2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

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

Quentin Rentmeesters
Department of Mathematical Engineering
Université catholique de Louvain
Avenue Georges Lemaître, 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 $SO(3)$ and the Grassmann manifold $\mathrm{Grass}(n,p)$ 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 $0 \leq t_1 < ... < t_m \leq 1$. 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:

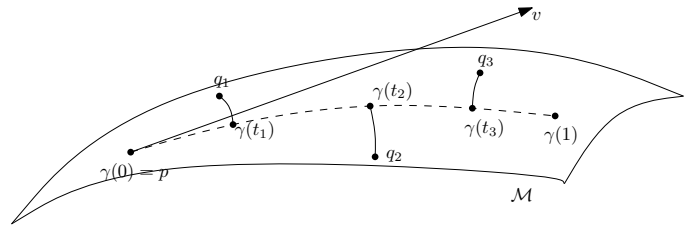

Fig. 1. Problem illustration. The dashed line is the geodesic curve we want to find and the continuous lines are geodesics whose length equals $d(q_i, \gamma(t_i))$.

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

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 ∇ 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 $\mathcal{TM} = \{(p,v) | p \in \mathcal{M},\ v \in T_p\mathcal{M}\}$. Notice that $\mathcal{TM}$ 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.

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its author(s).

In this paper, a gradient descent technique on $\mathcal{TM}$ 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 $\mathcal{TM}$ is described. The tangent space of $\mathcal{TM}$ at (p, v) denoted by $T_{(p,v)}\mathcal{TM}$ is a vector space of dimension 2 dim$(\mathcal{M})$. This tangent space can be expressed as the direct sum of two sub-vector spaces of dimension dim$(\mathcal{M})$ called the vertical space and the horizontal space. A tangent vector $V \in T_{(p,v)}\mathcal{TM}$ is vertical if it corresponds to a variation of v only, p remaining fixed. Equivalently, a tangent vector V is horizontal if it corresponds to a variation of p only, v remaining constant, i.e., parallel. Notice that the horizontal space depends on the definition of a connection on $\mathcal{M}$, the Riemannian connection in our case. This splitting of $T_{(p,v)}\mathcal{TM}$ makes it possible to represent a tangent vector $V \in T_{(p,v)}\mathcal{TM}$ by two tangent vectors $\Delta p, \Delta v \in T_p\mathcal{M}$: Δp representing a variation of p and Δv a variation of v. That is why, a tangent vector $V \in T_{(p,v)}\mathcal{TM}$ is denoted by a couple $(\Delta p, \Delta v)$. To turn $\mathcal{TM}$ into a Riemannian manifold, the Sasaki metric

$$G((\Delta p_1, \Delta v_1), (\Delta p_2, \Delta v_2))_{(p,v)} = g_p(\Delta p_1, \Delta p_2) + g_p(\Delta v_1, \Delta v_2) \quad (2)$$

is used. This metric makes horizontal tangent vectors $(\Delta p, 0)$ orthogonal to vertical tangent vectors $(0, \Delta v)$, see [9] for more details. Using this metric, we can define the gradient of the a function $f : \mathcal{TM} \to \Re$, which is a tangent vector $\mathrm{grad} f(p, v) = (gp, gv) \in T_{(p,v)}\mathcal{TM}$ such that

$$\begin{aligned} Df((p, v))[(\Delta p, \Delta v)] &\triangleq \frac{d}{dt} f(c(t))|_{t=0}, \\ &= G((gp, gv), (\Delta p, \Delta v))_{(p,v)}, \end{aligned}$$

where $c(t)$ is a curve in $\mathcal{TM}$ such that $\frac{d}{dt}c(t)|_{t=0} = (\Delta p, \Delta v)$.

To find the gradient of f, we need to compute the differential of f. The expressions for the differential of the objective function in (1) and the gradient are derived in section III. Using this gradient, it is possible to implement a first order optimization algorithm as in [10]. Algorithm 1 presents a simple version of such a technique.

