A Gradient-Descent Method for Curve Fitting on Riemannian Manifolds*

Chafik Samir† P.-A. Absil† Anuj Srivastava‡ Eric Klassen§

Abstract

Given data points $p_0, \ldots, p_N$ on a closed submanifold M of $\mathbb{R}^n$ and time instants $0 = t_0 < t_1 < \ldots < t_N = 1$, we consider the problem of finding a curve γ on M that best approximates the data points at the given instants while being as "regular" as possible. Specifically, γ is expressed as the curve that minimizes the weighted sum of a sum-of-squares term penalizing the lack of fitting to the data points and a regularity term defined, in the first case as the mean squared velocity of the curve, and in the second case as the mean squared acceleration of the curve. In both cases, the optimization task is carried out by means of a steepest-descent algorithm on a set of curves on M. The steepest-descent direction, defined in the sense of the first-order and second-order Palais metric, respectively, is shown to admit analytical expressions involving parallel transport and covariant integral along curves. Illustrations are given in $\mathbb{R}^n$ and on the unit sphere.

Keywords: curve fitting, steepest-descent, Sobolev space, Palais metric, geodesic distance, energy minimization, splines, piecewise geodesic, smoothing, Riemannian center of mass.

1 Introduction

We are interested in the problem of fitting smooth curves to given finite sets of points on Riemannian manifolds. Let $p_0, p_1, \ldots, p_N$ be a finite set of points on a Riemannian manifold M, and let $0 = t_0 < t_1 < ... < t_N = 1$ be distinct and ordered instants of time. The problem of fitting a smooth curve γ on M to the given points at the given times involves two goals of conflicting nature. The first goal is that the curve should fit the data as well as possible, as measured, e.g., by the real-valued function E_{d} defined by:

$$E_{\mathrm{d}}(\gamma) = \sum_{i=0}^{N} d^2(\gamma(t_i), p_i), \tag{1}$$

where d denotes the distance function on the Riemannian manifold M. The second goal is that the curve should be sufficiently "regular", as measured by a function $\gamma \mapsto E_{\mathrm{s}}(\gamma)$ such as (2) or (3) below. We are thus facing an optimization problem with two objective functions—a fitting function E_{d} and a regularity function E_{s}—whose domain is a suitable set of curves on the Riemannian manifold M.

Curve fitting problems on manifolds appear in various applications. To cite but one example, let $(I_i)_{i \leq N}$ be a temporal sequence of images of a 2D or 3D object motion, in which the object can appear and disappear at arbitrary times due to obscuration and other reasons. The task is to

*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 authors. This research was supported in part by AFOSR FA9550-06-1-0324 and ONR N00014-09-10664.

†Department of Mathematical Engineering, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (http://www.inma.ucl.ac.be/{~samir,~absil}).

‡Dept of Statistics, Florida State University, Tallahassee, FL 32306, USA (anuj@stat.fsu.edu).

§Dept of Mathematics, Florida State University, Tallahassee, FL 32306, USA (klassen@math.fsu.edu).

estimate the missing data and recover the motion of the object as well as possible. It is clear that focussing on the first goal (fitting the data) without concern for the second goal (regularity of the curve) would yield poor motion recovery, and that the result is likely to be improved if inherent regularity properties of the object motion are taken into account. The features of interest in such tracking problems are often represented as elements of nonlinear manifolds. An important feature in object tracking and recognition is the shape formed by its silhouette in an image.

1.1 Previous work

One possible way of tackling an optimization problem with two objective functions is to turn it into a classical optimization problem where one of the objective functions becomes *the* objective function and the other one is turned into a constraint.

Let us first discuss the case where the fitting objective function E_{d} is minimized under a regularity constraint. When $M = \mathbb{R}^n$, a classical regularity constraint is to restrict the curve γ to the family of polynomial functions of degree not exceeding m, ($m \leq N$). This least-squares problem cannot be straightforwardly generalized to an arbitrary Riemannian manifold M because the notion of polynomial does not carry to M in an obvious way. An exception is the case $m = 1$; the polynomial functions in $\mathbb{R}^n$ are then straight lines, whose natural generalization on Riemannian manifolds are geodesics. The problem of fitting a geodesic to data on Riemannian manifold M was considered in [MS06] for the case where M is the special orthogonal group $\mathrm{SO}(n)$ or the unit sphere $\mathbb{S}^n$.

The other case is when a regularity criterion E_{s} is optimized under a constraint on E_{d}, in which case it is natural to impose the *interpolation constraint* $E_{\mathrm{d}}(\gamma) = 0$. For example, when $M = \mathbb{R}^n$, minimizing the function E_{s} defined by

$$E_{\mathrm{s},1}(\gamma) = \frac{1}{2}\int_0^1 \|\dot{\gamma}(t)\|^2 \mathrm{d}t \tag{2}$$

yields the piecewise-linear interpolant for the given data points and time instants (this follows from [Mil63, p. 70]), while minimizing

$$E_{\mathrm{s},2}(\gamma) = \frac{1}{2}\int_0^1 \|\ddot{\gamma}(t)\|^2 \mathrm{d}t$$

yields solutions known as cubic splines. (From now on, we will frequently omit the variable t in the integrand when it is clear from the context.) For the case where M is a nonlinear manifold, several results on interpolation can be found in the literature. The generalization of cubic splines to more general Riemannian manifolds was pioneered by Noakes *et al.* [NHP89]. Cubic splines are then defined as curves that minimize the function

$$E_{\mathrm{s},2}(\gamma) = \frac{1}{2}\int_0^1 \left\langle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2} \right\rangle_{\gamma(t)} \mathrm{d}t, \tag{3}$$

where $\frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}$ denotes the (Levi-Civita) second covariant derivative of γ and $\langle \cdot, \cdot \rangle_p$ stands for the Riemannian metric on M at p. (The subscript may be omitted if there is no risk of confusion.) A necessary condition for optimality takes the form of a fourth-order differential equation. This variational approach was followed by other authors, including Crouch and Silva Leite [CS91, CS95]. Splines of class C^k were generalized to Riemannian manifolds by Camarinha *et al.* [CSC95]. Still in the context of interpolation on manifolds, but without a variational interpretation, we mention the literature on splines based on generalized Bézier curves, defined by a generalization to manifolds of the de Casteljau algorithm; see [CKS99, Alt00, PN07]. Recently, Jakubiak *et al.* [JSR06] presented a geometric two-step algorithm to generate splines of an arbitrary degree of smoothness in Euclidean spaces, then extended the algorithm to matrix Lie groups and applied it to generate smooth motions of 3D objects. Another approach to interpolation on manifolds consists of mapping the data points onto the affine tangent space at a particular point of M, then computing an

interpolating curve in the tangent space, and finally mapping the resulting curve back to the manifold. The mapping can be defined, e.g., by a rolling procedure, see [HS07, KDL07].

Another way of tackling an optimization problem with two objective functions is to optimize a weighted sum of the objective functions. Spherical smoothing splines on the two dimensional unit sphere were originally studied by Jupp and Kent [JK87]. This approach was followed on general manifolds by Machado *et al.* [MSH06] using the first-order smoothing term (2) and by Machado and Silva Leite [MS06] for the second-order smoothing term (3).

Specifically, in [MSH06], the objective function is defined to be

$$E_1 \equiv \frac{1}{2}\sum_{i=0}^{N} d^2(\gamma(t_i), p_i) + \frac{\lambda}{2}\int_0^1 \langle \dot\gamma, \dot\gamma\rangle \,\mathrm{d}t,$$

over the class of all piecewise smooth curves $\gamma : [0,1] \to M$, where λ (> 0) is a smoothing parameter. Solutions to this variational problem are piecewise geodesics that best fit the given data. As shown in [MSH06], when λ goes to $+\infty$, the optimal curve converges to a single point which is shown in [MLK10] to be the Riemannian mean of the data points. When λ goes to zero, the optimal curve goes to a broken geodesic on M interpolating the data points.

In [MS06], the objective function is defined to be

$$E_2 \equiv \frac{1}{2}\sum_{i=0}^{N} d^2(\gamma(t_i), p_i) + \frac{\lambda}{2}\int_0^1 \left\langle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2} \right\rangle \mathrm{d}t$$

over a certain set of admissible C^2 curves. The authors give a necessary condition of optimality that takes the form of a fourth-order differential equation involving the covariant derivative and the curvature tensor along with certain regularity conditions at the time instants t_i, $i = 0, \ldots, N$ [MS06, Th. 4.4]. The optimal curves are *approximating cubic splines*: they are approximating because in general $\gamma(t_i)$ differs from p_i, and they are cubic splines because they are obtained by smoothly piecing together segments of cubic polynomials on M, where "cubic polynomial on M" is understood in the sense of Noakes *et al.* [NHP89]. It is also shown in [MS06, Prop. 4.5] that, as the smoothing parameter λ goes to $+\infty$, the optimal curves converge to a geodesic curve on M fitting the given data points at the given instants of time. When λ goes to zero, the approximating cubic spline converges to an interpolating cubic spline [MS06, Prop. 4.6].

1.2 Our approach

In this paper, rather than trying to solve directly the fourth-order differential equation obtained in [MS06] (a feat that is not attempted there, except for $M = \mathbb{R}^n$), we propose to search for an optimizer of the objective function using a steepest-descent method in an adequate set of curves on the Riemannian manifold M. In this section, we present the essence of our approach, and delay the mathematical technicalities until Section 2.

As in [MS06], we consider the problem of minimizing the objective function

$$\begin{aligned} E_2 : \Gamma_2 \to \mathbb{R} : \gamma \mapsto E_2(\gamma) &= E_{\mathrm{d}}(\gamma) + \lambda E_{\mathrm{s},2}(\gamma) \\ &= \frac{1}{2}\sum_{i=0}^{N} d^2(\gamma(t_i), p_i) + \frac{\lambda}{2}\int_0^1 \left\langle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2} \right\rangle \mathrm{d}t, \end{aligned} \tag{4}$$

where Γ_2 is an adequate set of curves on M to be defined in Section 2. The steepest-descent direction for E_2 is defined with respect to the *second-order Palais metric* defined by

$$\langle\langle v, w\rangle\rangle_{2,\gamma} = \langle v(0), w(0)\rangle_{\gamma(0)} + \left\langle \frac{\mathrm{D}v}{\mathrm{d}t}(0), \frac{\mathrm{D}w}{\mathrm{d}t}(0) \right\rangle_{\gamma(0)} + \int_0^1 \left\langle \frac{\mathrm{D}^2 v}{\mathrm{d}t^2}, \frac{\mathrm{D}^2 w}{\mathrm{d}t^2} \right\rangle_{\gamma(t)} \mathrm{d}t, \tag{5}$$

where v and w are tangent vector fields along γ. We assume that M is a closed Riemannian submanifold of $\mathbb{R}^n$—a mild condition in virtue of Nash's isometric embedding theorem—since the

required theory on the Palais metric is only available in this case [Tro77, Prop. 6.1]. Consequently, the Levi-Civita covariant derivative $\frac{\mathrm{D}}{\mathrm{d}t}$ along γ reduces to the classical derivative in $\mathbb{R}^n$ followed by the orthogonal projection onto the tangent space to M at $\gamma(t)$; see, e.g., [Boo03, §VII.2].

