Intrinsic Polynomials for Regression on Riemannian Manifolds

Jacob Hinkle · P. Thomas Fletcher · Sarang Joshi

Abstract We develop a framework for polynomial regression on Riemannian manifolds. Unlike recently developed spline models on Riemannian manifolds, Riemannian polynomials offer the ability to model parametric polynomials of all integer orders, odd and even. An intrinsic adjoint method is employed to compute variations of the matching functional, and polynomial regression is accomplished using a gradient-based optimization scheme. We apply our polynomial regression framework in the context of shape analysis in Kendall shape space as well as in diffeomorphic landmark space. Our algorithm is shown to be particularly convenient in Riemannian manifolds with additional symmetry, such as Lie groups and homogeneous spaces with right or left invariant metrics. As a particularly important example, we also apply polynomial regression to time-series imaging data using a right invariant Sobolev metric on the diffeomorphism group. The results show that Riemannian polynomials provide a practical model for parametric curve regression, while offering increased flexibility over geodesics.

Keywords Polynomial · Riemannian geometry · Regression · Rolling maps · Lie groups · Shape space

1 Introduction

Comparative studies are essential to biomedical statistical analysis. In the context of shape, such analyses are used to discriminate between healthy and disease states based on observations of anatomical shapes within individuals in the two populations [35]. Commonly, in these methods the shape data are modelled on a Riemannian manifold and intrinsic coordinate-free manifold-based methods are used [8]. This prevents bias due to arbitrary choice of coordinates and avoids the influence of unwanted effects. For instance, by modelling shapes with a representation incapable of representing scale and rotation of an object and using intrinsic manifold-based methods, scale and rotation are guaranteed not to effect the analysis [19].

Many conditions such as developmental disorders and neurodegeneration are characterized not only by shape characteristics, but by abnormal *trends* in anatomical shapes over time. Thus it is often the temporal dependence of shape that is most useful for comparative shape analysis. The field of regression analysis involves studying the connection between independent variables and observed responses [34]. In particular, this includes the study of temporal trends in a observed data.

In this work, we extend the recently developed geodesic regression model [12] to higher order polynomials using intrinsic Riemannian manifold-based methods. We show that this Riemannian polynomial model is able to provide increased flexibility over geodesics, while remaining in the parametric regression setting. The increase in flexibility is particularly important, as it enables a more accurate description of shape trends and, ultimately, more useful comparative regression analysis.

While our primary motivation is shape analysis, the Riemannian polynomial model is applicable in a variety of applications. For instance, directional data is commonly modelled as points on the sphere $\mathbb{S}^2$, and video sequences representing human activity are modelled in Grassmannian manifolds [36].

In computational anatomy applications, the primary objects of interest are elements of a group of symmetries acting

J. Hinkle (✉) · P.T. Fletcher · S. Joshi
SCI Institute, University of Utah, 72 Central Campus Dr., Salt Lake City, Utah 84112, USA
e-mail: jacob@sci.utah.edu

on the space of observable data. For instance, rigid motion is studied using the groups SO(3) and SE(3), acting on a space of landmark points or scalar images. Non-rigid motion and growth is modelled using infinite-dimensional diffeomorphism groups, such as in the currents framework [37] for unlabelled landmarks or the large deformation diffeomorphic metric mapping (LDDMM) framework of deforming images [30]. We show that in the presence of a group action, optimization of our polynomial regression model using an adjoint method is particularly convenient.

This work is an extension of the Riemannian polynomial regression framework first presented by Hinkle et al. [15]. In Sects. 5–7, we give a new derivation of polynomial regression for Lie groups and Lie group actions with Riemannian metrics. By performing the adjoint optimization directly in the Lie algebra, the computations in these spaces are greatly simplified over the general formulation. We show how this Lie group formulation can be used to perform polynomial regression on the space of images acted on by groups of diffeomorphisms.

1.1 Regression Analysis and Curve-Fitting

The study of the relationship between measured data and descriptive variables is known as the field of regression analysis. As with most statistical techniques, regression analyses can be broadly divided into two classes: parametric and non-parametric. The most widely used parametric regression methods are linear and polynomial regression in Euclidean space, wherein a linear or polynomial function is fit in a least-squares fashion to observed data. Such methods are the staple of modern data analysis. The most common non-parametric regression approaches are kernel-based methods and spline smoothing approaches which provide great flexibility in the class of regression functions. However, their non-parametric nature presents a challenge to inference problems; if, for example, one wishes to perform a hypothesis test to determine whether the trend for one group of data is significantly different from that of another group.

In previous work, non-parametric kernel-based and spline-based methods have been extended to observations that lie on a Riemannian manifold with some success [8, 18, 22, 26], but intrinsic parametric regression on Riemannian manifolds has received limited attention. Recently, Fletcher [12] and Niethammer et al. [31] have each independently developed a form of parametric regression, geodesic regression, which generalizes the notion of linear regression to Riemannian manifolds. Geodesic models are useful, but are limited by their lack of flexibility when modelling complex trends.

Fletcher [12] defines a geodesic regression model by introducing a manifold-valued random variable Y,

$$Y = \text{Exp}\big(\text{Exp}(p, Xv), \epsilon\big), \tag{1}$$

where $p \in M$ is an initial point and $v \in T_p M$ an initial velocity. The geodesic curve $\text{Exp}(p, Xv)$ then relates the independent variable $X \in R$ to the dependent random variable Y, via this equation and the Gaussian random vector $\epsilon \in T_{\text{Exp}(p, Xv)} M$. In this paper, we extend this model to a polynomial regression model

$$Y = \text{Exp}\big(\gamma(X), \epsilon\big), \tag{2}$$

where the curve $\gamma(X)$ is a Riemannian polynomial of integer order k. In the case that M is Euclidean space, this model is simply

$$Y = p + \sum_{i=1}^{k} \frac{v_i}{i!} X^i + \epsilon, \tag{3}$$

where the point p and vectors v_i constitute the parameters of our model.

In this work we use the common term *regression* to describe methods of fitting polynomial curves using a sum of squared error penalty function. In Euclidean spaces, this is equivalent to solving a maximum likelihood estimation problem using a Gaussian noise model for the observed data. In Riemannian manifolds, the situation is more nuanced, as there is no consensus on how to define Gaussian distributions on general Riemannian manifolds, and in general the least-squares penalty may not correspond to a log likelihood. Many of the examples we will present are symmetric spaces: Kendall shape space in two dimensions, the rotation group, and the sphere, for instance. As Fletcher [12, Sect. 4] explains, least-squares regression in symmetric spaces does, in fact, correspond to maximum likelihood estimation of model parameters, using a natural definition of Gaussian distribution.

1.2 Previous Work: Cubic Splines

Noakes et al. [32] first introduced the notion of Riemannian cubic splines. They fix the endpoints $y_0, y_1 \in M$ of a curve, as well as the derivative of the curve at those points $y'_0 \in T_{y_0} M$, $y'_1 \in T_{y_1} M$. A Riemannian cubic spline is then defined as any differentiable curve $\gamma : [0, 1] \to M$ taking on those endpoints and derivatives and minimizing

$$\Phi(\gamma) = \int_0^1 \left\langle \nabla_{\frac{d}{dt}\gamma} \frac{d}{dt} \gamma(t), \nabla_{\frac{d}{dt}\gamma} \frac{d}{dt} \gamma(t) \right\rangle dt. \tag{4}$$

As is shown by Noakes et al. [14, 32], between endpoints, cubic splines satisfy the following Euler-Lagrange equation:

$$\nabla_{\frac{d}{dt}\gamma} \frac{d}{dt} \gamma + R\left(\nabla_{\frac{d}{dt}\gamma} \frac{d}{dt} \gamma, \frac{d}{dt} \gamma \right) \frac{d}{dt} \gamma = 0. \tag{5}$$

Cubic splines are useful for interpolation problems on Riemannian manifolds. However, cubic splines provide an

insufficient model for parametric curve regression. For instance, by increasing the order of derivatives in Eq. (4), cubic splines are generalizable to higher order curves. Still, only odd order splines may be defined in this way, and there is no clear way to define even order splines.