In step 8, this algorithm requires a retraction, i.e., a smooth mapping $r_{(p,v)} : T_{(p,v)}\mathcal{TM} \to \mathcal{TM}$ that satisfies: $r_{(p,v)}((0, 0)) = (p, v)$ and $Dr_{(p,v)}((0, 0)) = I$ where I denotes the identity operator. Such a retraction can be built using a smooth curve $\alpha(t)$ in $\mathcal{TM}$, i.e., a curve in $\mathcal{M}$ and a vector field along this curve, such that $\alpha(0) = (p, v)$ and $\dot{\alpha}(0) = (\Delta p, \Delta v)$. The following curve in $\mathcal{TM}$ has been chosen:

Algorithm 1 Gradient descent

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

$$\alpha(t) = (\exp_p(t\Delta p), \Gamma_{p\to\exp_p(t\Delta p)}(v + t\Delta v)), \quad (3)$$

where $\exp_p : T_p\mathcal{M} \to \mathcal{M}$ is the exponential map and $\Gamma_{p\to q} : T_p\mathcal{M} \to T_q\mathcal{M}$ is the parallel transport along the geodesic joining p to q. As a retraction, we take $r_{(p,v)}((\Delta p, \Delta v)) = \alpha(1)$. This means that we move along the geodesic curve (in $\mathcal{M}$) starting at p in the direction Δp 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:

$$f : \mathcal{TM} \to \Re,\ (p, v) \mapsto \frac{1}{2} d(q_i, \gamma(p, v, t_i))^2, \quad (4)$$

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

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

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

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

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

$$Dg(c)[\Delta c] = g_c(\Delta c, \dot{\beta}(1)),$$

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, $Dh : T_{(p,v)}\mathcal{TM} \to T_{\gamma(p,v,t_i)}\mathcal{M}$ is given by*

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

where $J(t_i)$ is the Jacobi field along the geodesic curve $t \mapsto \gamma(p,v,t)$ such that

$$D_t^2 J(t) + R(J(t), \dot\gamma(p,v,t), \dot\gamma(p,v,t)) = 0, \quad (7)$$
$$J(0) = \Delta p,$$
$$\dot J(0) = \Delta v,$$

where $D_t J(t)$ is the covariant derivative of the vector field $J(t)$ along $\gamma(p,v,t)$ and R stands for the curvature endomorphism defined by:

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

See [11] or [12] for more details.

Using these two theorems, the differential of f at $(p,v) \in \mathcal{TM}$ in the direction $(\Delta p, \Delta v)$ is then given by

$$\begin{aligned} Df((p,v))[(\Delta p, \Delta v)] &= g_{\gamma(p,v,t_i)}(J(t_i), \dot\beta(1)) \\ &= g_p(\Gamma_{\gamma(p,v,t_i)\to p}(J(t_i)), \Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1))) \end{aligned} \quad (8)$$

since the parallel transport is an isometry. Notice that $\dot\beta(1) = -\exp^{-1}_{\gamma(p,v,t_i)}(q_i)$ where $\exp_p^{-1}(q) : \mathcal{M} \to T_p\mathcal{M}$ 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 $\mathcal{M}$ is a symmetric Riemannian manifold, which implies $\nabla R = 0$, i.e., the curvature tensor is parallel along $\gamma(p,v,t)$, then (7) becomes a second order constant coefficients linear differential equation that is easy to solve, see [12]. Let us define the Jacobi operator

$$L_{\dot\gamma(p,v,0)} : J \mapsto R(J, \dot\gamma(p,v,0))\dot\gamma(p,v,0), \quad (9)$$

that acts on tangent vectors J at p. Since this operator is symmetric, it admits an orthogonal basis of eigenvectors and real eigenvalues. Let $E_1 = \frac{\dot\gamma}{\|\dot\gamma\|}, ..., E_{\dim(\mathcal{M})}$ be a basis of eigenvectors and $\lambda_1 = 0, ..., \lambda_{\dim(\mathcal{M})}$, the corresponding eigenvalues, i.e.,

$$L_{\dot\gamma}(E_k) = \lambda_k E_k \text{ for } k = 1, ..., \dim(\mathcal{M}).$$

Then, the Jacobi equation (7) can be expressed in this basis of eigenvectors. Let $J(t) = \sum_{k=1}^{\dim(\mathcal{M})} u_k(t) E_k(t)$ where $E_k(t)$ is the extension of $E_k = E_k(0)$ along the curve γ by parallel transport. By replacing $J(t)$ in (7), one obtains

$$\sum_{k=1}^{\dim(\mathcal{M})} (\ddot u_k(t) + u_k(t)\lambda_k)E_k(t) = 0,$$

since $D_t E_k(t) = 0$. So, we end up with $\dim(\mathcal{M})$ 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

