphere \(S^2\) have shown that the Karcher mean is close to the best geodesic if the data points are not located too far apart.

Our best geodesic fitting method was also tested on the Grassmann manifold with 10 data points. Table I shows that the number of iterations of our gradient technique, required to reach an accuracy of \(\epsilon = 10^{-6}\), does not depend on the dimension of the Grassmann manifold. This is interesting...
TABLE I

<table>
<thead>
<tr>
<th>( \epsilon = 10^{-6} ) for different values of ( n ) and ( p ) on ( \text{Grass}(n,p) ).</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>326</td>
</tr>
<tr>
<td>344</td>
</tr>
</tbody>
</table>

since it frequently occurs in object tracking problems on a video sequence that \( n \approx 1000 \gg p \).

To illustrate the filtering capability of the proposed method, 5 data points \( p_1, \ldots, p_5 \) on \( P_2^+ \), i.e., the set of ellipsoids in \( \mathbb{R}^2 \), were generated on a geodesic curve according to the model:

\[
p_k = \exp_p(t_k v) \text{ for } t_k = k \cdot 0.1,
\]

where \( v \) is a tangent vector at \( p \in P_2^+ \). To perturb these points, for each point \( p_k \), a Gaussian random vector of mean 0 was generated on \( T_p P_2^+ \). Then, this random vector was parallel translated along the geodesic curve \( t \mapsto \exp_p(tv) \) to \( T_{p_k} P_2^+ \) and a perturbed version of \( p_k \) was computed by moving in the direction of this tangent vector along a geodesic curve. This ensures that all the points \( p_k \) are perturbed with the same probability distribution with respect to the Riemannian metric (11) since the parallel transport is an isometry.

To reduce the noise, our geodesic fitting method was applied to these perturbed data. The points \( p_1, \ldots, p_5 \) are drawn in blue in Fig.3. Since the filtered data points are close to the true data points \( p_k \), the true geodesic is almost recovered.

VI. CONCLUSIONS AND FURTHER WORK

A gradient descent technique has been presented to solve the problem of best geodesic data fitting on some symmetric spaces of interest in practical applications. Since this gradient technique requires an efficient way to diagonalize the Jacobi operator, explicit solutions for the eigenvalues and eigenvectors of this operator have been given. Notice that these explicit solutions can also be used to build efficiently the Hessian of the Riemannian distance function. The study of a second order method for geodesic data fitting and the uniqueness of the solution of (1) in function of the data distribution are the subject of current research.

REFERENCES


