MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Fitting Curves on Riemannian Manifolds Using Energy Minimization

Chafik Samir[†], P.-A. Absil[†], Anuj Srivastava[‡], and Eric Klassen[§]

[†]Department of Mathematical Engineering, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.

[‡]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA.

[§]Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA.

Abstract

Given data points $p_0, \ldots, p_N$ on a Riemannian manifold $\mathcal{M}$ and time instants $0 = t_0 < t_1 < \ldots < t_N = 1$, we consider the problem of finding the curve γ on $\mathcal{M}$ that best approximates the data points at the given instants. In this work, γ is expressed as the curve that minimizes the weighted sum of a least-squares term penalizing the lack of fitting to the data points and a regularity term defined as the mean squared velocity of the curve. The optimization task is carried out by means of a steepest-descent algorithm on the set of continuous paths on $\mathcal{M}$. The steepest-descent direction, defined in the sense of the Palais metric, is shown to admit a simple formula based on parallel translation.

1 Introduction

1.1 Motivation

Let $p_0, p_1, \ldots, p_N$ be a finite set of points in $\mathbb{R}^n$, or more generally on a Riemannian manifold $\mathcal{M}$, and let $0 = t_0 < t_1 < \ldots < t_N = 1$ be different instants of time. The problem of fitting a curve γ in $\mathcal{M}$ 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, e.g., in the sense of the functional E_d defined by:

$$E_d(\gamma) = \sum_{i=0}^{N} d^2(\gamma(t_i), p_i),$$

where d denotes the geodesic distance function on the Riemannian manifold $\mathcal{M}$. (In the case where $\mathcal{M}$ is the Euclidean space $\mathbb{R}^n$ endowed with its canonical metric, $E_d(\gamma)$ is simply $\sum_{i=0}^{N} \|\gamma(t_i) - p_i\|^2$, where $\|\cdot\|$ denotes the Euclidean norm). The second goal is that the curve should be sufficiently "regular", in certain sense, e.g., the length of the curve should be as small as possible, or the changes in velocity should be minimized.

Curve fitting problems on manifolds appear in various applications. To cite but one example, let $(I_i)_{i \le N}$ 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 estimate the missing data and recover the motion of the object as well as possible [6]. 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 result is likely to be improved if inherent regularity properties of the object motion are taken into account [5].

1.2 Previous work

One extreme way of handling the two conflicting goals to the curve fitting problem is to impose a *regularity constraint* and minimize the criteria E_d. When $\mathcal{M} = \mathbb{R}^n$, a classical regularity constraint consists of restricting the curve γ to the family $\mathcal{P}_m$ of polynomial functions of degree not exceeding m, $(m \le N)$. This least-square problem, introduced by Lagrange $(1736 - 1813)$, cannot be straightforwardly generalized to an arbitrary Riemannian manifold $\mathcal{M}$ because the notion of polynomial does not carry to $\mathcal{M}$ 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 [6]. The problem of fitting geodesics to data on Riemannian manifold $\mathcal{M}$ was considered by several authors ([13], [11],[4]) for the case where $\mathcal{M}$ is the special orthogonal group $\mathrm{SO}(n)$ or the unit sphere $\mathbb{S}^n$.

The other extreme approach to the curve fitting problem is to seek γ that optimizes a regularity criterion E_s under the *interpolation constraint* $E_d(\gamma) = 0$. For example, when $\mathcal{M} = \mathbb{R}^n$, minimizing the functional E_s defined by

$$E_s(\gamma) := \int_0^1 \|\dot{\gamma}(t)\|^2 dt \qquad (1)$$

yields an optimal curve γ composed of line segments between the data points $p_0, \ldots, p_N$. Another example is the criterion

$$E_s(\gamma) := \int_0^1 \|\ddot{\gamma}(t)\|^2 dt, \qquad (2)$$

which yields solutions known as cubic splines. This concept has been investigated by several authors when $\mathcal{M}$ is a nonlinear manifold, motivated by different types of application. Crouch *et al.* [3] implemented the de Casteljau algorithm on Lie groups and on m-dimensional spheres under some boundary conditions. More recently, Jakubiak *et al.* [5] presented a geometric algorithm to generate splines of an arbitrary degree of smoothness in Euclidean spaces and then extended it to matrix Lie groups. They applied their algorithm to design a smooth motion of a 3D object in space. Using an unrolling and unwarpping procedure in Riemannian manifolds, Kume *et al.* [10] developed a new method to fit smooth curves through a series of shape of landmarks.