- $c_k(t) = 1$ and $s_k(t) = t$ if $\lambda_k = 0$,
- $c_k(t) = \cos(\sqrt{\lambda_k} t)$ and $s_k(t) = \frac{\sin(\sqrt{\lambda_k} t)}{\sqrt{\lambda_k}}$ if $\lambda_k > 0$,
- $c_k(t) = \cosh(\sqrt{-\lambda_k} t)$ and $s_k(t) = \frac{\sinh(\sqrt{-\lambda_k} t)}{\sqrt{-\lambda_k}}$ if $\lambda_k < 0$,

where $\Delta p_k = g_p(\Delta p, E_k)$ and $\Delta v_k = g_p(\Delta v, E_k)$. With this notation, we can write

$$J(t_i) = \sum_{k=1}^{\dim(\mathcal{M})} (c_k(t_i)\Delta p_k + s_k(t_i)\Delta v_k)E_k(t_i),$$

and the differential (8) becomes

$$\begin{aligned} & g_p\Big(\sum_{k=1}^{\dim(\mathcal{M})} (c_k(t_i)\Delta p_k + s_k(t_i)\Delta v_k)E_k(0), \Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1))\Big) \\ &= g_p\Big(\Delta p, \sum_{k=1}^{\dim(\mathcal{M})} c_k(t_i) g_p(\Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1)), E_k)E_k\Big) \\ &+ g_p\Big(\Delta v, \sum_{k=1}^{\dim(\mathcal{M})} s_k(t_i) g_p(\Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1)), E_k)E_k\Big). \end{aligned}$$

Consequently, the gradient (gp, gv) in the Sasaki metric is given by

$$(gp, gv) = \Big(\sum_{k=1}^{\dim(\mathcal{M})} c_k(t_i) g_p(T, E_k)E_k, \sum_{k=1}^{\dim(\mathcal{M})} s_k(t_i) g_p(T, E_k)E_k\Big), \quad (10)$$

where $T = \Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1))$. 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 $\Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1))$, the parallel transport of $\beta(1)$ at $T_p\mathcal{M}$ along the geodesic curve $t \mapsto \gamma(p,v,t)$,
3) diagonalization of the operator $L_{\dot\gamma}$,
4) computation of $(gp, gv) \in T_{(p,v)}\mathcal{TM}$ 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 $\mathrm{Grass}(n,p)$. Then, explicit expressions for the eigenpairs of the Jacobi operator are given.

A. The sphere S^2

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

$$T_p S^2 = \{v \in \Re^3 | p^\top v = 0\}.$$

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

$$\begin{aligned}
\exp_p(v) &= p\cos(\|v\|) + \frac{v}{\|v\|}\sin(\|v\|),\\
\exp_p^{-1}(q) &= \frac{(I_3 - pp^\top)q}{\sqrt{1-(p^\top q)^2}}\text{acos}(p^\top q),\\
\Gamma_{p\to q}(w) &= (-p\sin(\|v\|) + \frac{v}{\|v\|}\cos(\|v\|))v^\top w\\
&\quad + (I_3 - \frac{vv^\top}{\|v\|^2})w \text{ where } v = \exp_p^{-1}(q).
\end{aligned}$$

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: $L_A(X) = R(X,A)A = A^\top AX - AA^\top X = \|A\|^2(I - \frac{AA^\top}{\|A\|^2})X$. So, the eigenpairs (λ, X) of L_A are $(0, \frac{A}{\|A\|})$, 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

$$g_p(X,Y) = \text{trace}(Xp^{-1}Yp^{-1}), \tag{11}$$

defined on the tangent space at p:

$$T_p P_n^+ = \{v \in \Re^{n\times n} | v = v^\top\}.$$

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

$$\begin{aligned}
\exp_p(v) &= p^{1/2}\,\text{Exp}(p^{-1/2}vp^{-1/2})p^{1/2},\\
\exp_p^{-1}(q) &= p^{1/2}\,\text{Log}(p^{-1/2}qp^{-1/2})p^{1/2},\\
\Gamma_{p\to q}(w) &= p^{1/2}rp^{-1/2}wp^{-1/2}rp^{1/2}\\
&\text{where } r = \text{Exp}(p^{-1/2}\frac{v}{2}p^{-1/2}) \text{ and } v = \exp_p^{-1}(q),
\end{aligned}$$

where Exp and Log stand for the matrix exponential and the matrix 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],

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

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

$$L_{A_I}(X_I) = \frac{1}{4}(-A_I^2X_I + 2A_IX_IA_I - X_IA_I^2).$$

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

- $\lambda = 0$, $X = Ue_ie_i^\top U^\top$ for $1 \le i \le n$;
- $\lambda = -\frac{1}{4}(d_i - d_j)^2$, $X = U\frac{1}{\sqrt{2}}(e_ie_j^\top + e_je_i^\top)U^\top$ for $1 \le i < j \le n$;

where e_i stands for the i-th identity vector.