Riemannian splines are parametrized by the endpoint conditions, meaning that the space of curves is naturally explored by varying control points. This is convenient if control points such as observed data are given at the outset. However, for parametric curve regression, curve models are preferred that don't depend on the data, such as the initial conditions of a geodesic [12]. Although Eq. (5) provides an ODE which could be used as such a parametric model in a "spline shooting" algorithm, estimating initial position and derivatives as parameters, the curvature term complicates integration and optimization.

1.3 Contributions in This Work

The goal of the current work is to extend the geodesic regression model in order to accommodate more flexibility while remaining in the parametric setting. The increased flexibility introduced by the methods in this manuscript allow a better description of the variability in the data. The work presented in this paper allows one to fit polynomial regression curves on a general Riemannian manifold, using intrinsic methods and avoiding the need for unwrapping and rolling. Since our model includes time-reparametrized geodesics as a special case, information about time dependence is also obtained from the regression without explicit modeling by examining the collinearity of the estimated parameters.

We derive practical algorithms for fitting polynomial curves to observations in Riemannian manifolds. The class of polynomial curves we use, described by Leite & Krakowski [24], is more suited to parametric curve regression than are spline models. These polynomials curves are defined for any integer order and are naturally parametrized via initial conditions instead of control points. We derive explicit formulas for computing derivatives with respect to the initial conditions of these polynomials in a least-squares curve-fitting setting.

In the following sections, we describe our method of fitting polynomial curves to data lying in various spaces. We develop the theory for general Riemannian manifolds, Lie groups with right invariant metrics, and finally for spaces acted on by such Lie groups. In order to keep each application somewhat self-contained, results will be shown in each case in the section in which the associated space is treated, instead of in a separate results section following all the methods.

2 Riemannian Geometry Preliminaries

Before defining Riemannian polynomials, we first review a few basic results from Riemannian geometry and establish a common notation. For a more in-depth treatment of this background material see, for instance, do Carmo [9]. Let (M, g) be a Riemannian manifold. At each point $p \in M$, the metric g defines an inner product on the tangent space $T_p M$. The metric also provides a method to differentiate vector fields with respect to one another, referred to as the covariant derivative. For smooth vector fields $v, w \in \mathfrak{X}(M)$ and a smooth curve $\gamma : [0, 1] \to M$ the covariant derivative satisfies the following product rule:

$$\frac{d}{dt}\langle v(\gamma(t)), w(\gamma(t))\rangle = \langle \nabla_{\frac{d}{dt}\gamma} v(\gamma(t)), w(\gamma(t))\rangle$$
$$+ \langle v(\gamma(t)), \nabla_{\frac{d}{dt}\gamma} w(\gamma(t))\rangle. \quad (6)$$

A geodesic $\gamma : [0, 1] \to M$ is characterized (for instance) by the conservation of kinetic energy along the curve:

$$\frac{d}{dt}\left\langle \frac{d}{dt}\gamma, \frac{d}{dt}\gamma \right\rangle = 0 = 2\left\langle \nabla_{\frac{d}{dt}\gamma} \frac{d}{dt}\gamma, \frac{d}{dt}\gamma \right\rangle. \quad (7)$$

which leads to the differential equation

$$\nabla_{\frac{d}{dt}\gamma} \frac{d}{dt}\gamma = 0. \quad (8)$$

This is called the geodesic equation and uniquely determines geodesics, parametrized by the initial conditions $(\gamma(0), \frac{d}{dt}\gamma(0)) \in TM$. The mapping from the tangent space at p into the manifold M, defined by integration of the geodesic equation, is called the exponential map and is written $\mathrm{Exp}_p : T_p M \to M$. The exponential map is injective on a zero-centered ball B in $T_p M$ of some non-zero radius. Thus, for a point q within a neighborhood of p, there exists a unique vector $v \in T_p M$ corresponding to a minimal length path under the exponential map from p to q. The mapping of such points q to their associated tangent vectors v at p is called the log map of q at p, denoted $v = \mathrm{Log}_p q$.

Given a curve $\gamma : [0, 1] \to M$, the covariant derivative $\nabla_{\frac{d}{dt}\gamma}$ provides a way to relate tangent vectors at different points along γ. A vector field w is said to be parallel transported along γ if it satisfies the parallel transport equation,

$$\nabla_{\frac{d}{dt}\gamma} w(\gamma(t)) = 0. \quad (9)$$

Notice that the geodesic equation is a special case of parallel transport, under which the velocity is parallel along the curve itself.

3 Riemannian Polynomials

We now introduce Riemannian polynomials as a generalization of geodesics [15]. Geodesics are generalizations to the

Riemannian manifold setting of curves in $\mathbb{R}^d$ with constant first derivative. In the previous section we briefly reviewed how the covariant derivative provides a way to define vector fields which are analogous to constant vector fields along γ, via parallel transport.

We refer to the vector field $\nabla_{\frac{d}{dt}\gamma}\frac{d}{dt}\gamma(t)$ as the acceleration of the curve γ. Curves with parallel acceleration are generalizations of curves in $\mathbb{R}$ whose coordinates are second order polynomials, and satisfy the second order polynomial equation,

$$(\nabla_{\frac{d}{dt}\gamma})^2 \frac{d}{dt}\gamma(t) = 0. \tag{10}$$

Extending this idea, a cubic polynomial is a curve with parallel jerk (time derivative of acceleration), and so on. Generally, a kth order polynomial in M is defined as a curve $\gamma : [0, 1] \to M$ satisfying

$$(\nabla_{\frac{d}{dt}\gamma})^k \frac{d}{dt}\gamma(t) = 0 \tag{11}$$

for all times $t \in [0, 1]$. As with polynomials in Euclidean space, polynomials are fully determined by initial conditions at $t = 0$:

$$\gamma(0) \in M, \tag{12}$$

$$\frac{d}{dt}\gamma(0) \in T_{\gamma(0)}M, \tag{13}$$

$$(\nabla_{\frac{d}{dt}\gamma})^i \frac{d}{dt}\gamma(0) \in T_{\gamma(0)}M, \quad i = 1, \dots, k-1. \tag{14}$$

Introducing vector fields $v_1(t), \dots, v_k(t) \in T_{\gamma(t)}M$, we write the following system of covariant differential equations, which is equivalent to Eq. (11):

$$\frac{d}{dt}\gamma(t) = v_1(t) \tag{15}$$

$$\nabla_{\frac{d}{dt}\gamma}v_i(t) = v_{i+1}(t), \quad i = 1, \dots, k-1 \tag{16}$$

$$\nabla_{\frac{d}{dt}\gamma}v_k(t) = 0. \tag{17}$$

In this notation, the initial conditions that determine the polynomial are $\gamma(0), v_i(0), i = 1, \dots, k$.

The Riemannian polynomial equations cannot, in general, be solved in closed form, and must be integrated numerically. In order to discretize this system of covariant differential equations, we implement a covariant Euler integrator, depicted in Algorithm 1. A time step Δt is chosen and, at each step of the integrator, $\gamma(t + \Delta t)$ is computed using the exponential map:

$$\gamma(t + \Delta t) = \mathrm{Exp}_{\gamma(t)}\big(\Delta t v_1(t)\big). \tag{18}$$

Each vector v_i is incremented within the tangent space at $\gamma(t)$ and the results are parallel transported infinitesimally

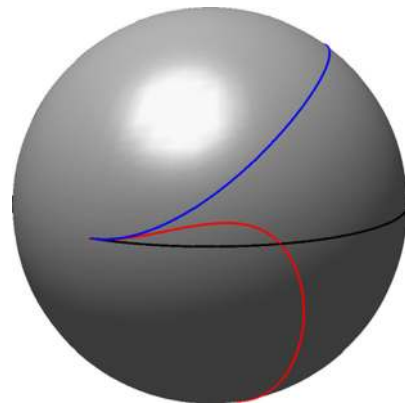**Algorithm 1** Pseudocode for forward integration of kth order Riemannian polynomial

```
γ ← γ(0)
for i = 1, ..., k do
    v_i ← v_i(0)
end for
t ← 0
repeat
    w ← v_1
    for i = 1, ..., k − 1 do
        v_i ← ParTrans(γ, Δt w, v_i + Δt v_{i+1})
    end for
    v_k ← ParTrans(γ, Δt w, v_k)
    γ ← Exp_γ(Δt w)
    t ← t + Δt
until t=T
```


Fig. 1 Sample polynomial curves emanating from a common basepoint on the sphere (*black* = geodesic, *blue* = quadratic, *red* = cubic)

along a geodesic from $\gamma(t)$ to $\gamma(t + \Delta t)$. For a proof that this algorithm approximates the polynomial equations, see Appendix A. The only ingredients necessary to integrate a polynomial are the exponential map and parallel transport on the manifold.