As we shall see in Section 4, the choice of the second-order Palais metric (5) ensures that the gradient of E_2 at γ, represented by a vector field G along γ, admits an analytical expression involving parallel transport and covariant integral along γ. This expression makes it possible to implement a steepest-descent algorithm on Γ_2, where the next iterate is obtained from the current iterate γ using a line-search procedure along the path $\tau \mapsto \gamma^\tau$ on Γ_2 defined by

$$\gamma^\tau(t) = \exp_{\gamma(t)}(-\tau G(t));$$

see Section 5. We use an Armijo backtracking procedure, but other stepsize selection methods would be suitable.

We also present a gradient-descent approach for the objective function of [MSH06], namely

$$\begin{aligned} E_1 : \Gamma_1 \to \mathbb{R} : \gamma \mapsto E_1(\gamma) &= E_{\mathrm{d}}(\gamma) + \lambda E_{\mathrm{s},1}(\gamma) \\ &= \frac{1}{2}\sum_{i=0}^{N} d^2(\gamma(t_i), p_i) + \frac{\lambda}{2}\int_0^1 \langle \dot\gamma, \dot\gamma \rangle \,\mathrm{d}t, \end{aligned} \tag{6}$$

where Γ_1 is another adequate set of curves on M defined in Section 2. For E_1, the steepest-descent direction is considered with respect to the *first-order Palais metric* defined by

$$\langle\langle v, w \rangle\rangle_{1,\gamma} = \langle v(0), w(0) \rangle_{\gamma(0)} + \int_0^1 \left\langle \frac{\mathrm{D}v}{\mathrm{d}t}, \frac{\mathrm{D}w}{\mathrm{d}t} \right\rangle_{\gamma(t)} \mathrm{d}t, \tag{7}$$

where v and w are tangent vector fields along γ.

This choice confers a simple expression to the gradient; see Section 3.

Observe that the parameter λ makes it possible to balance between the two conflicting goals mentioned above: when λ is large, a higher emphasis is on the regularity condition relative to the fitting condition, whereas when λ is small, the fitting condition dominates.

The rest of the paper is organized as follows. Section 2 deals with the choice of the curve spaces Γ_1 and Γ_2. An expression for the gradient of E_1, resp. E_2, is given in Section 3, resp. 4. The steepest-descent method is presented in Section 5. Numerical illustrations are given in Section 6 for $M = \mathbb{R}^2$ and $M = \mathbb{S}^2$. Section 7 contains final remarks.

2 Preliminaries

In this section, we exploit results of Palais [Pal63, §13] and Tromba [Tro77, §6] to define the domain Γ of the objective function E in such a way that the gradient of E with respect to the Palais metric is guaranteed to exist and to be unique.

2.1 First-order case

We first consider the objective function E_1 defined in (6). Let I denote the unit interval $[0, 1]$ and let $H^0(I, \mathbb{R}^n)$ denote the set of square integrable functions from I to $\mathbb{R}^n$. The set $H^0(I, \mathbb{R}^n)$ is a Hilbert space under pointwise operations and with the inner product $\langle\langle \cdot, \cdot \rangle\rangle_0$ defined by

$$\langle\langle v, w \rangle\rangle_0 = \int_0^1 \langle v(t), w(t) \rangle \,\mathrm{d}t,$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product in $\mathbb{R}^n$. Let $H^1(I, \mathbb{R}^n)$ denote the set of absolutely continuous maps $\gamma : I \to \mathbb{R}^n$ such that $\dot\gamma \in H^0(I, \mathbb{R}^n)$. Note that absolute continuity is equivalent to requiring that $\dot\gamma(t)$ exists for almost all $t \in I$, that $\dot\gamma$ is summable, and that

$$\gamma(t) = \gamma(0) + \int_0^t \dot\gamma(s) \,\mathrm{d}s.$$

Then $H^1(I, \mathbb{R}^n)$ is a Hilbert space under the inner product $\langle\langle \cdot, \cdot \rangle\rangle_1$ defined by

$$\langle\langle v, w \rangle\rangle_1 = \langle v(0), w(0) \rangle + \int_0^1 \langle \dot{v}(t), \dot{w}(t) \rangle \, \mathrm{d}t \tag{8}$$

This inner product belongs to a class of Riemannian structures proposed by Linnér [Lin03, §3].

Let M be a closed C^{k+4}-submanifold of $\mathbb{R}^n$ ($k \geq 1$). Define $H^1(I, M)$ to be the set of all $\gamma \in H^1(I, \mathbb{R}^n)$ such that $\gamma(I) \subseteq M$. Then $H^1(I, M)$ is a closed C^k-submanifold of the Hilbert space $H^1(I, \mathbb{R}^n)$. We set

$$\Gamma_1 = H^1(I, M), \tag{9}$$

which ensures that E_1 (6) is a well defined C^k map around each $\gamma \in \Gamma_1$ such that, for all i, p_i is in the image of the domain of injectivity of the exponential mapping at $\gamma(t_i)$ (see Lazard and Tits [LT66] for the case where the manifold is a Lie group).

The tangent space to $H^1(I, M)$ at a curve $\gamma \in H^1(I, M)$ is given by

$$T_\gamma H^1(I, M) = \{v \in H^1(I, TM) : v(t) \in T_{\gamma(t)} M \text{ for all } t \in I\},$$

where TM denotes the tangent bundle of M. Moreover, $H^1(I, M)$ is a complete C^k-Riemannian manifold in the Riemannian structure induced on it as a closed C^k-submanifold of $H^1(I, \mathbb{R}^n)$.

Note that the induced Riemannian structure on $H^1(I, M)$ induced by (8) is the "extrinsic" structure given by

$$\langle v(0), w(0) \rangle + \int_0^1 \langle \dot{v}(t), \dot{w}(t) \rangle \, \mathrm{d}t$$

where $\dot{v}$ and $\dot{w}$ are the derivatives in the sense of the embedding space $\mathbb{R}^n$. It thus differs from the "intrinsic" first-order Palais metric defined in (7). However, the extrinsic and intrinsic Riemannian structures are equivalent on bounded sets [Tro77, Prop. 6.1].

From this, it follows that, given $\gamma \in H^1(I, M)$, the tangent space $T_\gamma H^1(I, M)$ endowed with the inner product (7) is a Hilbert space. This fact will be exploited in Section 2.3.

2.2 Second-order case

We now turn to the objective function E_2 defined in (4). Let $H^2(I, \mathbb{R}^n)$ be the set of maps $\gamma : I \to \mathbb{R}^n$ with $\gamma \in H^1(I, \mathbb{R}^n)$ and $\dot{\gamma} \in H^1(I, \mathbb{R}^n)$. Then $H^2(I, \mathbb{R}^n)$ is a vector space under pointwise operations, and the map

$$\Phi : \mathbb{R}^n \times \mathbb{R}^n \times H^0(I, \mathbb{R}^n) \to H^2(I, \mathbb{R}^n) : (\gamma_0, \dot{\gamma}_0, h) \mapsto \gamma,$$

defined by $\gamma(0) = \gamma_0$, $\dot{\gamma}(0) = \dot{\gamma}_0$, $\ddot{\gamma}(t) = h(t)$ for all $t \in I$, is an isomorphism. In $H^2(I, \mathbb{R}^n)$, consider the inner product $\langle\langle \cdot, \cdot \rangle\rangle_2$ defined by

$$\langle\langle v, w \rangle\rangle_2 = \langle v(0), w(0) \rangle + \langle \dot{v}(0), \dot{w}(0) \rangle + \int_0^1 \langle \ddot{v}(t), \ddot{w}(t) \rangle \, \mathrm{d}t.$$

Then Φ is an isometry and $H^2(I, \mathbb{R}^n)$ is a Hilbert space.

Let M be a closed C^{k+4}-submanifold of $\mathbb{R}^n$ ($k \geq 1$). Define $H^2(I, M)$ to be the set of all $\gamma \in H^2(I, \mathbb{R}^n)$ such that $\gamma(I) \subseteq M$. Then, by restricting the proof of [Pal63, Th. 6.6] to $H^2(I, M)$, one obtains that $H^2(I, M)$ is a closed C^k-submanifold of the Hilbert space $H^2(I, \mathbb{R}^n)$. We set

$$\Gamma_2 = H^2(I, M), \tag{10}$$

which ensures that E_2 is well defined. The tangent space to $H^2(I, M)$ at a curve $\gamma \in H^2(I, M)$ is given by

$$T_\gamma H^2(I, M) = \{v \in H^2(I, TM) : v(t) \in T_{\gamma(t)} M \text{ for all } t \in I\}.$$

Given $\gamma \in H^2(I, M)$, consider the mapping

$$\Phi : T_{\gamma(0)} M \times T_{\gamma(0)} M \times H^0(I, T_{\gamma(0)} M) \to T_\gamma H^2(I, M)$$

that maps $(v_0, \dot{v}_0, g)$ to the vector field v along γ defined by

$$v(0) = v_0, \quad \frac{\mathrm{D}v}{\mathrm{d}t}(0) = \dot{v}_0, \quad \frac{\mathrm{D}^2 v}{\mathrm{d}t^2}(t) = P_\gamma^{t \leftarrow 0} g(t),$$

where $P_\gamma^{t \leftarrow 0}$ is the parallel transport along γ. Recall that the parallel transport is an isometry. The map Φ is an isomorphism of vector spaces between its domain and image, and it is an isometry with the obvious metric on the domain and the second-order Palais metric (5) on the image. Since the domain of Φ is a Hilbert space, its image is also a Hilbert space endowed with the inner product (5). Hence the Riesz representation theorem applies.

2.3 Existence and uniqueness of the gradient of E

For $i = 1, 2$, the gradient of E_i at γ is defined to be the unique $G_\gamma \in T_\gamma H^i(I, M)$ that satisfies, for all $w \in T_\gamma H^i(I, M)$,

$$\langle\langle G_\gamma, w \rangle\rangle_{i,\gamma} = \mathrm{D}E_i(\gamma)[w],$$

where $\mathrm{D}E_i(\gamma)[w]$ denotes the derivative of E_i at γ along w. The existence and uniqueness of G are guaranteed by the Riesz representation theorem. We will use the notation $\nabla E_i(\gamma)$ for the gradient of a function E_i at γ, or simply G_γ when E_i and γ are clear from the context.

3 Gradient of E_1 in the first-order Palais metric

We derive an expression for the gradient of $E_1 = E_\mathrm{d} + \lambda E_{\mathrm{s},1}$ (6) over Γ_1 (9) endowed with the first-order Palais metric (7). The gradient evaluated at a curve γ involves the operations of parallel transport and covariant integral along γ.