C. The special orthogonal group SO(3)

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

$$g(X,Y) = \frac{1}{2}\text{trace}(X^\top Y). \tag{12}$$

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

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

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:

$$\begin{aligned}
\exp_p(v) &= p\,\text{Exp}(v),\\
\exp_p^{-1}(q) &= \text{Log}(p^\top q),\\
\Gamma_{p\to q}(w) &= q^\top p\,\text{Exp}(v/2)w\,\text{Exp}(v/2) \text{ where } v = \exp_p^{-1}(q).
\end{aligned}$$

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 $L_A(X) = \frac{1}{4}\left(-A^2X + 2AXA - XA^2\right)$. Since $A \in so(3)$ is skew-symmetric, we can consider its Schur decomposition $A = U\beta(e_3e_2^\top - e_2e_3^\top)U^\top$ and the eigenpairs of the Jacobi operator are described by

- $\lambda = 0, U(e_3e_2^\top - e_2e_3^\top)U^\top$;
- $\lambda = \frac{\beta^2}{4}, U(e_2e_1^\top - e_1e_2^\top)U^\top, U(e_3e_1^\top - e_1e_3^\top)U^\top$.

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:

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

turns $\mathrm{Grass}(n,p)$ 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]):

$$\exp_P(V) = (PW\cos(\Sigma) + U\sin(\Sigma))\,W^\top$$
$$\text{where } V = U\Sigma W^\top \text{ (compact SVD)},$$
$$\exp_P^{-1}(Q) = W_2\Sigma W_1^\top \text{ with}$$
$$\begin{bmatrix} X^\top Q \\ (I_n - XX^\top)Q \end{bmatrix} = \begin{bmatrix} W_1\cos(\Sigma)Z^\top \\ W_2\sin(\Sigma)Z^\top \end{bmatrix} \text{ (CS decomposition)},$$
$$\Gamma_{P\to Q}(T) = (-XW\sin(\Sigma) + U\cos(\Sigma))\,U^\top T + (I_n - UU^\top)T$$
$$\text{where } \exp_P^{-1}(Q) = U\Sigma W^\top \text{(compact SVD)}.$$

Let X, Y and Z be $n\times p$ matrices representing tangent vectors at $P \in \mathrm{Grass}(n,p)$. According to [17], the curvature endomorphism is given by

$$R(X,Y)Z = (XY^\top - YX^\top)Z + Z(Y^\top X - X^\top Y),$$

and thus, the Jacobi operator is

$$L_A(X) = XA^\top A - 2AX^\top A + AA^\top X,$$

where A is an $n\times p$ matrices representing $\dot\gamma(p,v,0)$. Let $A = U\Sigma W^\top$ where $\Sigma = \mathrm{diag}(\sigma_1,...,\sigma_p)$ be the compact SVD of A and define $U_\perp \in \Re^{n\times(n-2p)}$ such that $[P|U|U_\perp]$ is an $n\times n$ orthogonal matrix. It can be shown that the eigenpairs (λ, X) of the Jacobi operator (9) are given by

- $\lambda = 0$, $X = UD_iW^\top$ with $D_i = e_ie_i^\top$ for $1\le i\le p$;
- $\lambda = (\sigma_i - \sigma_j)^2$, $X = US_{ij}W^\top$ with $S_{ij} = \frac{1}{\sqrt{2}}(e_ie_j^\top + e_je_i^\top)$ for $1\le i<j\le p$;
- $\lambda = (\sigma_i + \sigma_j)^2$, $X = UR_{ij}W^\top$ with $R_{ij} = \frac{1}{\sqrt{2}}(e_ie_j^\top - e_je_i^\top)$ for $1\le i<j\le p$;
- $\lambda = \sigma_i^2$, $X = U_\perp\tilde e_je_i^\top W^\top$ for $1\le i\le p$ and $1\le j\le (n-2p)$;