Figure 1 shows the result of integrating polynomials of order one, two, and three on the sphere. The parameters, the initial velocity, acceleration, and jerk, were chosen a priori and a cubic polynomial was integrated to obtain the blue curve. Then the initial jerk was set to zero and the blue quadratic curve was integrated, followed by the black geodesic whose acceleration was also set to zero.

3.1 Polynomial Time Reparametrization

Geodesic curves propagate at a constant speed as a result of their extremal action property. Polynomials provide flexibility not only in the class of paths that are possible, but in the time dependence of the curves traversing those paths. If the parameters of a polynomial γ consist of collinear vectors

$v_i(0) \in T_{\gamma(0)}M$, then the path of γ (the image of the mapping γ) matches that of a geodesic, but the time dependence has been reparametrized by some polynomial transformation $t \mapsto c_0 + c_1 t + c_2 t^2 + c_3 t^3$. This generalizes the existence of polynomials in Euclidean space which are merely polynomial transformations of a straight line path. Regression models could even be implemented in which the operator wishes to estimate geodesic paths, but is unsure of parametrization, and so enforces the estimated parameters to be collinear.

4 Polynomial Regression via Adjoint Optimization

In order to regress polynomials against observed data $J_j \in M$, $j = 1, \ldots, N$ at known times $t_j \in \mathbb{R}$, $j = 1, \ldots, N$, we define the following objective function

$$E_0\big(\gamma(0), v_1(0), \ldots, v_k(0)\big) = \frac{1}{N} \sum_{j=1}^{N} d\big(\gamma(t_j), J_j\big)^2 \qquad (19)$$

subject to the constraints given by Eqs. (15)–(17). Note that in this expression d represents the geodesic distance: the minimum length of a path from the curve point $\gamma(t_j)$ to the data point J_j. The function E_0 is minimized in order to find the optimal initial conditions $\gamma(0), v_i(0), i = 1, \ldots, k$, which we will refer to as the parameters of our model.

In order to determine the optimal parameters of the polynomial, we introduce Lagrange multiplier vector fields λ_i for $i = 0, \ldots, k$, often called the adjoint variables, and define the augmented Lagrangian function

$$E\big(\gamma, \{v_i\}, \{\lambda_i\}\big)$$

$$= \frac{1}{N} \sum_{j=1}^{N} d\big(\gamma(t_j), J_j\big)^2$$

$$+ \int_0^T \Big\langle \lambda_0(t), \frac{d}{dt}\gamma(t) - v_1(t) \Big\rangle dt$$

$$+ \sum_{i=1}^{k-1} \int_0^T \big\langle \lambda_i(t), \nabla_{\frac{d}{dt}\gamma} v_i(t) - v_{i+1}(t) \big\rangle dt$$

$$+ \int_0^T \big\langle \lambda_k(t), \nabla_{\frac{d}{dt}\gamma} v_k(t) \big\rangle dt. \qquad (20)$$

As is standard practice, the optimality conditions for this equation are obtained by taking variations with respect to all arguments of E, integrating by parts when necessary. The resulting variations with respect to the adjoint variables yield the original dynamic constraints: the polynomial equations. Variations with respect to the primal variables gives rise to

the following system of equations, termed the adjoint equations (see B for derivation).

$$\nabla_{\frac{d}{dt}\gamma}\lambda_i(t) = -\lambda_{i-1}(t) \quad i = 1, \ldots, k \qquad (21)$$

$$\nabla_{\frac{d}{dt}\gamma}\lambda_0(t) = -\sum_{i=1}^{k} R\big(v_i(t), \lambda_i(t)\big)v_1(t), \qquad (22)$$

where R is the Riemannian curvature tensor and the adjoint variable λ_0 takes jump discontinuities at time points where data is present:

$$\lambda_0\big(t_j^-\big) - \lambda_0\big(t_j^+\big) = \mathrm{Log}_{\gamma(t_j)} J_j. \qquad (23)$$

Note that this jump discontinuity corresponds to the variation of E with respect to $\gamma(t_j)$. The Riemannian curvature tensor is defined by the formula [9]

$$R(u, v)w = \nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u,v]}w, \qquad (24)$$

and can be computed in closed form for many manifolds. Gradients of E with respect to initial and final conditions give rise to the terminal endpoint conditions for the adjoint variables,

$$\lambda_i(1) = 0, \quad i = 0, \ldots, k \qquad (25)$$

as well as expressions for the gradients with respect to the parameters $\gamma(0), v_i(0)$:

$$\delta_{\gamma(0)}E = -\lambda_0(0), \qquad (26)$$

$$\delta_{v_i(0)}E = -\lambda_i(0). \qquad (27)$$

In order to determine the value of the adjoint vector fields at $t = 0$, and thus the gradients of the functional E_0, the adjoint variables are initialized to zero at time 1, then Eq. (22) is integrated backward in time to $t = 0$.

Given the gradients with respect to the parameters, a simple steepest descent algorithm is used to optimize the functional. At each iteration, $\gamma(0)$ is updated using the exponential map and the vectors $v_i(0)$ are updated via parallel translation. This algorithm is depicted in Algorithm 2.

Note that in the special case of a zero-order polynomial ($k = 0$), the only gradient λ_0 is simply the mean of the log map vectors at the current estimate of the Fréchet mean. So this method generalizes the common method of Fréchet averaging on manifolds via gradient descent [13]. In the case of geodesic polynomials, $k = 1$, the curvature term in Eq. (22) indicates that λ_1 is a sum of Jacobi fields. So this approach subsumes geodesic regression as presented by Fletcher [12]. For higher order polynomials, the adjoint equations represent a generalization of Jacobi field.

As we will see later, in some cases these adjoint equations take a simpler form not involving curvature. In the case that the manifold M is a Lie group, the adjoint equations can be computed by taking variations in the Lie algebra, avoiding explicit curvature computation.

Algorithm 2 Pseudocode for reverse integration of adjoint equations for k^{th} order Riemannian polynomial

---

$\gamma \leftarrow \gamma(T)$
for $i = 0, \ldots, k$ **do**
   $\lambda_i \leftarrow 0$
end for
$t \leftarrow T$
repeat
   $w \leftarrow v_1(t)$
   $\lambda_0 \leftarrow \lambda_0 + \Delta t \sum_{i=1}^{k} R(v_i, \lambda_i) v_1$
   if $t = t_i$ **then**
      $\lambda_0 \leftarrow \lambda_0 + \frac{2}{N} \text{Log}_\gamma J_i$
   end if
   for $i = k, \ldots, 1$ **do**
      $\lambda_i \leftarrow \text{ParTrans}(\gamma, -\Delta t w, \lambda_i + \Delta t \lambda_{i-1})$
   end for
   $\lambda_0 \leftarrow \text{ParTrans}(\gamma, -\Delta t w, \lambda_0)$
   $\gamma \leftarrow \text{Exp}_\gamma(-\Delta t w)$
   $t \leftarrow t - \Delta t$
until t=0
$\delta_{\gamma(0)} E \leftarrow -\lambda_0$
for $i = 1, \ldots, k$ **do**
   $\delta_{v_i(0)} E \leftarrow -\lambda_i$
end for

4.1 Coefficient of Determination (R^2) in Metric Spaces

In order to characterize how well our model fits a given set of data, we define the coefficient of determination of our regression curve $\gamma(t)$, denoted R^2 [12]. As with the usual definition of R^2, we first compute the variance of the data. Naturally, as the data lie on a non-Euclidean metric space, instead of the standard sample variance, we substitute the Fréchet variance, defined as

$$\text{var}\{y_1, \ldots, y_N\} = \frac{1}{N} \min_{\bar{y} \in M} \sum_{j=1}^{N} d(\bar{y}, y_j)^2. \tag{28}$$

The sum of squared error for a curve γ is the value $E_0(\gamma)$:

$$SSE = \frac{1}{N} \sum_{j=1}^{N} d(\gamma(t_j), y_j)^2. \tag{29}$$

We then define R^2 as the amount of variance that has been reduced using the curve γ:

$$R^2 = 1 - \frac{SSE}{\text{var}\{y_1, \ldots, N\}}. \tag{30}$$

Clearly a perfect fit will remove all error, resulting in an R^2 value of one. The worst case ($R^2 = 0$) occurs when no polynomial can improve over a stationary point at the Fréchet mean, which can be considered a zero-order polynomial regression against the data.