3.1 Derivative of E_d

We first give an expression for the derivative of the i^{th} term in E_d, namely,

$$f_i : \Gamma_1 \to \mathbb{R} : \gamma \mapsto \frac{1}{2} d^2(\gamma(t_i), p_i).$$

Let $\exp_p$ denote the Riemannian exponential map at $p \in M$; see, e.g., [Boo03, dC92]. Since M is a closed Riemannian submanifold of $\mathbb{R}^n$, it follows that M is complete (see [Pal63, p. 326]), which means that $\exp_p \xi$ exists for all $\xi \in T_p M$. If $q \in M$ is not in the cut locus of p, then there exists a unique minimizing geodesic α_{pq} with $\alpha_{pq}(0) = p$ and $\alpha_{pq}(1) = q$ (see [dC92, corollary 13.2.8]), and we define $\exp_p^{-1}(q) = \dot{\alpha}_{pq}(0)$. Note that in this case, it also holds that p is not in the cut locus of q (see [dC92, corollary 13.2.7]), and we have $\exp_q^{-1}(p) = -\dot{\alpha}_{pq}(1)$. An expression for the derivative of f_i is readily obtained from the following result. A proof is given in Section A.

Theorem 3.1 (Karcher, 1977)**.** *Let M be a complete Riemannian manifold, let p be a point of M and let q be a point of M that is not in the cut locus of p. Then the squared distance function to p is differentiable at q and we have, for all $\xi \in T_q M$,*

$$\frac{1}{2}\mathrm{D}d^2(p, \cdot)(q)[\xi] = \left\langle \xi, -\exp_q^{-1} p \right\rangle.$$

In view of this result, we have that the derivative of f_i at γ along $w \in T_\gamma(\Gamma_1)$ is

$$\mathrm{D}f_i(\gamma)[w] = \langle w(t_i), v_i \rangle, \tag{11}$$

where

$$v_i = -\exp_{\gamma(t_i)}^{-1}(p_i),$$

provided that $\gamma(t_i)$ is not in the cut locus of p_i. This is a mild condition, since the cut locus has measure zero [GHL04, lemma 3.96]. Finally, the derivative of E_{d} is given by

$$\mathrm{D}E_{\mathrm{d}}(\gamma)[w] = \sum_{i=0}^{N} \langle w(t_i), v_i \rangle .$$

3.2 Gradient of E_{d}

The gradient of f_i at γ, with respect to the first-order Palais metric (7), is the unique element g_i of $T_\gamma\Gamma_1$ such that, for all $w \in T_\gamma\Gamma_1$,

$$\langle\langle g_i, w \rangle\rangle_1 = \mathrm{D}f_i(\gamma)[w]. \tag{12}$$

The next theorem gives an expression for g_i.

Theorem 3.2. *The gradient of the function $f_i : \Gamma_1 \to \mathbb{R} : \gamma \mapsto \frac{1}{2}d^2(\gamma(t_i), p_i)$ evaluated at $\gamma \in \Gamma_1$ is the vector field g_i along γ defined by*

$$g_i(t) = \begin{cases} (1+t)\tilde{v}_i(t), & 0 \le t \le t_i \\ (1+t_i)\tilde{v}_i(t), & t_i \le t \le 1 \end{cases},$$

where $v_i = -\exp^{-1}_{\gamma(t_i)}(p_i) \in T_{\gamma(t_i)}M$ and $\tilde{v}_i$ is the parallel transport of v_i along γ.

Proof. In view of (7), (11), and (12), g_i is the unique element of $T_\gamma\Gamma_1$ (vector field along γ) that satisfies

$$\langle g_i(0), w(0) \rangle + \int_0^1 \left\langle \frac{\mathrm{D}g_i}{\mathrm{d}t}(t), \frac{\mathrm{D}w}{\mathrm{d}t}(t) \right\rangle \mathrm{d}t = \langle v_i, w(t_i) \rangle \tag{13}$$

for all $w \in T_\gamma\Gamma_1$. Let $\tilde{v}_i$ denote the parallel transport of v_i along γ, and observe that $\tilde{v}_i(t_i) = v_i$. We have

$$\langle v_i, w(t_i) \rangle = \langle \tilde{v}_i(0), w(0) \rangle + \int_0^{t_i} \frac{\mathrm{d}}{\mathrm{d}t} \langle \tilde{v}_i(t), w(t) \rangle \, \mathrm{d}t$$

since $t \mapsto \langle \tilde{v}_i(t), w(t) \rangle$ is absolutely continuous,

$$= \langle \tilde{v}_i(0), w(0) \rangle + \int_0^{t_i} \left\langle \frac{\mathrm{D}\tilde{v}_i}{\mathrm{d}t}(t), w(t) \right\rangle + \left\langle \tilde{v}_i(t), \frac{\mathrm{D}w}{\mathrm{d}t}(t) \right\rangle \mathrm{d}t$$

by a defining property of the Levi-Civita covariant derivative,

$$= \langle \tilde{v}_i(0), w(0) \rangle + \int_0^{t_i} \left\langle \tilde{v}_i(t), \frac{\mathrm{D}w}{\mathrm{d}t}(t) \right\rangle \mathrm{d}t$$

since $\frac{\mathrm{D}}{\mathrm{d}t}\tilde{v}_i = 0$ by definition of parallel transport. It follows that g_i satisfies

$$\begin{aligned} g_i(0) &= \tilde{v}_i(0), \\ \frac{\mathrm{D}g_i}{\mathrm{d}t}(t) &= \begin{cases} \tilde{v}_i(t), & 0 < t < t_i, \\ 0, & t_i < t < 1. \end{cases} \end{aligned}$$

The solution is given by

$$g_i(t) = \begin{cases} (1+t)\tilde{v}_i(t), & 0 \le t \le t_i, \\ (1+t_i)\tilde{v}_i(t), & t_i \le t \le 1. \end{cases} \tag{14}$$

□

Observe that g_i is covariantly linear from 0 to t_i, and is covariantly constant from t_i to 1. In other words, the covariant derivative of g_i is covariantly constant ($\tilde{v}_i$) until t_i, and it is 0 after that. Note also that $\tilde{v}_i(t_i) = v_i$.

Once we have the gradient for each of the terms in E_{d}, the gradient of E_{d}, under the first-order Palais metric, is simply their sum

$$G_1 = \sum_{i=0}^{N} g_i. \tag{15}$$

3.3 Derivative of $E_{\mathrm{s},1}$

The derivative and gradient of $E_{\mathrm{s},1}$ (2) can be readily deduced, e.g., from [Tro77, §6] or [KS06, Th. 1]. We give a full development here for convenience.

Recall that

$$E_{\mathrm{s},1}(\gamma) = \frac{1}{2}\int_0^1 \langle \dot{\gamma}(t), \dot{\gamma}(t)\rangle \, \mathrm{d}t.$$

Define a *variation* of γ to be a smooth function $h : [0,1] \times (-\epsilon, \epsilon) \to M$: $(t,s) \mapsto h(t,s)$ such that $h(t,0) = \gamma(t)$ for all $t \in [0,1]$. The variational vector field corresponding to h is given by $w(t) = h_s(t,0)$, where $h_s(t,s)$ stands for $\frac{\partial h}{\partial s}(t,s)$. Thinking of h as a path of curves in M, we define $F(s)$ as the energy of the curve obtained by restricting h to $[0,1] \times \{s\}$. That is,

$$F(s) = \frac{1}{2}\int_0^1 \langle h_t(t,s), h_t(t,s)\rangle \, dt \ .$$

We now compute,

$$F'(0) = \int_0^1 \left\langle \frac{\mathrm{D}h_t}{\mathrm{d}s}(t,0), h_t(t,0) \right\rangle \mathrm{d}t = \int_0^1 \left\langle \frac{\mathrm{D}h_s}{\mathrm{d}t}(t,0), h_t(t,0) \right\rangle \mathrm{d}t = \int_0^1 \left\langle \frac{\mathrm{D}w}{\mathrm{d}t}(t), \dot{\gamma}(t) \right\rangle \mathrm{d}t,$$

since $h_t(t,0)$ is simply $\dot{\gamma}(t)$. Hence the derivative of $E_{\mathrm{s},1}$ at γ along w is given by

$$\mathrm{D}E_{\mathrm{s},1}(\gamma)[w] = \int_0^1 \left\langle \frac{\mathrm{D}w}{\mathrm{d}t}(t), \dot{\gamma}(t) \right\rangle \mathrm{d}t. \tag{16}$$

3.4 Gradient of $E_{\mathrm{s},1}$

In view of the above expression for the derivative of $E_{\mathrm{s},1}$, we derive in the next theorem an expression for the gradient of $E_{\mathrm{s},1}$.

Theorem 3.3. *The vector field H_1 along γ that provides the gradient of the function $E_{\mathrm{s},1}$ (2) on Γ_1 (9) with respect to the first-order Palais metric (7) satisfies the equation:*

$$\frac{\mathrm{D}H_1}{\mathrm{d}t}(t) = \dot{\gamma}(t), \quad H_1(0) = 0. \tag{17}$$

Proof. In view of (16), H_1 is the unique vector field along γ that satisfies

$$\langle w(0), H_1(0)\rangle + \int_0^1 \left\langle \frac{\mathrm{D}w}{\mathrm{d}t}(t), \frac{\mathrm{D}H_1}{\mathrm{d}t}(t) \right\rangle dt = \int_0^1 \left\langle \frac{\mathrm{D}w}{\mathrm{d}t}(t), \dot{\gamma}(t) \right\rangle \mathrm{d}t,$$

for all $w \in T_\gamma\Gamma_1$. The result follows. □

In the case $M = \mathbb{R}^n$, the gradient vector field is simply $H_1(t) = \gamma(t) - \gamma(0)$.

3.5 Gradient of E_1

Since $E_1 = E_{\mathrm{d}} + \lambda E_{\mathrm{s},1}$, the gradient of E_1 follows directly from the gradients of E_{d} and $E_{\mathrm{s},1}$ computed above. We thus have that $\nabla E_1 = G_1 + \lambda H_1$, with G_1 given by (15) and H_1 given by (17).

4 Gradient of E_2 in the second-order Palais metric

Recall that $E_2 = E_\mathrm{d} + \lambda E_{\mathrm{s},2}$ is defined on Γ_2 (10) by

$$E_2(\gamma) = \frac{1}{2}\sum_{i=0}^{N} d^2(\gamma(t_i), p_i) + \frac{\lambda}{2}\int_0^1 \left\langle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2} \right\rangle \mathrm{d}t. \tag{18}$$

The purpose of this section is to obtain an expression for the gradient of E_2 with respect to the second-order Palais metric (5).

4.1 Gradient of E_d

The derivative does not depend on the metric, in contrast to the gradient. Thus we have, as in Section 3,

$$\mathrm{D}f_i(\gamma)[w] = \langle w(t_i), v_i \rangle ,$$

where f_i denotes the function $\gamma \mapsto \frac{1}{2}d^2(\gamma(t_i), p_i)$ and $v_i = -\exp^{-1}_{\gamma(t_i)}(p_i)$.

Theorem 4.1. *The gradient of the function $f_i : \Gamma_2 \to \mathbb{R} : \gamma \to d^2(p_i, \gamma(t_i))$ at $\gamma \in \Gamma_2$ with respect to the second-order Palais metric (5) is given by the vector field g_i along γ defined by*

$$g_i(t) = \begin{cases} (1 + t_i t + \frac{1}{2}t_i t^2 - \frac{1}{6}t^3)\tilde{v}_i(t) & 0 \le t \le t_i \\ (1 + t t_i + \frac{1}{2}t t_i^2 - \frac{1}{6}t_i^3)\tilde{v}_i(t) & t_i \le t \le 1, \end{cases}$$

where $\tilde{v}_i$ is the parallel transport of v_i along γ.