where e_i is the i-th identity vector in $\Re^p$ and $\tilde e_j$, the j-th identity vector in $\Re^{n-2p}$. 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 $X = U_\perp\tilde e_je_i^\top W^\top$ is $(W\otimes U_\perp)\mathrm{vec}(\tilde e_je_i^\top)$, and the components of the gradient in (10), in $\mathrm{col}(U_\perp)$, are of the form:

$$(W\otimes U_\perp)\left(\underbrace{\begin{bmatrix} c_1^2 & & \\ & \ddots & \\ & & c_p^2 \end{bmatrix}}_{C}\otimes I_{(n-2p)}\right)(W\otimes U_\perp)^\top\mathrm{vec}(T), \tag{13}$$

where $T = \Gamma_{\gamma(p,v,t_i)\to p}(\dot\beta(1))$. Using $(A\otimes B)(C\otimes D) = (AC\otimes BD)$ and $(W\otimes U_\perp)^\top = (W^\top\otimes U_\perp^\top)$, (13) becomes

$$(WCW^\top\otimes U_\perp U_\perp^\top)\mathrm{vec}(T) = (WCW^\top\otimes(I_n - [P|U][P|U]^\top)\mathrm{vec}(T),$$

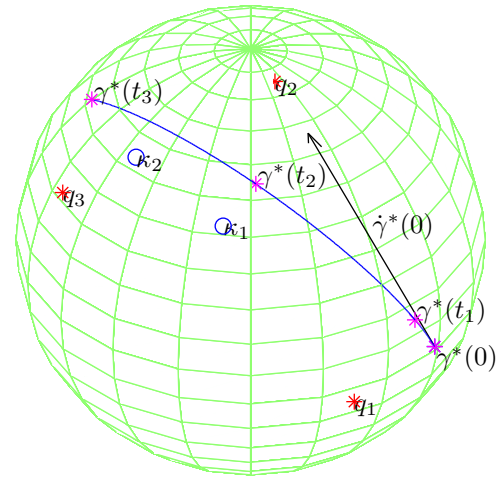

Fig. 2. The data points q_i $1\le i\le 3$ are in red. The best geodesic $\gamma^*(t)$ is the blue curve and the black arrow represents $\dot\gamma^*(0)$. The points $\gamma^*(t_i)$ for $t_1 = 0.1, t_2 = 0.5, t_3 = 0.9$, are in magenta. The Karcher mean of the q_i's is κ_1, and κ_2 is the weighted Karcher mean.

whose matrix form is

$$(I_n - [P|U][P|U]^\top)T(WCW^\top).$$

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 $\Re^n$, the best geodesic approximation passes through the two following points:

$$\bar t = \frac{1}{m}\sum_{i=1}^m t_i \quad \gamma(\bar t) = \frac{1}{m}\sum_{i=1}^m q_i, \tag{14}$$

$$\bar{\bar t} = \frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m t_i^2 \quad \gamma(\bar{\bar t}) = \frac{1}{\sum_{i=1}^m t_i}\sum_{i=1}^m t_iq_i. \tag{15}$$

Notice that $\gamma(\bar t)$ is the center of mass of the data points, and $\gamma(\bar{\bar t})$ 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 $\Re^n$ 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 κ_1 nor to the weighted Karcher mean κ_2. But this is still interesting because the geodesic passing through these two points κ_1 and κ_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

NUMBER OF ITERATIONS REQUIRED TO REACH AN ACCURACY OF $\epsilon = 10^{-6}$ FOR DIFFERENT VALUES OF n AND p ON $\mathrm{Grass}(n, p)$.

p \ n	10	100	1000
2	316	317	321
5	326	334	335
10	-	344	346

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 $p_1, ..., p_5$ on P_2^+, i.e., the set of ellipsoids in $\Re^2$, were generated on a geodesic curve according to the model:

$$p_k = \exp_p(t_k v) \text{ for } t_k = k\ 0.1,$$

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 $p_1, ..., p_5$ 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.