4.2 Example: Kendall Shape Space

A common challenge in medical imaging is the comparison of shape features which are independent of easily explained differences such as differences in pose (relative position and rotation). Additionally, scale is often uninteresting as it is easily characterized by volume calculation and explained mostly by intersubject variability or differences in age. It was with this perspective that Kendall [19] originally developed his theory of shape space. Here we briefly describe Kendall's shape space of m-landmark point sets in $\mathbb{R}^d$, denoted Σ_d^m. For a complete treatment of Kendall's shape space, the reader is encouraged to consult Kendall and Le [20, 23].

Given a point set $x = (x_i)_{i=1,\ldots,m}, x_i \in \mathbb{R}^d$, translation and scaling effects are removed by centering and uniform scaling. This is achieved by translating the point set so that the centroid is at zero, then scaling so that $\sum_{i=1}^{m} \|x_i\|^2 = 1$. After this standardization, x constitutes a point in the sphere $\mathbb{S}^{(m-1)d-1}$. This representation of shape is not yet complete as it is effected by global rotation, which we wish to ignore. Thus points on $\mathbb{S}^{(m-1)d-1}$ are referred to as *preshapes* and the sphere $\mathbb{S}^{(m-1)d-1}$ is referred to as *preshape space*. Kendall shape space Σ_d^m is obtained by taking the quotient of the preshape space by the action of the rotation group $\text{SO}(d)$. In practice, points in the quotient (referred to as *shapes*) are represented by members of their equivalence class in preshape space. We describe now how to compute exponential maps, log maps, and parallel transport in shape space, using representatives in $\mathbb{S}^{(m-1)d-1}$. The work of O'Neill [33] concerning Riemannian submersions characterizes the link between the shape and preshape spaces.

The case $d > 2$ is complicated in that these spaces contain degeneracies: points at which the mapping from preshape space to Σ_d^m fails to be a submersion [1, 11, 17]. Despite these pathologies, outside of a singular set, the shape spaces are described by the theory of Riemannian submersions. We assume the data lie within a single "manifold part" away from any singularities, and show experiments in two dimensions, so that these technical issues can be safely ignored.

Each point p in preshape space projects to a point $\pi(p)$ in shape space. The shape $\pi(p)$ is the orbit of p under the action of $\text{SO}(d)$. Viewed as a subset of $\mathbb{S}^{(m-1)d-1}$, this orbit is a submanifold whose tangent space is a subspace of that of the sphere. This subspace is called the vertical subspace of $T_p\mathbb{S}^{(m-1)d-1}$ and its orthogonal complement is the horizontal subspace. Projections onto the two subspaces of a vector $v \in T_p\mathbb{S}^{(m-1)d-1}$ are denoted by $\mathcal{V}(v)$ and $\mathcal{H}(v)$, respectively. Curves moving along vertical tangent vectors result in rotations of a preshape, and so do not indicate any change in actual shape.

A vertical vector in preshape space arises as the derivative of a rotation of a preshape. The derivative of such a rotation

is a skew-symmetric matrix W, and its action on a preshape x has the form $(Wx_1, \ldots, Wx_n) \in T\mathbb{S}^{(m-1)d-1}$. The vertical subspace is then spanned by such tangent vectors arising from any linearly independent set of skew-symmetric matrices. The projection $\mathcal{H}$ is performed by taking such a spanning set, performing Gram-Schmidt orthonormalization, and removing each component.

The horizontal projection allows one to relate the covariant derivative on the sphere to that on shape space. Lemma 1 of O'Neill [33] states that if X, Y are horizontal vector fields at some point p in preshape space, then

$$\mathcal{H}\nabla_X Y = \nabla^*_{X^*} Y^*, \tag{31}$$

where ∇ denotes the covariant derivative on preshape space and ∇^*, X^*, and Y^* are their counterparts in shape space.

For the manifold part of a general shape space Σ^m_d, the exponential map and parallel translation are performed using representatives preshapes in $\mathbb{S}^{(m-1)d-1}$. For $d > 2$, this must be done in a time-stepping algorithm, in which at each time step an infinitesimal spherical parallel transport is performed, followed by the horizontal projection. The resulting algorithm can be used to compute the exponential map as well. Computation of the log map is less trivial, as it requires an iterative optimization routine. A special case arises in the case when $d = 2$, in which case the entire space Σ^m_d is a manifold. In this case the exponential map, parallel transport and log map are computed in closed form [12]. With the exponential map, log map, and parallel transport, one performs polynomial regression on Kendall shape space via the adjoint method described previously.

4.2.1 Rat Calivaria Growth

We have applied polynomial regression in Kendall shape space to the data first analyzed by Bookstein [2], which consists of $m = 8$ landmarks on a midsagittal section of rat calivaria (skulls excluding the lower jaw). The positions of eight identifiable positions on the skull are available for 18 rats and at of eight ages apiece. Figure 2 shows Riemannian polynomial fits of orders $k = 0, 1, 2, 3$. Curves of the same color indicate the synchronized motion of landmarks within a preshape, and the collection of curves for all eight landmarks represents a curve in shape space. While the geodesic curve in Kendall shape space shows little curvature, the quadratic and cubic curves are less linear which demonstrates the added flexibility provided by higher order polynomials. The R^2 values agree with this qualitative difference: the geodesic regression has $R^2 = 0.79$, while the quadratic and cubic regressions have R^2 values of 0.85 and 0.87, respectively. While this shows that there is a clear improvement in the fit due to increasing k from one to two, it also shows that little is gained by increasing the order of the polynomial beyond $k = 2$. Qualitatively, Fig. 2 shows that the slight increase in R^2 obtained by moving from a quadratic to cubic model corresponds to a marked difference in the curves, indicating that the cubic curve is likely overfitting the data. As seen in Table 1, increasing the order of polynomial to four or five has very little effect on R^2 as well.

These results indicate that moving from a geodesic to quadratic model provides an important improvement in fit quality. This is consistent with the results of Kenobi et al. [21], who also found that quadratic and possibly cubic curves are necessary to fit this dataset. However, whereas Kenobi et al. use polynomials defined in the tangent space
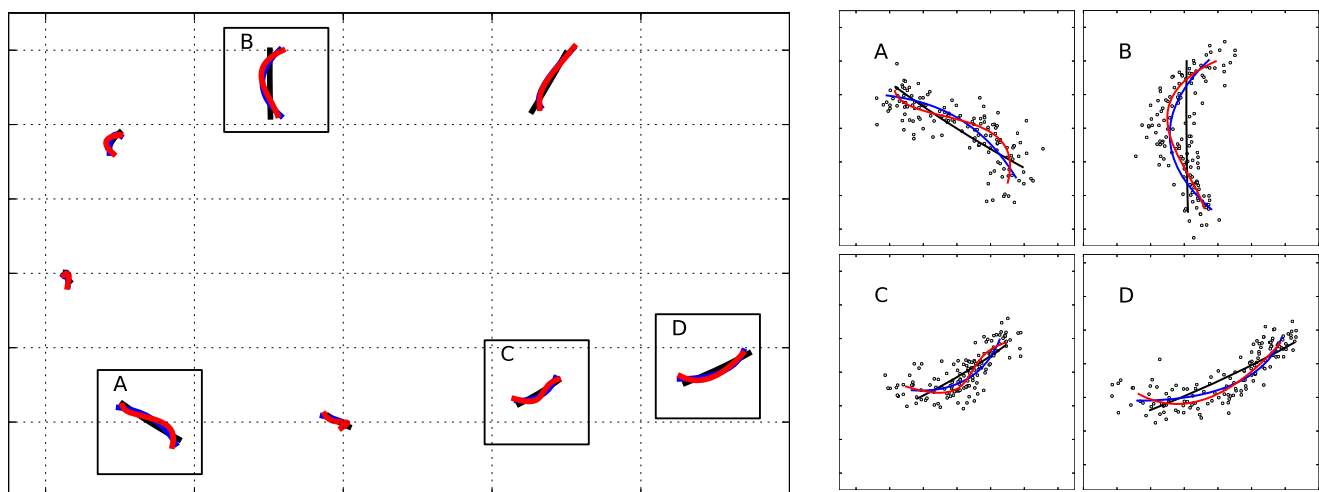
Fig. 2 Bookstein rat calivaria data after uniform scaling and Procrustes alignment. The colors of lines indicate order of polynomial used for the regression (*black = geodesic, blue = quadratic, red = cubic*). Zoomed views of individual rectangles are shown at right, along with data points in *gray*. Note that the axes are arbitrary, due to scale-invariance of Kendall shape space, but that they are the same for the horizontal and vertical axes in these figures

Table 1 R^2 for regression of rat dataset

Polynomial order k	R^2
1	0.79
2	0.85
3	0.87
4	0.87
5	0.87

Table 2 R^2 for regression of corpus callosum dataset

Polynomial order k	R^2
1	0.11
2	0.14
3	0.20
4	0.21
5	0.22

at the Fréchet mean of the data points, the polynomials we use are defined intrinsically, independent of base point.