Proof. In view of (5) and (11), g_i is the unique tangent vector field along γ that satisfies

$$\langle g_i(0), w(0)\rangle + \left\langle \frac{\mathrm{D}g_i}{\mathrm{d}t}(0), \frac{\mathrm{D}w}{\mathrm{d}t}(0) \right\rangle + \int_0^{t_i} \left\langle \frac{\mathrm{D}^2 g}{\mathrm{d}t^2}(t), \frac{\mathrm{D}^2 w}{\mathrm{d}t^2}(t) \right\rangle \mathrm{d}t = \langle v_i, w(t_i)\rangle .$$

We have

$$\langle v_i, w(t_i)\rangle = \langle \tilde{v}_i(0), w(0)\rangle + t_i \frac{\mathrm{d}}{\mathrm{d}t}\langle \tilde{v}_i(t), w(t)\rangle \,|_{t=0} + \int_0^{t_i} (t_i - t)\frac{\mathrm{d}^2}{\mathrm{d}t^2}\langle \tilde{v}_i(t), w(t)\rangle \,\mathrm{d}t$$

by Taylor's theorem with remainder in integral form,

$$= \langle \tilde{v}_i(0), w(0)\rangle + t_i \left\langle \tilde{v}_i(0), \frac{\mathrm{D}w}{\mathrm{d}t}(0) \right\rangle + \int_0^{t_i} (t_i - t)\left\langle \tilde{v}_i(t), \frac{\mathrm{D}^2 w}{\mathrm{d}t^2}(t) \right\rangle \mathrm{d}t$$

for reasons already given in the proof of Theorem 3.2. It follows that

$$\begin{aligned} g_i(0) &= \tilde{v}_i(0), \\ \frac{\mathrm{D}g_i}{\mathrm{d}t}(0) &= t_i \tilde{v}_i(0), \\ \frac{\mathrm{D}^2 g_i}{\mathrm{d}t^2}(t) &= \begin{cases} (t_i - t)\tilde{v}_i(t), & 0 < t < t_i, \\ 0, & t_i < t < 1. \end{cases} \end{aligned}$$

Solving for g_i, we obtain:

$$g_i(t) = \begin{cases} (1 + t_i t + \frac{1}{2}t_i t^2 - \frac{1}{6}t^3)\tilde{v}_i(t), & 0 \le t \le t_i, \\ (1 + t t_i + \frac{1}{2}t t_i^2 - \frac{1}{6}t_i^3)\tilde{v}_i(t), & t_i \le t \le 1. \end{cases} \tag{19}$$

□

This gradient function is a covariantly cubic polynomial before t_i and is a covariantly linear polynomial after t_i. The total gradient is given by $G_2(t) = \sum_{i=0}^{N} g_i(t)$. Another way of writing this summation is, for $t_{i-1} \le t \le t_i$,

$$G_2(t) = \sum_{j=0}^{i-1}(1 + t t_j + \frac{1}{2}t t_j^2 - \frac{1}{6}t_j^3)\tilde{v}_j(t) + \sum_{j=i}^{N}(1 + t_j t + \frac{1}{2}t_j t^2 - \frac{1}{6}t^3)\tilde{v}_j(t). \tag{20}$$

4.2 Derivative of $E_{\mathrm{s},2}$

Let $h(t,s)$ be a collection, indexed by s, of elements of the curve set Γ_2. Hence, for a fixed s, we have a curve on M parameterized by t. For $s=0$, that curve is called γ. Then w defined by $w(t) = h_s(t,0)$ is a generic tangent vector to Γ_2 at γ. We have $\mathrm{D}E_{\mathrm{s},2}(\gamma)[w] = \frac{\mathrm{d}}{\mathrm{d}s}F(s)\big|_{s=0}$, where

$$F(s) = \frac{1}{2}\int_0^1 \left\langle \frac{\mathrm{D}}{\mathrm{d}t}(h_t(t,s)), \frac{\mathrm{D}}{\mathrm{d}t}(h_t(t,s)) \right\rangle \mathrm{d}t.$$

Taking the derivative with respect to s:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}F(s) &= \int_0^1 \left\langle \frac{\mathrm{D}}{\mathrm{d}s}(\frac{\mathrm{D}}{\mathrm{d}t}(h_t(t,s))), \frac{\mathrm{D}}{\mathrm{d}t}(h_t(t,s)) \right\rangle \mathrm{d}t \\ &= \int_0^1 \left\langle [R(h_s(t,s),h_t(t,s))(h_t(t,s)) + \frac{\mathrm{D}}{\mathrm{d}t}(\frac{\mathrm{D}}{\mathrm{d}s}(h_t(t,s)))], \frac{\mathrm{D}}{\mathrm{d}t}(h_t(t,s)) \right\rangle \mathrm{d}t,\end{aligned}$$

where R is the Riemannian curvature tensor defined as:

$$R(h_s,h_t)(v) = \frac{\mathrm{D}}{\mathrm{d}s}\frac{\mathrm{D}}{\mathrm{d}t}(v) - \frac{\mathrm{D}}{\mathrm{d}t}\frac{\mathrm{D}}{\mathrm{d}s}(v).$$

(Note that the curvature tensor is sometimes defined with the opposite sign in the literature.) Since $\frac{\mathrm{D}}{\mathrm{d}s}(h_t) = \frac{\mathrm{D}}{\mathrm{d}t}(h_s)$, the desired derivative at $s=0$ becomes:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}F(s)|_{s=0} &= \int_0^1 \left\langle [R(w,\dot\gamma)(\dot\gamma) + \frac{\mathrm{D}^2}{\mathrm{d}t^2}(w)], \frac{\mathrm{D}}{\mathrm{d}t}(\dot\gamma) \right\rangle \mathrm{d}t \\ &= \int_0^1 \left\langle R(w,\dot\gamma)(\dot\gamma), \frac{\mathrm{D}}{\mathrm{d}t}(\dot\gamma) \right\rangle \mathrm{d}t + \int_0^1 \left\langle \frac{\mathrm{D}^2}{\mathrm{d}t^2}(w), \frac{\mathrm{D}}{\mathrm{d}t}(\dot\gamma) \right\rangle \mathrm{d}t. \qquad (21)\end{aligned}$$

This is the sought expression for $\mathrm{D}E_{\mathrm{s},2}(\gamma)[w]$.

4.3 Gradient of $E_{\mathrm{s},2}$

We will analyze the two terms in (21) separately.

The Riemannian curvature tensor has certain symmetries: for vector fields a,b,c,d along γ,

$$\langle R(a,b)(c),d\rangle = -\langle R(b,a)(c),d\rangle = -\langle R(a,b)(d),c\rangle = \langle R(c,d)(a),b\rangle,$$

which allows us to rewrite the first term of (21) as

$$\int_0^1 \left\langle R(\frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}(t),\dot\gamma(t))(\dot\gamma(t)), w(t) \right\rangle \mathrm{d}t = \int_0^1 \langle A(t), w(t)\rangle \,\mathrm{d}t,$$

where A denotes the vector field along the curve γ given by $A(t) = R(\frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}(t),\dot\gamma(t))(\dot\gamma(t))$. We need a vector field H_2 along γ with the property that

$$\langle H_2(0), w(0)\rangle + \langle H_2'(0), w'(0)\rangle + \int_0^1 \langle H_2''(t), w''(t)\rangle \,\mathrm{d}t = \int_0^1 \langle A(t), w(t)\rangle \,\mathrm{d}t$$

for all $w \in T_\gamma\Gamma_2$, where we temporarily use the prime as simplified notation for $\frac{\mathrm{D}}{\mathrm{d}t}$. We now express the right-hand side of the latter equation in a form that matches the one of the left-hand side. To this end, we let $A^{[k]}$ denote the kth covariant integral of A with zero conditions at $t=0$. Moreover, in keeping with previous notation, we let $\widetilde{A^{[k]}(1)}$ denote the parallel transport of $A^{[k]}(1)$ along γ. We have

$$\int_0^1 \langle A(t), w(t)\rangle \,\mathrm{d}t = \left\langle A^{[1]}(t) - \widetilde{A^{[1]}(1)}(t), w(t) \right\rangle\Big]_0^1 - \int_0^1 \left\langle A^{[1]}(t) - \widetilde{A^{[1]}(1)}(t), w'(t) \right\rangle \mathrm{d}t$$

by integration by parts, where $A^{[1]} - \widetilde{A^{[1]}(1)}$ is the primitive of A that vanishes at $t = 1$,

$$= \left\langle \widetilde{A^{[1]}(1)}(0), w(0) \right\rangle - \int_0^1 \left\langle A^{[1]}(t) - \widetilde{A^{[1]}(1)}(t), w'(t) \right\rangle \mathrm{d}t$$

since $A^{[1]}(0) = 0$ by definition,

$$\begin{aligned} &= \left\langle \widetilde{A^{[1]}(1)}(0), w(0) \right\rangle - \left\langle A^{[2]}(t) - t\widetilde{A^{[1]}(1)}(t) - \widetilde{A^{[2]}(1)} + \widetilde{A^{[1]}(1)}, w'(t) \right\rangle \Big]_0^1 \\ &\quad + \int_0^1 \left\langle A^{[2]}(t) - t\widetilde{A^{[1]}(1)}(t) - \widetilde{A^{[2]}(1)}(t) + \widetilde{A^{[1]}(1)}(t), w''(t) \right\rangle \mathrm{d}t \end{aligned}$$

by integration by parts,

$$\begin{aligned} &= \left\langle \widetilde{A^{[1]}(1)}(0), w(0) \right\rangle + \left\langle -\widetilde{A^{[2]}(1)}(0) + \widetilde{A^{[1]}(1)}(0), w'(0) \right\rangle \\ &\quad + \int_0^1 \left\langle A^{[2]}(t) - t\widetilde{A^{[1]}(1)}(t) - \widetilde{A^{[2]}(1)}(t) + \widetilde{A^{[1]}(1)}(t), w''(t) \right\rangle \mathrm{d}t, \end{aligned}$$

which is the sought form. We can now conclude that

$$\begin{aligned} H_2(0) &= \widetilde{A^{[1]}(1)}(0), \\ H_2'(0) &= -\widetilde{A^{[2]}(1)}(0) + \widetilde{A^{[1]}(1)}(0), \\ H_2''(t) &= A^{[2]}(t) - t\widetilde{A^{[1]}(1)}(t) - \widetilde{A^{[2]}(1)}(t) + \widetilde{A^{[1]}(1)}(t),\ 0 \le t \le 1. \end{aligned}$$

The solution is given by

$$H_2(t) = \hat{H}_2(t) - \frac{1}{6}t^3\tilde{S}(t) - \frac{1}{2}t^2(\tilde{Q}(t) - \tilde{S}(t)) - t(\tilde{Q}(t) - \tilde{S}(t)) + \tilde{S}(t), \tag{22}$$

where $\hat{H}_2 = A^{[4]}$, $S = A^{[1]}(1) = \hat{H}_2'''(1)$ and $Q = A^{[2]}(1) = \hat{H}_2''(1)$.