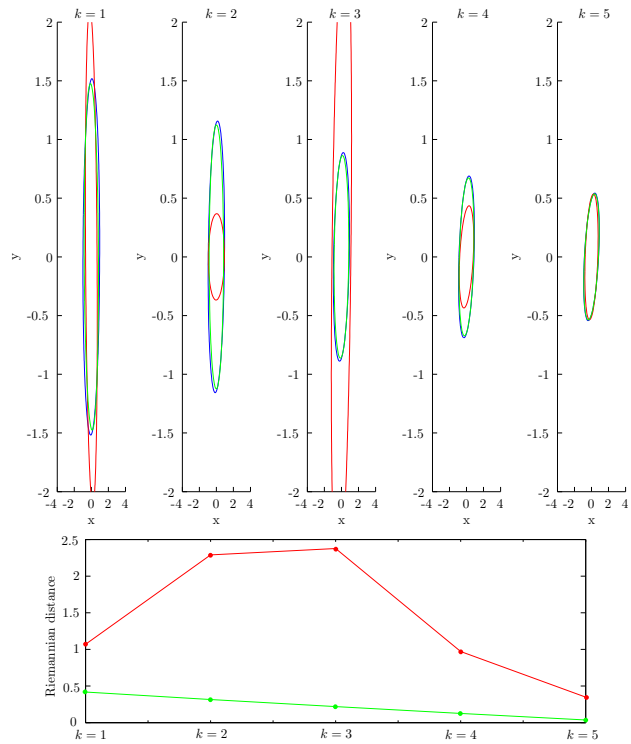

Fig. 3. (upper) The ellipsoids $p_1, ..., p_5$ are in blue. The ellipsoids in red represent the perturbed data and the ellipsoids in green are the filtered data. (lower) Riemannian distance between the p_k's and their perturbed version (in red) and their corresponding filtered version (in green).

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

[1] Q. Rentmeesters, P.-A. Absil, P. Van Dooren, K. Gallivan, and A. Srivastava, "An efficient particle filtering technique on the Grassmann manifold," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 3838–3841.

[2] R. Ferreira, J. Xavier, J.P. Costeira, and V. Barroso, "Newton method for Riemannian centroid computation in naturally reductive homogeneous spaces," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2006, vol. 3.

[3] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.

[4] P. Crouch and F. Silva Leite, "The dynamic interpolation problem: on Riemannian manifolds, Lie groups, and symmetric spaces," *Journal of Dynamical and control systems*, vol. 1, no. 2, pp. 177–202, 1995.

[5] F. Silva Leite and L. Machado, "Fitting smooth paths on Riemannian manifolds.," *International Journal of Applied Mathematics & Statistics,*, vol. 4, no. J06, pp. 25–53, June 2006.

[6] C. Samir, P.-A. Absil, A. Srivastava, and E. Klassen, "A gradient-descent method for curve fitting on Riemannian manifolds," *Foundations of Computational Mathematics*, pp. 1–25, 10.1007/s10208-011-9091-7.

[7] Q. Rentmeesters, P.-A. Absil, and P. Van Dooren, "A filtering technique on the Grassmann manifold," in *19th International Symposium on Mathematical Theory of Networks and Systems*, 2010.

[8] R. Mahony, T. Hamel, and J.M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *Automatic Control, IEEE Transactions on*, vol. 53, no. 5, pp. 1203–1218, 2008.

[9] S. Sasaki, "On the differential geometry of tangent bundles of Riemannian manifolds," *Tohoku Math. J. (2)*, vol. 10, no. 3, pp. 338–354, 1958.

[10] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.

[11] M.P. Do Carmo, *Riemannian geometry*, Birkhauser, 1992.

[12] J. Cheeger and D.G. Ebin, *Comparison theorems in Riemannian geometry*, American Mathematical Society, 2008.

[13] S. Kobayashi and K. Nomizu, "Foundations of Differential Geometry, Vol. 1 & 2," *Interscience Tracts in Pure and Applied Mathematics, Interscience Publ., New-York*, 1969.

[14] P.T. Fletcher and S. Joshi, "Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors," *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pp. 87–98, 2004.

[15] B. O'neill, *Semi-Riemannian geometry: with applications to relativity*, Academic Pr, 1983.

[16] A. Edelman, TA Arias, and ST Smith, "The Geometry of Algorithms with Orthogonality Constraints, Matrix Anal," *Applicant*, vol. 20, pp. 303–353, 1998.

[17] Y.-C. Wong, "Sectional curvatures of Grassmann manifolds," in *Proc Natl Acad Sci U S A.*, May 1968, vol. 60, pp. 75–79.