4.2.2 Corpus Callosum Aging

The corpus callosum, the major white matter bundle connecting the two hemispheres of the brain, is known to shrink during aging [10]. Fletcher showed [12] that more nuanced modes of shape change are observed using geodesic regression. In particular, the volume change observed in earlier studies corresponds to a thinning of the corpus callosum and increased curling of the anterior and posterior regions. In order to investigate even higher modes of shape change of the corpus callosum during normal aging, polynomial regression was performed on data from the OASIS brain database [27]. Magnetic resonance imaging (MRI) scans from 32 normal subjects with ages between 19 and 90 years were obtained from the database and a midsagittal slice was extracted from each volumetric image. The corpus callosum was then segmented on the 2D slices using the ITK-SNAP program [39]. Sets of 64 landmarks for each patient were obtained using the ShapeWorks program [6], which generates samplings of each shape boundary with optimal correspondences among the population.

Regression results for geodesic, quadratic, and cubic regression are shown in Fig. 3. At first glance the results appear similar for the three different models, since the motion envelopes each show the thinning and curling observed by Fletcher. Indeed, the optimal quadratic curve is quite similar to the optimal geodesic, as reflected by their similar R^2 values (0.13 and 0.12, respectively). However, moving from a quadratic to cubic polynomial model delivers a substantial increase in R^2 (from 0.13 to 0.21). This suggests that there are interesting third-order phenomena at work. However, as seen in Table 2, increasing the order beyond three results in very little increase in R^2, indicating that those orders overfit the data, as was the case in the rat calivaria study as well.

Inspection of the estimated parameters for the optimal cubic curve, shown in Fig. 4, reveals that the tangent vectors appear to be collinear. As discussed in Sect. 3.1, this suggests that the cubic curve is a geodesic that has undergone a cubic time reparametrization.

Note that the R^2 values are quite low in this study. Similar values were observed using geodesic regression in [12].
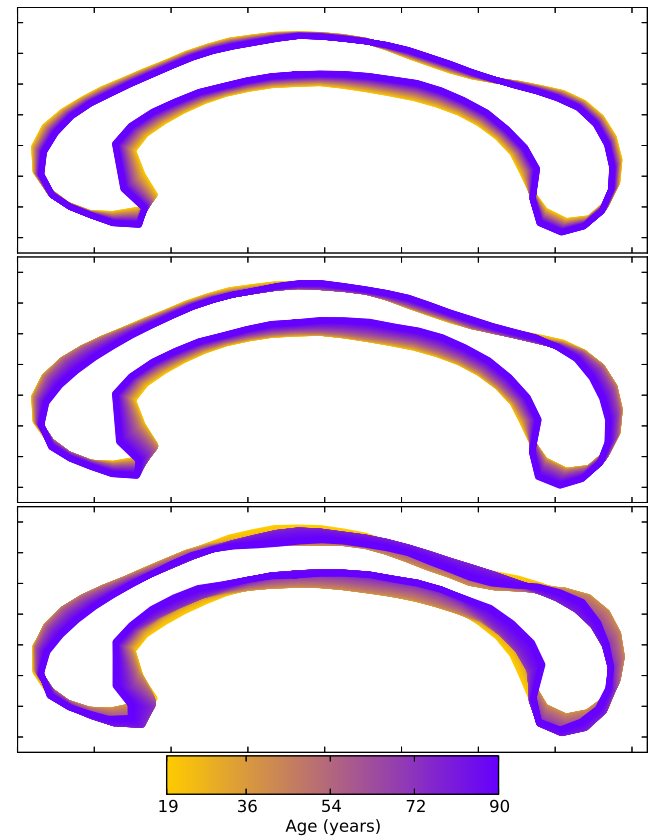

Fig. 3 Geodesic (*top*, $R^2 = 0.12$) quadratic (*middle*, $R^2 = 0.13$) and cubic (*bottom*, $R^2 = 0.21$) regression for corpus callosum dataset. Color represents age, with *yellow* indicating youth (age 19) and *purple* indicating old age (age 90)

As is noted, this is likely due to high inter-subject variability, and that age is only able to explain an effect which is small compared to differences between subjects. Fletcher [12] also notes that although the effect may be small, geodesic regression gives a result which is significant ($p = 0.009$) using a non-parametric permutation test.

Model selection, which in the case of polynomial regression amounts to the choice of polynomial order, is an important issue. R^2 always increases with increasing k, as we have seen in these two studies. As a result, other measures are sought which balance goodness of fit with complexity of the curve model. Tools often used for model selection in Euclidean polynomial regression, such as Akaike informa-
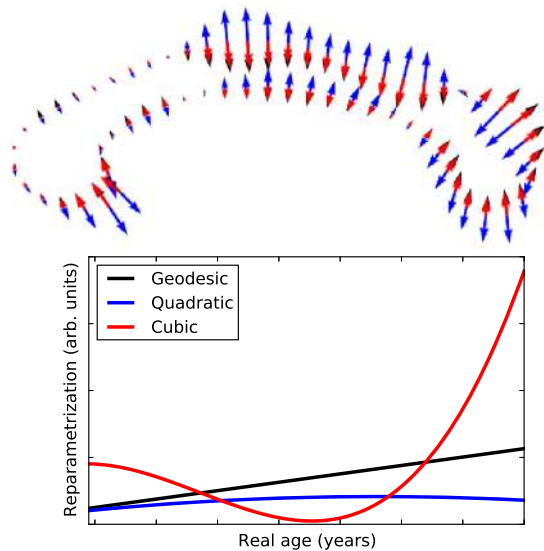
Fig. 4 Parameters for regression of corpus callosa using a cubic polynomial. The velocity (*black*), acceleration (*blue*) and jerk (*red*) are nearly collinear, indicating that the estimated path is essentially a geodesic with cubic time reparametrization. The time reparametrization is shown in the plot, for geodesic, quadratic, and cubic Riemannian polynomial regression

tion criterion and Bayesian information criterion [5] make assumptions about the distribution of data that are difficult to generalize to the manifold setting. Extension of permutation testing for geodesic regression to higher orders would be useful for this task, but such extension is not trivial on a Riemannian manifold. We expect that such an extension of permutation testing is possible in certain cases where it is possible to define "exchangeability" under the null hypothesis that the data follow a given order k trend. Currently, we select models based on qualitative analysis of the fit curves, as in the rat calivaria study, and R^2 values.

4.3 LDDMM Landmark Space

Analysis of landmarks is commonly done in an alternative fashion when scale and rotation invariance is not desired. In this section, we present polynomial regression using the large distance diffeomorphic metric mapping (LDDMM) framework. This framework consists of a Lie group of diffeomorphisms endowed with a right invariant Sobolev metric acting on a space of landmark configurations. For a more detailed description of the group action approach, the reader is encouraged to consult Bruveris et al. [4]. We will instead focus on the Riemannian structure of landmarks and use the formulas for general Riemannian manifolds.