We now consider the second term in (21), that is, $\int_0^1 \left\langle \frac{\mathrm{D}^2 w}{\mathrm{d}t^2}(t), \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}(t) \right\rangle \mathrm{d}t$. This term can be written as $\langle\langle H_3, w\rangle\rangle_2$ where H_3 satisfies

$$\frac{\mathrm{D}^2 H_3}{\mathrm{d}t^2} = \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \quad H_3(0) = \frac{\mathrm{D}H_3}{\mathrm{d}t}(0) = 0, \tag{23}$$

that is, H_3 is two times covariant integral of $\frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}$ with initial conditions $H_3(0) = \frac{\mathrm{D}H_3}{\mathrm{d}t}(0) = 0$.

In summary, we have the following result.

Theorem 4.2. *The vector field along γ that provides the gradient of the function $E_{\mathrm{s},2}$ (3) on Γ_2 (10) with respect to the second-order Palais metric (5) is given by*

$$H_2(t) + H_3(t), \tag{24}$$

as follows: H_2 admits the expression (22); $\hat{H}_2$ is given by $\frac{\mathrm{D}^4\hat{H}_2}{\mathrm{d}t^4}(t) = R(\frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}(t), \dot{\gamma}(t))(\dot{\gamma}(t))$ with initial conditions $\hat{H}_2(0) = \frac{\mathrm{D}\hat{H}_2}{\mathrm{d}t}(0) = \frac{\mathrm{D}^2\hat{H}_2}{\mathrm{d}t^2}(0) = \frac{\mathrm{D}^3\hat{H}_2}{\mathrm{d}t^3}(0) = 0$; $\tilde{Q}$ and $\tilde{S}$ are the parallel transport along γ of $Q = \frac{\mathrm{D}^2\hat{H}_2}{\mathrm{d}t^2}(1)$ and of $S = \frac{\mathrm{D}^3\hat{H}_2}{\mathrm{d}t^3}(1)$; H_3 is given by $\frac{\mathrm{D}^2 H_3}{\mathrm{d}t^2} = \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}$ with initial conditions $H_3(0) = \frac{\mathrm{D}H_3}{\mathrm{d}t}(0) = 0$.

In case $M = \mathbb{R}^n$, the two terms are simply $H_2(t) = 0$ and $H_3(t) = \gamma(t) - \dot{\gamma}(0)t - \gamma(0)$ for all $t \in I$.

4.4 Gradient of E_2

Combining the two gradient terms, we get the gradient of E_2 under the second-order Palais metric:

$$\nabla E_2 = G_2 + \lambda(H_2 + H_3),$$

where G_2 is given in (20), H_2 in (22), and H_3 in (23).

5 Steepest-descent algorithm on the curve spaces

Let E stand for E_1 (6), resp. E_2 (4), Γ for the set Γ_1 (9), resp. Γ_2 (10), of curves on the Riemannian manifold M, and let Γ be endowed with the first-order Palais metric (7), resp. second-order Palais metric (5). We propose the steepest-descent method for E described in Algorithm 1.

The algorithm creates a sequence of curves $(\gamma_k)_{k=0,1,\ldots} \subset \Gamma$ with decreasing energy $E(\gamma_k)$. The initialization step consists in choosing an arbitrary curve in Γ to be the starting curve γ_0. Then, given the current iterate γ_k, the algorithm computes the gradient $\nabla E(\gamma_k)$ and updates the curve to γ_{k+1} according to

$$\gamma_{k+1}(t) = \exp_{\gamma_k(t)}(-\hat{\rho}_k \nabla E(\gamma_k)(t)), \quad t \in I,$$

where $\hat{\rho}_k$ is a step size chosen using some step size selection rule (see, e.g., [Ber95]). For simplicity of exposition, Algorithm 1 is stated with the simple Armijo backtracking rule; in the numerical experiments presented in Section 6, we have used a more sophisticated strategy to satisfy the strong Wolfe conditions. In practice, the algorithm is stopped when a certain pre-determined stopping criterion is satisfied. The criterion can be a threshold on the norm of $\nabla E(\gamma_k)$, for example.

Algorithm 1 Gradient descent

1: Given an initial curve $\gamma_0 \in \Gamma$ and scalars $\bar{a} > 0$, $b, \sigma \in (0, 1)$;
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Compute $E(\gamma_k)$ and $\nabla E(\gamma_k)$;
4: Find the smallest integer $m > 0$ such that

$$E(\exp_{\gamma_k}(-\bar{a}b^m \nabla E(\gamma_k)) \leq E(\gamma_k) + \sigma \left\langle\!\left\langle \nabla E(\gamma_k), -\bar{a}b^m \nabla E(\gamma_k) \right\rangle\!\right\rangle_{\gamma_k}$$

and set $\hat{\rho}_k = \bar{a}b^m$;
5: Set $\gamma_{k+1}(t) = \exp_{\gamma_k(t)}(-\hat{\rho}_k \nabla E(\gamma_k)(t))$;
6: **end for**

Whereas analyzing the convergence of steepest-descent type methods on *finite-dimensional* manifolds is relatively simple (see [AG09]), the convergence analysis of steepest-descent methods on *infinite-dimensional* spaces is no trivial matter; see [SZ04] and references therein. Analyzing the convergence of Algorithm 1 is the object of ongoing research. Nevertheless, it is reasonable to expect that the algorithm behaves like steepest-descent methods in finite dimension: the sequence of iterates γ_k has a single limit (see [AMA05]) which, unless the initial curve is maliciously chosen, is a local minimizer of the objective function E. These expectations are corroborated by our numerical experiments; see Section 6.

6 Illustration on some specific manifolds: $M = \mathbb{R}^2$, $\mathbb{S}^2$

In this section we present some illustrations of our gradient descent approach to finding optimal curves. In the case of Euclidean spaces, it is sometimes possible to derive expressions for the optimal curves under E_1 and E_2 directly. In those situations, we can compare our numerical solutions to the analytical expressions, and characterize the performances. In the remaining cases, where the analytical solutions are not readily available, we will simply illustrate the results obtained using our procedures. Examples involving the analytical expressions will have $M = \mathbb{R}^n$ and while the other cases will have $M = \mathbb{S}^2$.

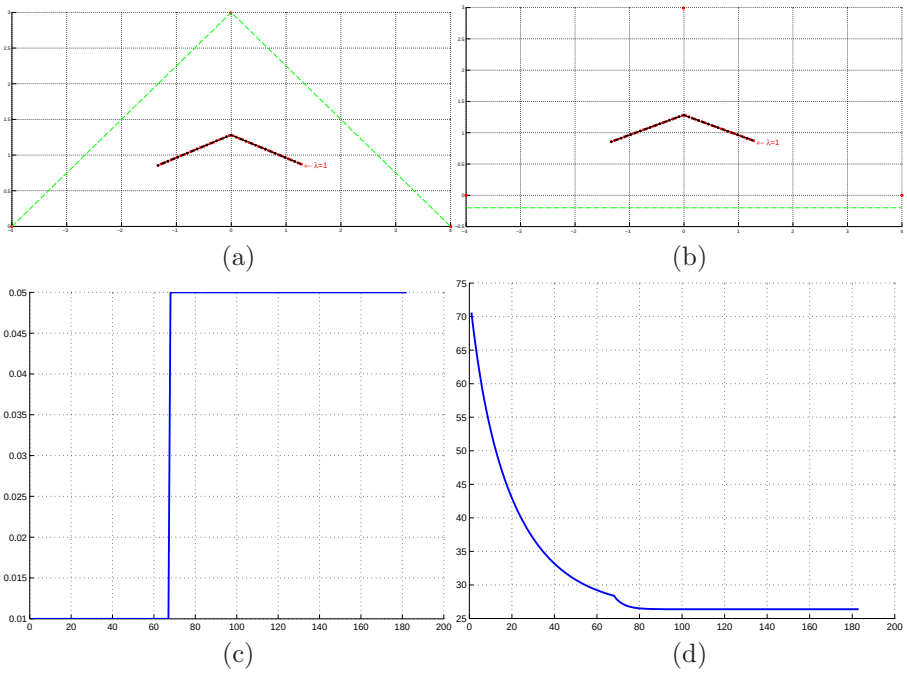


Figure 1: (a) and (b): The minimum of E_1 in $M = \mathbb{R}^2$ reached by the gradient descent method with respect to Palais metric using different starting curves for $\lambda = 1$, (c): the step length variation, and (d): the energy evolution versus iterations for the example shown in (a).

6.1 Analytical solution of E_1 in $\mathbb{R}^2$

As the first example we will consider the problem of finding the optimal curves under E_1 when $M = \mathbb{R}^2$. For simplicity, we will take $\lambda = 1$ in (6). This case is simple enough to seek an analytical expression as follows. Let $N = 2$ and let the three data points be given by $p_0 = (-A, 0)$, $p_1 = (0, B)$, $p_2 = (A, 0)$, at the time instants $t_0 = 0$, $t_1 = 0.5$, $t_2 = 1$, where A and B are two real variables. Using the symmetry of the given points, and since we know that the optimal curves of E_1 are piecewise geodesics, we can restrict our search to piecewise-geodesic curves γ connecting $q_0 = (-a, c)$, $q_1 = (0, b)$, and $q_2 = (a, c)$ at the three given time instants. Our goal is to find the values of a, b, and c in $\mathbb{R}$ such that γ is a minimum of E_1. By computing E_{d} and $E_{\mathrm{s},1}$ manually we get $E_1 = (A-a)^2 + c^2 + \frac{1}{2}(B-b)^2 + 2(a^2 + (b-c)^2)$. The critical points are given by the equation $\nabla E_1 = 0$, i.e., in terms of partial derivatives we have $\frac{\partial E_1}{\partial a} = \frac{\partial E_1}{\partial b} = \frac{\partial E_1}{\partial c} = 0$. This system has only one solution given by $a = \frac{1}{3}A$, $b = \frac{3}{7}B$, and $c = \frac{2}{7}B$, and the minimum of E_1 is given by the piecewise geodesic curve connecting points $q_0 = (-A/3, 2B/7)$, $q_1 = (0, 3B/7)$, and $q_2 = (A/3, 2B/7)$.

Shown in Figure 1((a) and (b)) are two optimal curves under E_1 obtained by our algorithm, for two different initial conditions. In each case, the green (dashed) curve shows the initial condition and the black (dotted) curve shows the final result obtained numerically. The red (continuous) curves show the optimal curve obtained using the analytical solution. The coincidence of black and red curves shows the accuracy and the stability of our algorithm. In Figure 1((c) and (d)) we show the variation of the step length, and the variation of the cost function E_1, respectively versus iterations, corresponding to the example shown in Figure 1(a).

In Figure 2 we present some additional results for $\mathbb{R}^2$, this time restricting only to our numerical solutions. These examples use a random set of points and different values of λ to demonstrate the strength of the algorithm. Each of these solutions are piecewise geodesics and the end points of the geodesic segments depend on the value of λ.