1.3 Our approach

In this paper, we choose a middle way to the two extreme approaches mentioned above. In the spirit of the work of Machado and Leite [13], we consider the problem of minimizing a functional of the form:

$$\begin{aligned} E: \quad & \Gamma \to \mathbb{R} \\ & \gamma \mapsto E(\gamma) \quad := E_d(\gamma) + \lambda E_s(\gamma) \end{aligned}$$

where Γ is the set $H^1([0,1],\mathcal{M})$ of all the continuous paths $\gamma : [0,1] \to \mathcal{M}$ whose weak first derivative is locally square integrable in every chart of $\mathcal{M}$, i.e.,

$$\Gamma = \{\gamma : [0,1] \to \mathcal{M} \mid \gamma \in C^0, \int_0^1 \|\dot{\gamma(t)}\|^2 dt < \infty\}, \tag{3}$$

and where λ is a positive real constant, termed *regularity parameter*. The parameter λ makes it possible to mitigate between the two conflicting goals mentioned above. When λ is large, the emphasis is laid on the regularity condition and less so on the fitting condition, whereas when λ is small, the fitting condition dominates. Observe that there is no constraint on γ further than belonging to Γ.

In [13], the regularity cost function E_s is chosen as

$$\int_0^1 \left\langle \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}, \frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2} \right\rangle_{\gamma(t)} \mathrm{d}t, \tag{4}$$

where $\langle \cdot, \cdot \rangle_x$ denotes the Riemannian metric at $x \in \mathcal{M}$ and $\frac{\mathrm{D}^2\gamma}{\mathrm{d}t^2}$ denotes the covariant derivative of the velocity vector field, $\dot{\gamma}$, along γ. This is a natural generalization of (2) to manifolds endowed with an affine connection. The choice of (2) is motivated by the fact that cubic splines in $\mathbb{R}^n$ can be viewed as extrema of (2) under the interpolation condition [13, Prop 4.6]. The main result in [13] is to give a necessary condition of optimality for γ to be a minimizer of $E_d + \lambda E_s$. The necessary condition takes the form of a fourth-order ordinary differential equation for γ involving the covariant derivative and the curvature tensor.

The present paper differs from [13] in two ways. First, instead of (4), we focus on the regularity cost function E_s defined as (1), which penalizes the average squared velocity of the curve. In summary, the function to be minimized is

$$\begin{aligned} & E : \Gamma \to \mathbb{R} : \gamma \mapsto E(\gamma) = E_d(\gamma) + \lambda E_s(\gamma) \\ = & \sum_{i=0}^{N} d^2(\gamma(t_i), p_i) + \lambda \int_0^1 \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} \, \mathrm{d}t. \end{aligned} \tag{5}$$

Other forms of E_s, such as (2) will be considered in later work. The second difference is that we propose an explicit numerical algorithm for computing the minimizer of $E_d+\lambda E_s$. The procedure is a steepest-descent iteration on the search space Γ. The chosen descent direction is steepest in the sense of the Palais metric [12], a Riemannian metric introduced recently to define gradient flows of various geometric minimal paths energies [9]. A major advantage of using the Palais metric in this context is that the gradient of E (whose negative provides the steepest-descent direction) admits a simple expression involving the parallel translation on the manifold $\mathcal{M}$. The step size of the steepest-descent iteration is selected using an Armijo backtracking procedure, but any other efficient step size selection method would have been suitable.

The paper is organized as follows. In section 1 we formulated the problem of fitting curves as an optimization problem for the functional E. Section 2 summarizes some definitions and properties, from differential geometry. In section 3 we compute the gradient of E with respect to the Palais metric, and in section 4 we formulate a gradient descent method to minimize the functional E. Some examples are performed in section 5, for $\mathcal{M} = \mathbb{R}^2$ and $\mathcal{M} = \mathbb{S}^2$, to assess the efficiency of our method. Conclusions and future work are presented in section 6.

2 Definitions

In this section, we recall some concepts of Riemannian geometry that we require for our analysis. For more details, we refer, e.g., to [12]. In what follows, $\mathcal{M}$ denotes a Riemannian manifold endowed with its Riemannian structure, Γ as defined in (3), and γ an element of Γ.