Given m landmarks in d dimensions, let $M \cong \mathbb{R}^{md}$ be the space of all possible configurations. We denote by $x_i \in \mathbb{R}^d$ the location of the ith landmark point. Tangent vectors are also represented as tuples of vectors, $v = (v_i)_{i=1,...,m} \in$

$\mathbb{R}^{md}$, as are cotangent vectors $\alpha = (\alpha_i)_{i=1,...,m} \in \mathbb{R}^{md}$. Contrasting ordinary differential geometric methods in which vectors and metrics are the objects of interest, it is more convenient to work with landmark covectors (which we refer to as momenta). In such case the inverse metric (also called the cometric) is generally written using a shift-invariant scalar kernel $K : \mathbb{R} \to \mathbb{R}$. The inner product of two covectors is given by

$$\langle \alpha, \beta \rangle_{T_x^* M} = \sum_{i,j} K(|x_i - x_j|^2) \alpha_i^T \beta_j. \tag{32}$$

The following Hamilton's equations describe geodesics in landmark space [38, Eq. (21)]:

$$\frac{d}{dt} x_i = \sum_j K(|x_i - x_j|^2) \alpha_j \tag{33}$$

$$\frac{d}{dt} \alpha_i = \sum_j 2(x_i - x_j) K'(|x_i - x_j|^2) \alpha_i^T \alpha_j \tag{34}$$

where K' denotes the derivative of the kernel.

Introducing tangent vectors $v = K\alpha$ and $w = K\beta$, parallel transport in LDDMM landmark space are computed in coordinates using the following formula, derived by Younes et al. [38, Eq. (25)]:

$$\frac{d}{dt} \beta_i = K^{-1} \left(\sum_{j=1}^N (x_i - x_j)^T (w_i - w_j) K'(|x_i - x_j|^2) \alpha_j \right.$$

$$\left. - \sum_{j=1}^N (x_i - x_j)^T (v_i - v_j) K'(|x_i - x_j|^2) \beta_j \right)$$

$$- \sum_{j=1}^N (x_i - x_j) \gamma'(|x_i - x_j|^2)(\alpha_j^T \beta_i + \alpha_i^T \beta_j). \tag{35}$$

In order to integrate the adjoint equations, it is also necessary to compute the Riemannian curvature tensor, which in this case is more complicated. For an in-depth treatment, see Micheli et al. [29, Theorem 2.2].

Using these approaches to computing parallel transport and curvature, we implemented the general polynomial adjoint optimization method. We applied this approach to the rat calivaria data, treating the data as absolute landmark positions (after Procrustes alignment) instead of as scale and rotation invariant Kendall shapes.

Shown in Fig. 5 are the results of LDDMM landmark polynomial regression. Notice that while the geodesic curve in this case corresponds to nonlinear trajectories for the individual landmarks, these paths do not fit the data quite as well as the quadratic curve. In particular, the point at the crown of the skull (labelled point A in Fig. 5) appears to change directions in the quadratic curve, which is not possible using
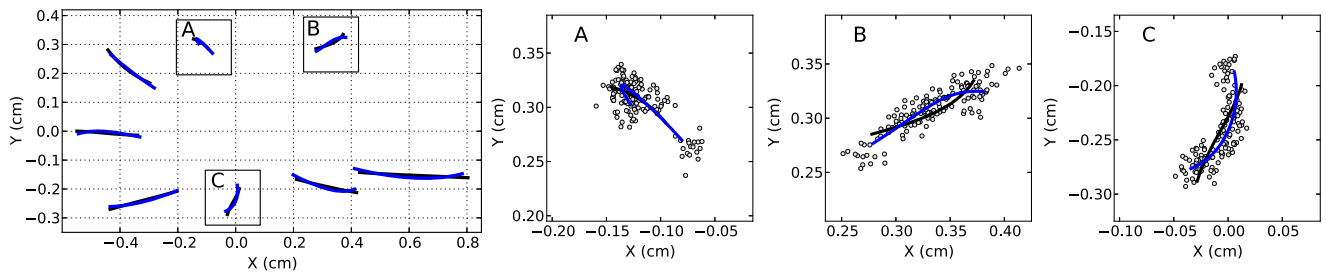
Fig. 5 Regression curves for Bookstein rat data using LDDMM landmark polynomials. The colors of lines indicate order of polynomial used for the regression (*black* = geodesic, *blue* = quadratic). Zoomed views of individual rectangles are shown at right, along with data points in *gray*. The data were aligned with respect to translation and rotation but not scaling, which explains the clear growth trend

a geodesic. These qualitative improvements correspond to a slight increase in R^2, from 0.92 with the geodesic to 0.94 with the quadratic curve.

5 Riemannian Polynomials in Lie Groups

In this section, we consider the case when the configuration manifold is a Lie group G. A tangent vector $v \in T_g G$ at a point $g \in G$ can be identified with a tangent vector at the identity element $e \in G$ via either right or left translation by g^{-1}. The resulting element of $T_e G$ is referred to as the right (respectively, left) trivialization of v. We call a vector field $X \in \mathfrak{X}(G)$ right (respectively, left) invariant if the right trivialization of $X(g)$ is constant for all g. Both left and right translation, considered as mappings $T_g G \to T_e G$ are linear isomorphisms, and we will use the common notation $\mathfrak{g}$ to refer to $T_e G$. The vector space $\mathfrak{g}$, endowed with the vector product given by the right trivialization of the negative Jacobi-Lie bracket of right invariant vector fields is called the Lie algebra of G.

Of particular importance to the study of Lie groups is the adjoint representation, which for each group element g determines a linear action Ad_g on $\mathfrak{g}$ called the adjoint action and its dual action Ad_g^* on $\mathfrak{g}^*$ which is called the coadjoint action of g. In a Riemannian Lie group, the inner product on $\mathfrak{g}$ can be used to compute the adjoint of the adjoint action, which we term the adjoint-transpose action $\mathrm{Ad}_g^\dagger$, defined by

$$\left\langle \mathrm{Ad}_g^\dagger X, Y \right\rangle = \left\langle X, \mathrm{Ad}_g Y \right\rangle \tag{36}$$

for all $X, Y \in \mathfrak{g}$. The infinitesimal version of these actions at the identity element are termed the infinitesimal adjoint action, ad_X, and the infinitesimal adjoint-transpose action, $\mathrm{ad}_X^\dagger$. These operators, along with the metric at the identity, encode all geometric properties such as covariant derivatives and curvature in a Lie group with right invariant Riemannian metric. For a more complete review of Lie groups and the adjoint representation, see [28]. Following [25], we introduce the symmetric product of two vectors $X, Y \in \mathfrak{g}$ as

$$\mathrm{sym}_X Y = \mathrm{sym}_Y X = -\left(\mathrm{ad}_X^\dagger Y + \mathrm{ad}_Y^\dagger X \right). \tag{37}$$

Extending X and Y to right invariant vector fields $\widetilde{X}, \widetilde{Y}$, the covariant derivative $\nabla_{\widetilde{X}} \widetilde{Y}$ is also right invariant (c.f. [7, Proposition 3.18]) and satisfies

$$(\nabla_{\widetilde{X}} \widetilde{Y}) g^{-1} = -\overline{\nabla}_X Y \tag{38}$$

where we have introduced the notation $\overline{\nabla}$ for the reduced Levi-Civita connection:

$$\overline{\nabla}_X Y = \frac{1}{2} \mathrm{ad}_X Y + \frac{1}{2} \mathrm{sym}_X Y. \tag{39}$$

Notice that in this notation, ad represents the skew-symmetric component of the Levi-Civita connection, while sym represents the symmetric component.

We use ξ_1 to denote the right trivialized velocity of the curve $\gamma(t) \in G$. Using our formula for the covariant derivative, one sees that the geodesic equation in a Lie group with right invariant metric is the right "Euler-Poincaré" equation:

$$\frac{d}{dt} \xi_1 = \overline{\nabla}_{\xi_1} \xi_1 = -\mathrm{ad}_{\xi_1}^\dagger \xi_1. \tag{40}$$

The left Euler-Poincaré equation is obtained by removing the negative sign from the right hand side. For polynomials, the Euler-Poincaré equation is generalized to higher order. Introducing $\xi_i, i = 1, \ldots, k$ to represent the right trivialized higher-order velocity vectors v_i,

$$v_i(t) = \xi_i(t) g(t), \tag{41}$$

the reduced Riemannian polynomial equations are

$$\frac{d}{dt} \gamma(t) = \xi_1 \gamma(t) \tag{42}$$

$$\frac{d}{dt} \xi_i(t) = \overline{\nabla}_{\xi_1} \xi_i(t) + \xi_{i+1}(t), \quad i = 1, \ldots, k-1 \tag{43}$$

$$\frac{d}{dt} \xi_k(t) = \overline{\nabla}_{\xi_1} \xi_k(t). \tag{44}$$

Notice that these equations correspond precisely to the polynomial equations (Eq. (15)).