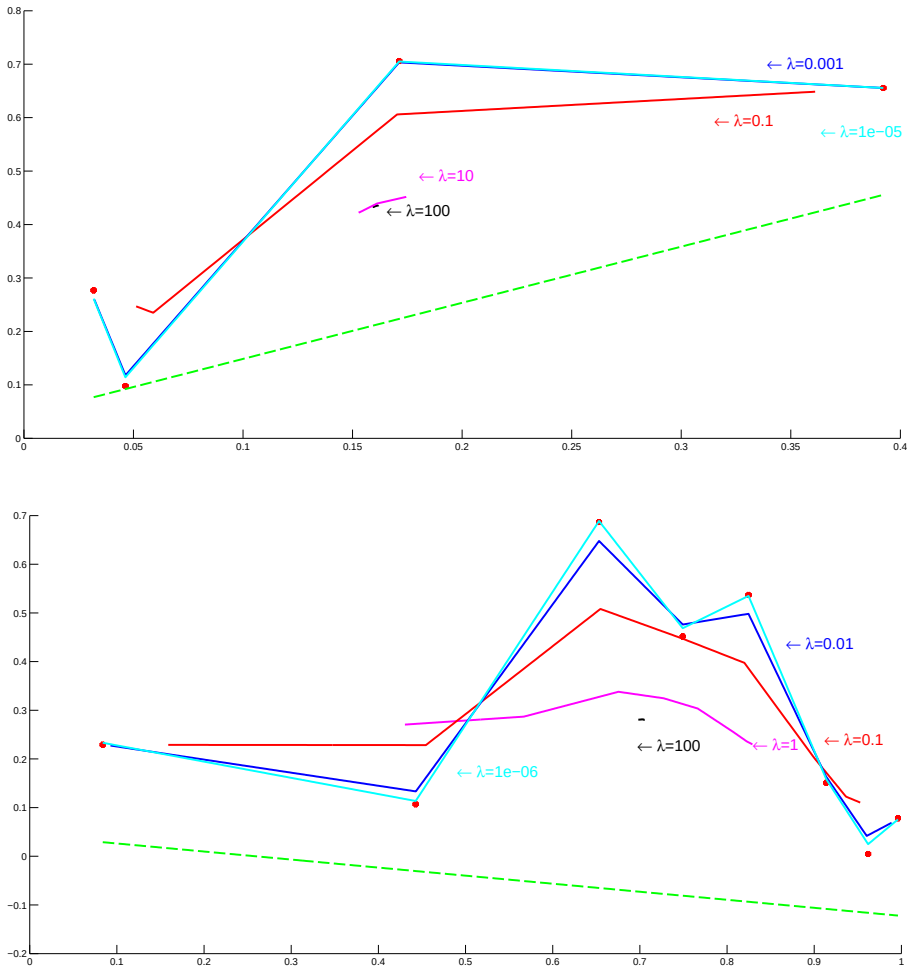


Figure 2: The minimum of E_1 in $M = \mathbb{R}^2$ reached by the gradient descent method with respect to first-order Palais metric using different values of λ.

6.2 Analytical solution of E_2 in $\mathbb{R}^n$

Next we derive the optimal curves under E_2 for Euclidean spaces. It is interesting to note that in this case the cost function has n components, each corresponding to a coordinate in $\mathbb{R}^n$. In other words, the problem breaks down into n independent problems, each being one-dimensional. Therefore, it is sufficient to illustrate the analytical solution for the one-dimensional case.

To derive an analytical solution to the one-dimensional problem, we will first establish a number of relations that this curve must satisfy and then use those relations to solve for the unknowns. We start with the fact that $\ddot{\gamma}(t) = \ddot{H}_3(t)$ for all t. Therefore, γ takes the form:

$$\gamma(t) = H_3(t) + rt + s \ ,$$

where r and s are two constants. Next, since γ is a critical point of E_2, we have $G_2(t) = -H_3(t)$ (assuming $\lambda = 1$) for all t which makes $\gamma(t) = -G_2(t)+rt+s$. Enumerating the different conditions on γ, we obtain the following constraints.

1. Since $H_3(0) = 0$, we have $G_2(0) = 0$ which implies:

$$G_2(0) = \sum_{j=1}^{n} v_j = 0 \ ,$$

where v_j is as defined in Section 4.1.

2. Also, since $\dot{H}_3(0) = 0$, we have $\dot{G}_2(0) = 0$ which means:

$$\dot{G}_2(0) = \sum_{j=1}^{n} v_j t_j = 0 \ .$$

3. Finally, since we know that $\gamma(t_i) = v_i + p_i$, we get:

$$-G_2(t_i) + rt_i + s = v_i + p_i \ ,$$

which give us the relations: for $i = 0, \ldots, N$

$$-\sum_{j=0}^{N}(1 + t_j t_i + \frac{1}{2}t_j t_i^2 - \frac{1}{6}t_i^3)v_j - \sum_{j=1}^{i-1}(1 + t_i t_j + \frac{1}{2}t_i t_j^2 - \frac{1}{6}t_j^3)v_j + rt_i + s = v_i + p_i \ . \quad (25)$$

Rearranging this equation, we reach

$$\sum_{j=1}^{n} \beta_{j,i} v_j + rt_i + s = p_i \ ,$$

where

$$\beta_{j,i} = \begin{cases} -[(1 - \frac{1}{6}t_j^3) + (t_j + \frac{1}{2}t_j^2)t_i] & j < i \\ -[(2 + t_j t_i + \frac{1}{2}t_j t_i^2 - \frac{1}{6}t_i^3)] & j = i \\ -[(1 + t_j t_i + \frac{1}{2}t_j t_i^2 - \frac{1}{6}t_i^3)] & j > i \end{cases}$$

Taking these three types of relations, we form a linear system of equations. We have $N + 3$ equations and $N + 3$ unknowns:

$$\begin{bmatrix} 1 & 1 & \ldots & 1 & 0 & 0 \\ t_0 & t_1 & \ldots & t_N & 0 & 0 \\ \beta_{0,0} & \beta_{1,0} & \ldots & \beta_{N,0} & t_0 & 1 \\ \ldots & & & & & \\ \beta_{0,N} & \beta_{1,N} & \ldots & \beta_{N,N} & t_N & 1 \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \ldots \\ v_N \\ r \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ p_0 \\ p_1 \\ \ldots \\ p_N \end{bmatrix} \ . \quad (26)$$

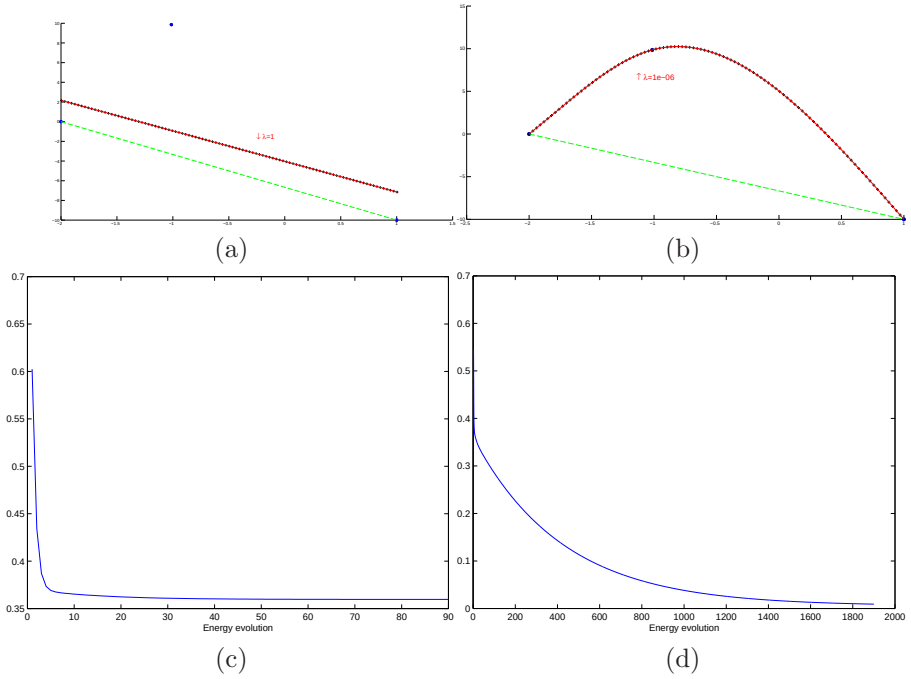


Figure 3: The panels (a) and (b) show two examples of optimal curves under E_2 obtained using our numerical approach (black curve) and the analytical solution (red curve). The panels in (c) and (d) plot the evolutions of E_2 versus iterations for the cases (a) and (b), respectively. The curves in green (dashed) are used as the initial curves for the optimization.

After solving for the v_js, r and s, we can evaluate the optimal curve $\gamma(t) = -G_2(t) + rt + s$.

We present some examples for comparing the numerical solutions with this analytical solution for $n = 2$. In the panels Figure 3 (a) and (b), we present two examples with three points each and solve for the optimal curves under different λs. In each case, the green (dashed) line shows the initial curve, the black line shows the optimal curve obtained numerically, and the red (continuous) line shows the analytical solution. We have used a dotted pattern for the black (dotted) curve since the two optimal curves match perfectly and one hides the other. As predicted by the theory, the optimal solution resembles a straight line when λ is sufficiently large, and an interpolating spline when λ is sufficiently small. The plots in panels (c) and (d) show the corresponding evolution of E_2 versus iterations.

In Figure 4, we present some additional examples of optimal curves (obtained using our numerical method) under E_2 in $\mathbb{R}^2$ for a variety of data points and λs. Each panel in this figure shows the optimal γs for different (mostly small) values of λ but with the same data points. In each case the initial curve for the gradient process is given by the green (dashed) curve.

6.3 Optimal Curves on the Unit Sphere

In this section we consider the case of $M = \mathbb{S}^2$ where the analytical expressions for the optimal curves are not readily available, and we apply our numerical approach to find the solutions. In these experiments, we first generate $N+1$ data points $p_0, p_1, \ldots, p_N$ randomly on $\mathbb{S}^2$ and associate them with different instants of time $0 = t_0 < t_1 < t_2 < \ldots < t_n = 1$. Then, we initialize our algorithm by an arbitrary continuous curve $\gamma_0 \in \Gamma$, and finally apply our gradient descent method to search for the optimal curve γ_* that minimizes E.