Definition 1 (Tangent space) *The tangent space of Γ at γ is given by*

$$T_\gamma(\Gamma) = \{\alpha \in \mathcal{H}^1([0,1], T\mathcal{M}) \mid \alpha(t) \in T_{\gamma(t)}(\mathcal{M}), \ \forall t \in [0,1]\}.$$

Definition 2 (Covariant derivative) *Let $w \in T_\gamma(\mathcal{M})$ be a vector field along γ. The covariant derivative of w (induced by the Riemannian connection) is a vector field denoted $\frac{Dw}{dt}$, and obtained by projecting $\frac{dw}{dt}(t)$ onto the tangent space $T_{\gamma(t)}(\mathcal{M})$.*

Definition 3 (Covariant integral) *A vector field $w \in T_\gamma(\mathcal{M})$ is called a covariant integral of v along γ if the covariant derivative of v is w, i.e. $\frac{Dw}{dt} = v$.*

Definition 4 (Covariantly constant) *A vector field w is covariantly constant or parallel along the curve γ if and only if $\frac{Dw}{dt}(t) = 0$ for all t.*

Definition 5 (Parallel transport) *A vector field $\tilde{v}$ is called the parallel transport of a vector field $v \in T_{\gamma(0)}(\mathcal{M})$, along γ if and only if $\tilde{v}(0) = v$ and $\frac{D\tilde{v}}{dt} = 0$, for all $t \in [0,1]$.*

Definition 6 (Critical point) *Let $E : \Gamma \to \mathbb{R}$ be a real valued function on Γ. A point $\gamma \in \Gamma$ is a critical point of E if the gradient of E vanishes at γ, i.e. $\nabla E(\gamma) = 0$.*

Observe that the definition of the (non-degenerate) Riemannian metric does not affect the critical points: if ∇E vanishes at γ in one Riemannian metric, then it does so in all Riemannian metrics. However, the basins of attraction of the local minima for a given steepest-decent algorithm will in general depend on the Riemannian metric.

3 Gradient with respect to the Palais metric

For the gradient of E (and thus the steepest-descent direction of E) to be well-defined, a Riemannian structure (i.e., a Riemannian metric) is needed on Γ. There is some freedom in this choice. Here we will use the Palais metric [12] defined as follows:

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

with $v(t), w(t) \in T_{\gamma(t)}(\mathcal{M})$. This turns Γ into a Hilbert Riemannian manifold, and the Riesz representation theorem ensures that gradients of differentiable real-valued functions on Γ are well defined (see [7, §6.11]). In this section, we show that the gradient of E (5) with respect to the Palais metric admits a simple expression involving the parallel transport on $\mathcal{M}$.

3.1 Derivation of data term

We first obtain an expression for the gradient of the i^{th} term in E_d, namely, $\gamma \mapsto d^2(\gamma(t_i), p_i)$. Let v_i be the smallest tangent vector at $\gamma(t_i)$ such that $\exp_{\gamma(t_i)}(v_i) = p_i$, where exp denotes the Riemannian exponential as defined in [2]. We assume that v_i exists and is unique. (Note that this is the case for almost all pairs $(\gamma(t_i), p_i)$ on $\mathbb{R}^n$, the unit sphere, the special orthogonal group, the Grassmann manifold of p-planes in $\mathbb{R}^n$, and the Stiefel manifold of orthogonal p-frames in $\mathbb{R}^n$ $(p < n)$.) Then the directional derivative of the function $\gamma \mapsto d^2(\gamma(t_i), p_i)$, in the direction of any $w \in T_\gamma(\Gamma)$, is $2\langle w(t_i), v_i\rangle$; see, e.g., [8]. The gradient of $\gamma \mapsto d^2(\gamma(t_i), p_i)$ at γ is thus the element α of $T_\gamma\Gamma$ such that $\langle\langle \alpha, w\rangle\rangle_\gamma = 2\langle w(t_i), v_i\rangle$ for all $w \in T_\gamma(\Gamma)$. The next theorem gives an expression for α.

Theorem 1 *The gradient of the function $\Gamma \to \mathbb{R} : \gamma \mapsto d^2(\gamma(t_i), p_i)$ evaluated at $\gamma \in \Gamma$ is represented by the following vector field along γ:*

$$\alpha_i(t) = \begin{cases} 2t\tilde{v}_i(t) + 2\tilde{v}_i(t), & 0 \le t \le t_i \\ 2t_i\tilde{v}_i(t) + 2\tilde{v}_i(t), & t_i \le t \le 1 \end{cases},$$

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

The proof is removed due to space limitation.

Observe that $\tilde{v}_i(t_i) = v_i$, α_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 α_i is covariantly constant ($\tilde{v}_i$) until t_i, after that it is 0.

Finally, the gradient of the full data term E_d, under the Palais metric, is $\alpha = \sum_{i=1}^N \alpha_i$.

3.2 Derivation of length term

As shown in [9] the gradient of the functional $E_s(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|^2 dt$, with respect to the Palais metric, is simply the covariant integral of the velocity vector $\dot{\gamma}$, vanishing at $t = 0$.