6 Polynomial Regression in Lie Groups

We have seen that the geodesic equation is simplified in a Lie group with right invariant metric, using the Euler-Poincaré equation. In this section, we derive the adjoint equations used to perform geodesic and polynomial regression in a Lie group. Using right-trivialized adjoint variables, we will see that the symmetries provided by the Lie group structure result in adjoint equations more amenable to computation than those in Sect. 4.

6.1 Geodesic Regression

Before moving on to polynomial regression, we first present an adjoint optimization approach to geodesic regression in a Lie group with right invariant metric. Suppose N data points $J_j \in G$ are observed at times $t_j \in [0, 1]$. Using the geodesic distance $d : G \times G \to \mathbb{R}$, the least squares geodesic regression problem is to find the minimum of

$$E(\gamma) = \frac{1}{2} \sum_{j=1}^{N} d\big(\gamma(t_j), J_j\big)^2, \tag{45}$$

subject to the constraint that the curve $\gamma : [0, 1] \to G$ is a geodesic.

In order to determine optimality conditions for γ, consider a variation of the geodesic $\gamma(t)$, which is a vector field along γ that we denote $\delta\gamma(t) \in T_{\gamma(t)}G$. We denote by $Z(t)$ the right trivialization of $\delta\gamma(t)$. The variation of γ induces the following variation in the trivialized velocity ξ_1 [16]:

$$\delta\xi_1(t) = \frac{d}{dt}Z(t) - \mathrm{ad}_{\xi_1} Z(t). \tag{46}$$

Constraining $\delta\gamma$ to be a Jacobi field, we use the following variation of the Euler-Poincaré equation to obtain

$$\frac{d}{dt}\delta\xi_1 = \delta\left(\frac{d}{dt}\xi_1\right) = \delta\big(-\mathrm{ad}_{\xi_1}^{\dagger}\xi_1\big) = \mathrm{sym}_{\xi_1}\delta\xi_1. \tag{47}$$

Combining these results, we write the ordinary differential equation (ODE) that determines, along with initial conditions, the vector field Z:

$$\frac{d}{dt}\begin{pmatrix} Z \\ \delta\xi_1 \end{pmatrix} = \begin{pmatrix} \mathrm{ad}_{\xi_1} & I \\ 0 & \mathrm{sym}_{\xi_1} \end{pmatrix}\begin{pmatrix} Z \\ \delta\xi_1 \end{pmatrix}. \tag{48}$$

This ODE constitutes a general perturbation of a geodesic and the vector field $Z(t)$ is a right trivialized Jacobi field. In order to compute the variations of E with respect to the initial position $\gamma(0)$ and velocity $\xi_1(0)$ of the geodesic $\gamma(t)$, the variations of E with respect to $\gamma(1)$ and $\xi_1(1)$ are transported backward to $t = 0$ by the adjoint ODE. Introducing

adjoint variables $\lambda_0(t), \lambda_1(t) \in \mathfrak{g}$, the left trivialized variation of E with respect to $\gamma(t)$ and the variation with respect to $\xi_1(t)$ are given by

$$\delta_{\gamma(0)}E = -\lambda_0(0) \tag{49}$$

$$\delta_{\xi_1(0)}E = -\lambda_1(0). \tag{50}$$

These variations are computed by initializing $\lambda_0(1) = \lambda_1(1) = 0$ and integrating the adjoint ODE backward to $t = 0$. The adjoint ODE is obtained by simply computing the adjoint of the ODE governing geodesic perturbations, Eq. (48), with respect to the $L^2([0, 1] \to \mathfrak{g})$ inner product. The resulting adjoint ODE is

$$\frac{d}{dt}\begin{pmatrix} \lambda_0 \\ \lambda_1 \end{pmatrix} = \begin{pmatrix} -\mathrm{ad}_{\xi_1}^{\dagger} & 0 \\ -I & -\mathrm{sym}_{\xi_1}^{\dagger} \end{pmatrix}\begin{pmatrix} \lambda_0 \\ \lambda_1 \end{pmatrix}, \tag{51}$$

where the adjoint of the symmetric product is given by

$$\mathrm{sym}_X^{\dagger} Y = -\mathrm{ad}_X Y + \mathrm{ad}_Y^{\dagger} X. \tag{52}$$

The adjoint variable λ_0 takes jump discontinuities when passing over data points:

$$\lambda_0\big(t_j^-\big) - \lambda_0\big(t_j^+\big) = (\mathrm{Log}_{\gamma(t_j)} J_j)\gamma(t_j)^{-1}. \tag{53}$$

The jumps represent the residual vectors, obtained by right trivialization of the Riemannian log map from the predicted point $\gamma(t_j)$ to the data J_j. Notice that the adjoint variable λ satisfies an equation resembling the Euler-Poincaré equation and can likewise be solved in closed form:

$$\lambda_0(t) = \sum_{j, t_j > t} \mathrm{Ad}_{\gamma^{-1}(t)\gamma(t_j)}^{\dagger} \mathrm{Log}_{\gamma(t_j)} J_j. \tag{54}$$

This is particularly useful because it reduces the second order ODE, Eq. (51), to an ODE of first order, since the first equation is solved in closed form. We will soon see that this simplification occurs even when using higher order polynomials.

Finally, minimization of E is performed using the variations $\delta_{\gamma(0)}E$, $\delta_{\xi_1(0)}E$ using, for example the following gradient descent steps:

$$\gamma(0)^{k+1} = \mathrm{Exp}(-\alpha\delta_{\gamma(0)^k}E)\gamma(0)^k \tag{55}$$

$$\xi_1(0)^{k+1} = \xi_1(0)^k - \alpha\delta_{\xi(0)^k}E \tag{56}$$

for some positive step size α, where k denotes the step of the iterative optimization process. Note that commonly the Riemannian exponential map Exp in the above expression is replaced by a numerically efficient approximation such as the Cayley map [3].

6.2 Example: Rotation Group SO(3)

As an example, in this section we derive the algorithm for polynomial regression in the group of rotations in three dimensions, SO(3). This group consists of orthogonal matrices with determinant one, and has associated the Lie algebra $\mathfrak{so}(3)$ of skew-symmetric 3-by-3 matrices. Skew-symmetric matrices can be bijectively identified with vectors in $\mathbb{R}^3$ using the following mapping $*$:

$$* : \mathbb{R}^3 \leftrightarrow \mathfrak{so}(3), \qquad * \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}. \quad (57)$$

We use a star to indicate both this mapping $\mathbb{R}^3 \to \mathfrak{so}(3)$ and its inverse, a notation which emphasizes that it is the Hodge dual in $\mathbb{R}^3$, though it is also commonly written using a hat symbol [28]. Using the cross product on $\mathbb{R}^3$, the star map is also a Lie algebra isomorphism, so that

$$* \operatorname{ad}_{*x} * y = x \times y. \quad (58)$$

The adjoint action under the star is also quite convenient, as it is given simply by matrix-vector multiplication:

$$\operatorname{Ad}_g(*x) = *(gx) \quad (59)$$

for any $g \in SO(3), x \in \mathbb{R}^3$.