1. **Case 1**: In the case $E = E_1$ we apply our algorithm as described in Section 3 and examples are shown in Figure 5. Similar to the Euclidean case, the solutions are piecewise geodesic

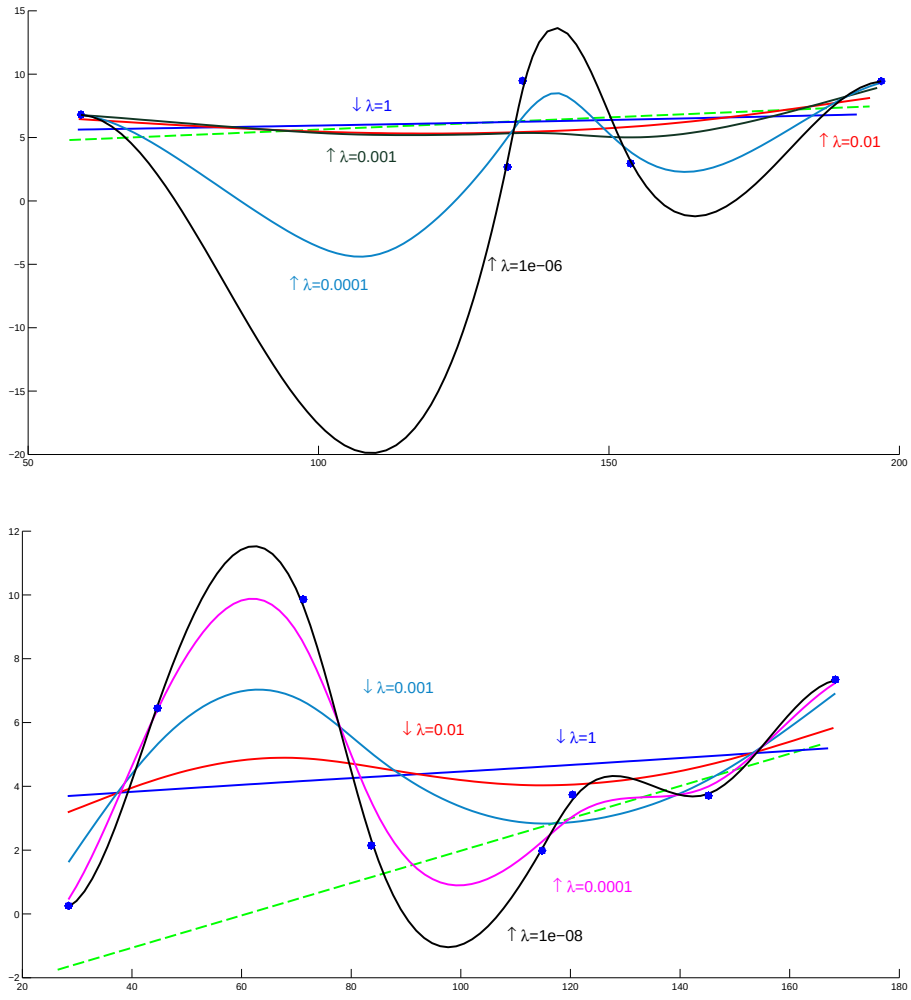
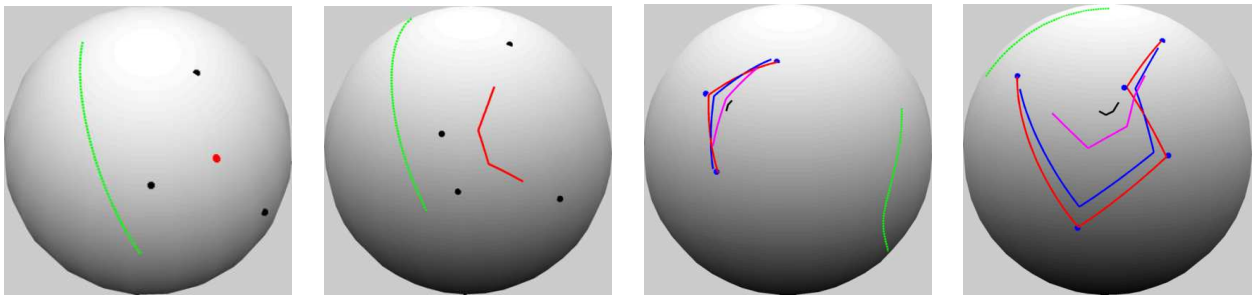
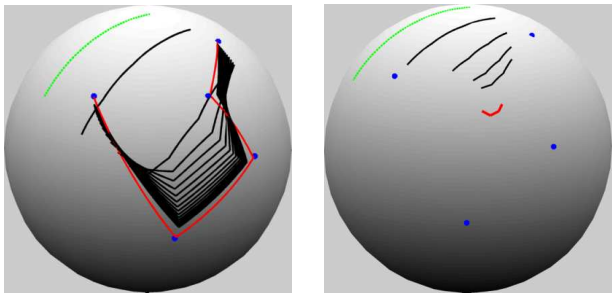


Figure 4: The optimal curves under E_2 for different combinations of data points and λs.



(a) (b) (c) (d)

Figure 5: Optimal curves under E_1 for $M = \mathbb{S}^2$ obtained by our gradient descent method with respect to the first-order Palais metric. (a): $\lambda = 100$, (b): $\lambda = 1$, (c) and (d): $\lambda = 10,\ 1,\ 0.1$, and 0.0001. In each case the green curve shows the initial condition.



(a) (b)

Figure 6: The evolution of curves under the gradient iterations for minimizing E_1. (a): $\lambda = 10^{-5}$, (b): $\lambda = 100$.

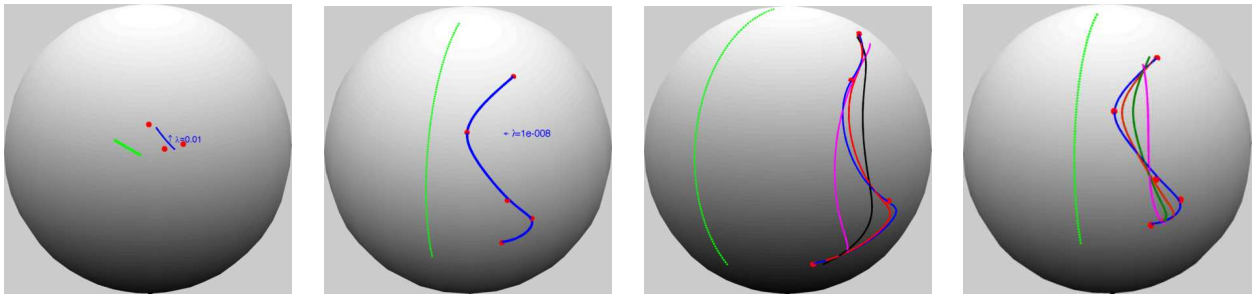
curves. Since geodesics on $\mathbb{S}^2$ are arcs that lie on great circles, these optimal curves are piecewise arcs. The panels (a) and (b) show examples of optimal γ for $N = 2$ (three data points) and $N = 3$ (four data points) with λ values being 100 and 1, respectively. For $\lambda = 100$, the resulting optimal curve looks like a point. The remaining two panels (c) and (d) show several optimal curves, each corresponding to different λs, for the same set of data points. As earlier, the initial condition for the gradient descent is given by the green curve.

The Figure 6 shows two examples of the actual optimization process where the iterative updates for γ under the gradient of E_1 are shown. The process starts with the green curves as the initial conditions and the updates are shown in black. The final curves in each case are shown in red.

2. **Case 2**: In the case $E = E_2$, we need to obtain an expression for the tangent vector field A defined in Section 4.3, which involves the Riemannian curvature tensor R on M. Since the unit sphere has constant sectional curvature $k = 1$, the curvature tensor is given by $R(X, Y)Z = k(\langle Y, Z\rangle X - \langle X, Z\rangle Y)$, $k = 1$; see, e.g., [Lee07]. Alternatively, this formula is easily obtained from the material in Section B. Thus, for the vector field A defined in Section 4.3, we have the expression

$$A = \langle \dot\gamma, \dot\gamma\rangle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2} - \left\langle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \dot\gamma \right\rangle \dot\gamma. \tag{27}$$

Using this expression, we first integrate $A(t)$ covariantly to determine the term H_2 of the gradient of E_2, and then use the gradient descent method of Algorithm 1 to minimize E_2.



(a) (b) (c) (d)

Figure 7: Optimal curves under E_2 for $M = \mathbb{S}^2$ obtained by our gradient descent method with respect to the second-order Palais metric. (a): $\lambda = 0.01$, (b): $\lambda = 10^{-8}$, (c) and (d): $\lambda = 0.01,\ 0.001,\ 10^{-4}$, and 10^{-6}. In each case the green curve shows the initial condition.

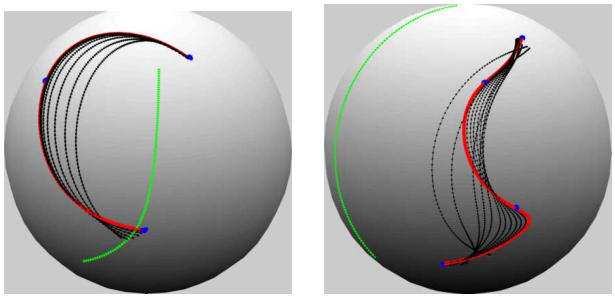


Figure 8: Evolutions of γ under the gradient of E_2. The green curves are the initial conditions and the red curves are the final states.

Shown in Figure 7 are some examples of our approach applied to different sets of points generated randomly on $\mathbb{S}^2$. The pictures in (a) and (b) show examples of optimal curves for three and five points with λs as indicated there. The remaining two panels show examples of optimal curves obtained for fixed data points under different λs. Curves in different colors are obtained by using different values of λ. The values of λ used in (c) and (d) are $10^{-2},\ 10^{-3},\ 10^{-4}$, and 10^{-6}. As the value of λ increases, we can see the optimal curves straightening and shortening into single arcs.

Figure 8 shows two examples of the iterative process by displaying the intermediate curves also. The initial curves are shown in green, the iterations are shown in black and the final curves are shown in red.

Asymptotics on λ: Our numerical experiments corroborate the following theoretical results mentioned in Section 1.1:

- Let λ go to zero. When $E = E_1$ (6), the optimal curve is the piecewise geodesic passing through the given points. When $E = E_2$ (4), the optimal curve is a piecewise cubic polynomial (in the sense of [NHP89]) interpolating the given set of points when $E = E_2$.
- Let λ go to infinity. When $E = E_1$, the optimal curve shrinks to one point in M, precisely the Riemannian center of mass (also called Fréchet mean or Karcher mean) of the given set of points $p_0, p_1, \ldots, p_N$. When $E = E_2$, the optimal curve approaches a geodesic.

7 Concluding remarks

We have addressed the problem of fitting a curve to data points on a closed Riemannian submanifold M of some Euclidean space by means of a Palais-based steepest-descent algorithm applied to the weighted sum of a fitting-related and a regularity-related cost function. As a proof of concept, we have used the simple regularity cost function (2) based on the first derivative. We have also considered the more challenging case of the regularity cost function (3), whose derivative involves the Riemannian curvature tensor on M, and for which the optimal curves are generalized cubic splines. We have illustrated the proposed method on fitting problems in $\mathbb{R}^2$ and $\mathbb{S}^2$. In future work, we will consider other nonlinear manifolds with applications in pattern recognition and image analysis.