Theorem 2 *The vector field β along γ that provides the gradient of the function $\Gamma \to \mathbb{R} : \gamma \mapsto E_s(\gamma) = \int_0^1 \langle\dot{\gamma}(t), \dot{\gamma}(t)\rangle_{\gamma(t)}$ dt is given by the following equation:*

$$\frac{D\beta(t)}{dt} = \dot{\gamma}(t), \quad \beta(0) = 0$$

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

4 Energy Minimization

In this section, we describe a steepest-descent algorithm for minimizing the energy function E (5). The algorithm is described conceptually in the space Γ of continuous curves on $\mathcal{M}$.

4.1 Conceptual algorithm in Γ

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

$$\gamma_{k+1} = \gamma_k - \hat{\rho}_k \nabla E(\gamma_k),$$

where $\hat{\rho}_k$ is a step size chosen using some step size selection rule (see, e.g., [1]).

A formal specification of our gradient descent algorithm is given in algorithm 1.

Algorithm 1 : Gradient descent

Require: a scalar $\epsilon \in]0, 1[$.
Input: initial iterate γ_0, an arbitrary curve in Γ.
Output: an optimal curve $\hat{\gamma}$.
1: $k = 0$.
2: ***Repeat*** *until* $\|\nabla E(\gamma_k)\| \le \epsilon$
3: $k = k + 1$.
4: *Compute* $E(\gamma_k)$ *and* $\nabla E(\gamma_k)$.
5: *Find the step size* $\hat{\rho}_k$.
6: *Set* $\gamma_k = \gamma_{k-1} - \hat{\rho}_k \nabla E(\gamma_{k-1})$.
7: ***End repeat.***
8: $\hat{\gamma} = \gamma_k$.

5 Experimental evaluation

In this section we will show some illustrations of our gradient descent method on the Euclidean plane $\mathcal{M} = \mathbb{R}^2$ and on the sphere $\mathcal{M} = \mathbb{S}^2$.

In what follows, and in both cases ($\mathcal{M} = \mathbb{R}^2$ and $\mathcal{M} = \mathbb{S}^2$), we first generate $N + 1$ control data points $p_0, p_1, \ldots, p_N$ randomly on the given manifold $\mathcal{M}$ at different instants of time $0 = t_0 < t_1 < t_2 < \ldots t_n \le t_N = 1$. Then, we initialize our algorithm with any arbitrary continuous curve $\gamma_0 \in \Gamma$, and finally apply our gradient descent method to search for the optimal curve $\hat{\gamma}$ that minimizes E.

5.1 Case 1: $\mathcal{M} = \mathbb{R}^2$

Shown in figure 1 (left panel) are some examples of our approach applied to different starting sets of points generated randomly on $\mathcal{M} = \mathbb{R}^2$ using different values of λ. If λ is very small, we have $E \simeq E_d$, and the solution is the linear piecewise curve passing through the given points. If λ is very large, then $E \simeq E_s$ and the curve will shrink to one point in $\mathbb{R}^2$, namely, the center of mass of the given points $p_0, p_1, \ldots, p_N$. Note that the problem on the top left in figure 1 has a simple closed-form solution. We checked that our solution corresponds exactly to this closed-form.

5.2 Case 2: $\mathcal{M} = \mathbb{S}^2$

Here we show some results on a non-linear manifold, the sphere $\mathcal{M} = \mathbb{S}^2$. It is well known that geodesics on $\mathbb{S}^2$ are shorter arcs connecting two points. Shown in figure 1 (right panel) are some examples of our approach applied to different sets of data points generated randomly on $\mathcal{M} = \mathbb{S}^2$. Curves in different colors are obtained with different values of λ. If λ is very small, we have $E \simeq E_d$, and the solution is the geodesic piecewise curve passing through the given points. If λ is very large, $E \simeq E_s$ and the curve will shrink to one point in $\mathbb{S}^2$, precisely the Karcher mean of the given set of points $p_0, p_1, \ldots, p_N$.

6 Summary

We have addressed the problem of fitting a curve to data points on a Riemannian manifold $\mathcal{M}$ by means of a Palais-based steepest-descent algorithm applied to weighted sum of fitting-related and a regularity-related cost function. As a proof of concept, we have used the simple regularity cost function (1), for which the optimal curves are piecewise geodesics. In future work, we will use cost function (4), which is related to curvature.

Acknowledgments

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.

References

[1] Dimitri P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1995.