An important feature of our approach is that the discretization takes place as late as possible in the implementation. The gradient of the cost function at a curve γ is a (continuous-time) vector field along γ expressed by means of the Riemannian logarithm, parallel transport, covariant differentiation, and covariant integrals. It is these operations that are approximated by discretization in the algorithm implementation. The advantage of using a continuous formulation is that tools from functional analysis become available. We are able to use the Palais metrics and, thus, simplify the gradient vector fields only because of this continuous formulation. An alternate approach would be to consider a discretization $\hat{\gamma}$ of γ using Q points and discretize the function E accordingly to obtain a new objective function $\hat{E} : M^Q \to \mathbb{R} : \hat{\gamma} \mapsto \hat{E}(\hat{\gamma})$ that we would optimize on the finite-dimensional product manifold M^Q using, e.g., a steepest-descent method described in [AMS08]. The two approaches yield considerably different expressions for the gradient. In particular, in the approach on $\hat{E}$, the gradient of the fitting term $d^2(\gamma(t_i), p_i)$, with respect to the product Riemannian metric on M^Q, vanishes everywhere except at time t_i, whereas with the approach proposed here the influence of the ith fitting term is spread along the whole curve in the expression of its gradient. Although we have not compared the results, one should expect a better performance with the approach where the discretization is delayed until the implementation step.

A Proof of Theorem 3.1

This proof is essentially a restriction of the proof of [Kar77, Th. 1.2]. Let α be defined by $\alpha(t) = \exp_q(t\xi)$. Consider the family of geodesics from p to $\alpha(t)$: $c_p(s,t) = \exp_p(s \exp_p^{-1} \alpha(t))$. Since the cut locus is closed [dC92, corollary 13.2.10], this expression is well defined for all t in a neighborhood of 0 and all $s \in [0,1]$. Denote $c_p' = \frac{\mathrm{d}}{\mathrm{d}s} c_p(s,t)$ and $\dot{c}_p = \frac{\mathrm{d}}{\mathrm{d}t} c_p(s,t)$. We know that $d(p, \alpha(t)) = \|c_p'(s,t)\|$ is independent of s. We have successively

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} d^2(p, \alpha(t)) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} \left\langle c_p'(s,t), c_p'(s,t) \right\rangle$$

which does not depend on s,

$$\begin{aligned} &= \left\langle \frac{\mathrm{D}}{\mathrm{d}t} c_p'(s,t), c_p'(s,t) \right\rangle \\ &= \left\langle \frac{\mathrm{D}}{\mathrm{d}s} \dot{c}_p(s,t), c_p'(s,t) \right\rangle \end{aligned}$$

which still does not depend on s, thus

$$\begin{aligned} &= \int_0^1 \left\langle \frac{\mathrm{D}}{\mathrm{d}s} \dot{c}_p(s,t), c_p'(s,t) \right\rangle \mathrm{d}s \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}s} \left\langle \dot{c}_p(s,t), c_p'(s,t) \right\rangle \mathrm{d}s \end{aligned}$$

since $\frac{\mathrm{D}}{\mathrm{d}s}c'_p(s,t) = 0$ (geodesic property),

$$= \left\langle \dot{c}_p(1,t), c'_p(1,t) \right\rangle$$

since $\dot{c}_p(0,t) = 0$,

$$= \left\langle \dot{\alpha}(t), -\exp_{\alpha(t)}^{-1} p \right\rangle .$$

Since $\frac{1}{2}\mathrm{D}d^2(p,\cdot)(q)[\xi] = \frac{1}{2}\,\frac{\mathrm{d}}{\mathrm{d}t} d^2(p,\alpha(t))\big|_{t=0}$, the result follows.

B An extrinsic expression for the curvature tensor

The expression of the derivative of $E_{\mathrm{s},2}$ obtained in Section 4.2 involves the Riemannian curvature tensor of the Riemannian manifold M. When M is a Riemannian submanifold of a Riemannian manifold N, the curvature tensor admits an extrinsic expression in terms of the second fundamental form and of the Weingarten map, which turns out to be handy in certain cases. In this section, we present this extrinsic expression, then we work out in detail the particular case where N is a Euclidean space.

Let $\mathrm{D}_\eta\xi$ denote the (Levi-Civita) covariant derivative of ξ along η on the embedding manifold M, and let $\overline{\mathrm{D}}$ denote that derivative on N. The curvature tensor R is defined by

$$R(\xi,\eta)\zeta = \mathrm{D}_\xi\mathrm{D}_\eta\zeta - \mathrm{D}_\eta\mathrm{D}_\xi\zeta - \mathrm{D}_{[\xi,\eta]}\zeta, \tag{28}$$

where ξ, η, ζ are tangent vector fields on M and $[\xi,\eta]$ denotes the Lie bracket. Given $x \in M$, let $P_x : T_xN \to T_xM$ denote the orthogonal projector onto T_xM. Let $T_x^\perp M$ denote the normal space to M at x, and let $P_x^\perp : T_xN \to T_x^\perp M$ denote the orthogonal projector onto the normal space $T_x^\perp M$. The *shape operator* (also called *second fundamental form*) is the object II defined as follows: for all $x \in M$ and all $\xi_x, \eta_x \in T_xM$,

$$II(\eta_x,\xi_x) := P_x^\perp \overline{\mathrm{D}}_{\eta_x}\xi, \tag{29}$$

where ξ is any tangent vector field that extends ξ_x. This definition can be found, e.g., in [O'N83, Cha06]. The *Weingarten map* is the object $\mathfrak{U}$ defined as follows. For all $x \in M$, $\eta_x \in T_xM$, $v_x \in T_x^\perp M$,

$$\mathfrak{U}_{\eta_x}(v_x) := -P_x\overline{\mathrm{D}}_{\eta_x}v, \tag{30}$$

where v is any normal vector field that extends v_x. This definition can be found, e.g., in [Cha06, p. 62]. Then the curvature tensor can be expressed as follows:

$$R(\xi,\eta)\zeta = \mathfrak{U}_\xi II(\eta,\zeta) - \mathfrak{U}_\eta II(\xi,\zeta). \tag{31}$$

Let us now assume that N is a Euclidean space. Then the projector field P can be viewed as a matrix-valued function, the shape operator admits the expression

$$II(\eta_x,\xi_x) = (\overline{\mathrm{D}}_{\eta_x}P)\xi_x, \tag{32}$$

where $\overline{\mathrm{D}}$ now reduces to the classical derivative, and the Weingarten map takes the form

$$\mathfrak{U}_{\eta_x}(v_x) = (\overline{\mathrm{D}}_{\eta_x}P)v_x. \tag{33}$$

We refer to [ATMA09] for details. These formulas are particularly useful when the projector field P admits a simple expression.

References

[AG09] P.-A. Absil and K. A. Gallivan. Accelerated line-search and trust-region methods. *SIAM J. Numer. Anal.*, 47(2):997–1018, 2009.

[Alt00] C. Altafini. The de Casteljau algorithm on $SE(3)$. In *Book chapter, Nonlinear control in the Year 2000*, pages 23–34, 2000.

[AMA05] P.-A. Absil, R. Mahony, and B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM J. Optim.*, 6(2):531–547, 2005.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[ATMA09] P.-A. Absil, Jochen Trumpf, Robert Mahony, and Ben Andrews. All roads lead to Newton: Feasible second-order methods for equality-constrained optimization, 2009. Technical report UCL-INMA-2009.024.

[Ber95] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific, Belmont, Massachusetts, 1995.

[Boo03] William M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 2003. Revised Second Edition.

[Cha06] Isaac Chavel. *Riemannian geometry*, volume 98 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2006. A modern introduction.

[CKS99] P. Crouch, G. Kun, and F. Silva Leite. The de Casteljau algorithm on the Lie group and spheres. In *Journal of Dynamical and Control Systems*, volume 5, pages 397–429, 1999.

[CS91] P. Crouch and F. Silva Leite. Geometry and the dynamic interpolation problem. In *Proc. Am. Control Conf., Boston, 26–29 July, 1991*, pages 1131–1136, 1991.

[CS95] P. Crouch and F. Silva Leite. The dynamic interpolation problem: on Riemannian manifolds, Lie groups, and symmetric spaces. *J. Dynam. Control Systems*, 1(2):177–202, 1995.

[CSC95] M. Camarinha, F. Silva Leite, and P. Crouch. Splines of class C^k on non-Euclidean spaces. *IMA J. Math. Control Inform.*, 12(4):399–410, 1995.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[GHL04] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*. Universitext. Springer-Verlag, Berlin, third edition, 2004.

[HS07] K. Hüper and F. Silva Leite. On the geometry of rolling and interpolation curves on S^n, SO_n, and Grassmann manifolds. *J. Dyn. Control Syst.*, 13(4):467–502, 2007.

[JK87] Peter E. Jupp and John T. Kent. Fitting smooth paths to spherical data. *J. Roy. Statist. Soc. Ser. C*, 36(1):34–46, 1987.

[JSR06] Janusz Jakubiak, Fátima Silva Leite, and Rui C. Rodrigues. A two-step algorithm of smooth spline generation on Riemannian manifolds. *J. Comput. Appl. Math.*, 194(2):177–191, 2006.

[Kar77] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure. Appl. Math.*, 30(5):509–541, 1977.

[KDL07] Alfred Kume, Ian L. Dryden, and Huiling Le. Shape-space smoothing splines for planar landmark data. *Biometrika*, 94(3):513–528, 2007.

[KS06] E. Klassen and A. Srivastava. Geodesic between 3D closed curves using path straightening . In A. Leonardis, H. Bischof, and A. Pinz, editors, *European Conference on Computer Vision*, pages 95–106, 2006.

[Lee07] John M. Lee. *Riemannian Manifolds: An Introduction to Curvature*, volume 176 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, NY, 2007.

[Lin03] Anders Linnér. Symmetrized curve-straightening. *Differential Geom. Appl.*, 18(2):119–146, 2003.

[LT66] Michel Lazard and Jacques Tits. Domaines d'injectivité de l'application exponentielle. *Topology*, 4:315–322, 1965/1966.

[Mil63] J. W. Milnor. *Morse Theory*. Princeton University Press, New Jersey, 1963.

[MLK10] L. Machado, F. Silva Leite, and K. Krakowski. Higher-order smoothing splines versus least squares problems on Riemannian manifolds. *J. Dyn. Control Syst.*, 16(1):121–148, 2010.

[MS06] Luís Machado and F. Silva Leite. Fitting smooth paths on Riemannian manifolds. *Int. J. Appl. Math. Stat.*, 4(J06):25–53, 2006.

[MSH06] Luís Machado, F. Silva Leite, and Knut Hüper. Riemannian means as solutions of variational problems. *LMS J. Comput. Math.*, 9:86–103 (electronic), 2006.

[NHP89] Lyle Noakes, Greg Heinzinger, and Brad Paden. Cubic splines on curved spaces. *IMA J. Math. Control Inform.*, 6(4):465–473, 1989.

[O'N83] Barrett O'Neill. *Semi-Riemannian Geometry*, volume 103 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.

[Pal63] Richard S. Palais. Morse theory on Hilbert manifolds. *Topology*, 2:299–340, 1963.

[PN07] Tomasz Popiel and Lyle Noakes. Bézier curves and C^2 interpolation in Riemannian manifolds. *J. Approx. Theory*, 148(2):111–127, 2007.

[SZ04] G. Smyrlis and V. Zisis. Local convergence of the steepest descent method in Hilbert spaces. *J. Math. Anal. Appl.*, 300(2):436–453, 2004.

[Tro77] A. J. Tromba. A general approach to Morse theory. *J. Differential Geometry*, 12(1):47–85, 1977.