[2] W. M. Boothby. An introduction to differential manifolds and Riemannian geometry. In *Academic Press, New york*, 1975.

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

[4] S. Florya and M. Hofer. Constrained curve fitting on manifolds. *Computer-Aided Design*, Volume 40, to appear:25–34, 2008.

[5] J. Jakubiak, F. S. Leite, and R.C. Rodrigues. A two-step algorithm of smooth spline generation on Riemannian manifolds. In *Journal of Computational and Applied Mathematics*, pages 177–191, 2006.

[6] S. Joshi, A. Srivastava, and W. Mio. Elastic shape models for interpolations of curves in image sequences. *Information Processing in Medical Imaging (IPMI)*, 3565:541–552, July 2005.

[7] Jürgen Jost. *Riemannian geometry and geometric analysis.* Universitext. Springer-Verlag, Berlin, fourth edition, 2005.

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

[9] E. Klassen and A. Srivastava. Geodesic between 3d closed curves using path straightening. In *European Conference on Computer Vision*, 2006.

[10] A. Kume, I. L. Dryden, H. Le, and A. T. A. Wood. Fitting cubic splines to data in shape spaces of planar configurations. In *Proceedings in Statistics of Large Datasets, LASR, 119-122*, 2002.

[11] P. Lancaster and K. Salkauskas. Curve and surface fitting. In *Academic Press*, 1986.

[12] J. Langer and D. A. Singer. Curve-straghtening in Riemannian manifolds. In *Ann. Global Anal. Geom.*, pages 133–150, 1987.

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

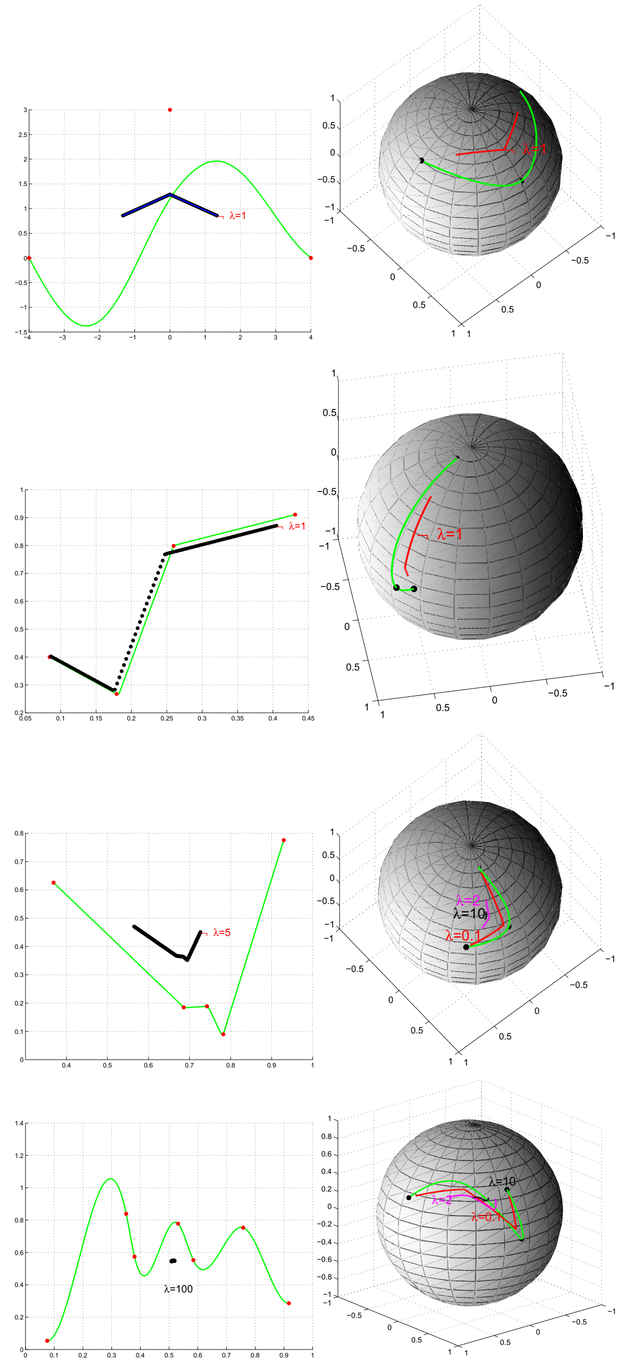

Figure 1: The minimum of E in $\mathcal{M} = \mathbb{R}^2$ in the left side and $\mathcal{M} = \mathbb{S}^2$ in the right, reached by the gradient descent method with respect to Palais metric using different values of λ.