We will use a left invariant metric given by a symmetric positive definite 3-by-3 matrix A. For vectors $x, y \in \mathbb{R}^3$, the inner product is

$$\langle *x, *y \rangle_{\mathfrak{g}} = x^T A y. \quad (60)$$

With this inner product, the infinitesimal adjoint transpose action is

$$* \operatorname{ad}_{*x}^\dagger * y = -A^{-1}(x \times Ay). \quad (61)$$

The most natural metric is that in which A is the identity matrix. In that case, left invariance also implies right invariance and skew-symmetry of $\operatorname{ad}^\dagger$, so that for any $X, Y \in \mathfrak{so}(3)$:

$$\operatorname{sym}_X Y 0, \qquad \overline{\nabla}_X Y = \frac{1}{2} \operatorname{ad}_X Y. \quad (62)$$

The Euler-Poincaré equation in the biinvariant case is

$$\frac{d}{dt} \xi = \operatorname{ad}_\xi^\dagger \xi = - * \xi \times * \xi = 0, \quad (63)$$

implying that geodesics using the biinvariant metric have constant trivialized velocity. The geodesic can then be integrated in closed form:

$$\frac{d}{dt} \gamma(t) = \xi \gamma(t) \quad \Longrightarrow \quad \gamma(t) = \exp(t\xi). \quad (64)$$

Notice that the adjoint-transpose action of a rotation matrix $g \in SO(3)$ on a 3-vector x is given by

$$* \operatorname{Ad}_g^\dagger(*x) = g^T x. \quad (65)$$

So the first adjoint equation is given by

$$\lambda_0(t) = \gamma(t)^T \gamma(1) \lambda_0(1) \quad (66)$$

$$= \exp(-t\xi) \exp(\xi) \lambda_0(1) \quad (67)$$

$$= \exp\big((1-t)\xi\big) \lambda_0(1) \quad (68)$$

$$= \lambda_0(1) \cos\big((1-t)\|\xi\|\big)$$

$$\quad - \frac{1}{\|\xi\|}\big(*\xi \times \lambda_0(1)\big) \sin\big((1-t)\|\xi\|\big)$$

$$\quad + \frac{1}{\|\xi\|^2} * \xi \big(*\xi \cdot \lambda_0(1)\big)\big(1 - \cos\big((1-t)\|\xi\|\big)\big). \quad (69)$$

where the last line is Rodrigues' rotation formula. The second adjoint equation, which determines the variation used to update the velocity, is obtained by integrating this. For geodesic regression with biinvariant metric, a closed form solution is available for the second adjoint variable as well:

$$\frac{d}{dt} \lambda_1(t) = -\lambda_0(t) \quad (70)$$

$$\lambda_1(t) = \int_t^1 \lambda_0(s) \, ds \quad (71)$$

$$= \lambda_0(1) \frac{1}{\|\xi\|} \sin\big((1-t)\|\xi\|\big) \quad (72)$$

$$\quad - \frac{1}{\|\xi\|^2}\big(*\xi \times \lambda_0(1)\big)\big(1 - \cos\big((1-t)\|\xi\|\big)\big)$$

$$\quad + \frac{1}{\|\xi\|^3} * \xi \big(*\xi \cdot \lambda_0(1)\big)\big(1 - t - \sin\big((1-t)\|\xi\|\big)\big).$$

6.3 Polynomial Regression

We apply a method similar to that of the previous section to derive an adjoint optimization scheme for Riemannian polynomial regression in a Lie group with right invariant metric. A variation of the first equation gives Eq. (46). Taking variations of the other equations, noting that $\overline{\nabla}$ is linear in each argument, we have

$$\frac{d}{dt} \delta\xi_i = \overline{\nabla}_{\delta\xi_1} \xi_i + \overline{\nabla}_{\xi_1} \delta\xi_i + \delta\xi_{i+1}. \quad (73)$$

Along with Eq. (46), these provide the essential equations for a polynomial perturbation Z of γ, which can be considered a kind of higher-order Jacobi field. Introducing adjoint variables $\lambda_0, \ldots, \lambda_k \in \mathfrak{g}$, the adjoint system is (see Appendix C for derivation)

$$\frac{d}{dt} \lambda_0 = - \operatorname{ad}_{\xi_1}^\dagger \lambda_0 \quad (74)$$

$$\frac{d}{dt}\lambda_1 = -\lambda_0 - \text{sym}^\dagger_{\xi_1}\lambda_1 + \sum_{i=2}^{k}(-\overline{\nabla}_{\xi_i} - \text{sym}^\dagger_{\xi_i})\lambda_i \qquad (75)$$

$$\frac{d}{dt}\lambda_i = -\lambda_{i-1} + \overline{\nabla}_{\xi_1}\lambda_i, \quad i = 2, \ldots, k, \qquad (76)$$

or, using only ad and ad†, as

$$\frac{d}{dt}\lambda_0 = -\text{ad}^\dagger_{\xi_1}\lambda_0 \qquad (77)$$

$$\frac{d}{dt}\lambda_1 = -\lambda_0 + \text{ad}_{\xi_1}\lambda_1 - \text{ad}^\dagger_{\lambda_1}\xi_1$$
$$+ \frac{1}{2}\sum_{i=2}^{k}\big(\text{ad}_{\xi_i}\lambda_i + \text{ad}^\dagger_{\xi_i}\lambda_i - \text{ad}^\dagger_{\lambda_i}\xi_i\big) \qquad (78)$$

$$\frac{d}{dt}\lambda_i = -\lambda_{i-1} + \frac{1}{2}\big(\text{ad}_{\xi_1}\lambda_i - \text{ad}^\dagger_{\xi_1}\lambda_i - \text{ad}^\dagger_{\lambda_i}\xi_1\big). \qquad (79)$$

For $i = 2, \ldots, k$, these equations resemble the original polynomial equations. However, the evolution of λ_1 is influenced by all adjoint variables and higher-order velocities in a non-trivial way. The first adjoint equation again resembles the Euler-Poincaré equation, and its solution is given by Eq. (54).

6.3.1 Polynomial Regression in SO(3)

Revisiting the rotation group, we can extend the geodesic regression results to polynomials. Representing Lie algebra elements as 3-vectors ξ_i, the equations for higher order polynomials in SO(3) are

$$\frac{d}{dt}\gamma(t) = \big(*\xi_1(t)\big)\gamma(t) \qquad (80)$$

$$\frac{d}{dt}\xi_1(t) = \xi_2(t) \qquad (81)$$

$$\frac{d}{dt}\xi_i(t) = \frac{1}{2}\xi_1(t) \times \xi_i(t) + \xi_{i+1}(t), \quad i = 2, \ldots, k-1 \qquad (82)$$

$$\frac{d}{dt}\xi_k(t) = \frac{1}{2}\xi_1(t) \times \xi_k(t). \qquad (83)$$

In this case, closed form integration isn't available, even with a biinvariant metric. Even for higher order polynomials, the first adjoint equation is integrated in closed form, giving

$$\lambda_0(t) = \gamma(t)^T\gamma(1)\lambda_0(1). \qquad (84)$$

7 Lie Group Actions

So far, we've seen that polynomial regression is particularly convenient in Lie groups with right invariant metrics, reducing the adjoint system from second to first order using the

closed form integral of λ_0. We now consider the case when a Lie group G acts on another manifold M which is itself equipped with a Riemannian metric. For our purposes, the group action need not be transitive, in which case the target space is called a "homogeneous space" for G.

Although the two approaches sometimes coincide, generally one must choose between using polynomials defined by the metric in M, ignoring the action of G, or using curves defined by the action of polynomials in G on points in M. In cases when a Riemannian Lie group is known to act on the space M, the primary object of interest is usually not the path in the object space M, but the path of symmetries described by the group elements. Therefore it is most natural to use the Lie group structure to define paths in object space. We employ this approach, in which polynomial regression under a Riemannian Lie group action is studied primarily using the Lie group elements.

Following this plan, we model a polynomial in M as a curve $p(t)$ defined using the group action:

$$p(t) = \gamma(t).p_0 \qquad (85)$$

where γ is a polynomial of order k in G with parameters

$$\gamma(0) \in G, \qquad \xi_1, \ldots, \xi_k \in \mathfrak{g} \qquad (86)$$

and $p_0 \in M$ is a base point in the object space. Invariance of the metric on G allows us to assume, without loss of flexibility in the model, that the base deformation is the identity: $\gamma(0) = e \in G$. Optimization is done by fixing $\gamma(0) = e \in G$ and minimizing a least squares objective function defined using the metric on M, with respect to the base point $p_0 \in M$ and the parameters of the Lie group polynomial, $\xi_1, \ldots, \xi_k \in \mathfrak{g}$. This is accomplished using a similar adjoint method to that presented in the previous sections, but where the jump discontinuities in λ_0 are modified due to this change in objective function. In the following sections, we discuss this in more detail and also derive the gradients with respect to the base point p_0.

7.1 Action on a General Manifold

A smooth group action can be differentiated to obtain a mapping from the Lie algebra $\mathfrak{g}$ to the tangent space T_pM at any point $p \in M$. Given a curve $g(t) : (-\epsilon, \epsilon) \to G$ such that $g(0) = e$ and $\frac{d}{dt}|_{t=0}g(t) = \xi \in \mathfrak{g}$, define the following mapping (c.f. [16]):

$$\rho_p(\xi) := \frac{d}{dt}\bigg|_{t=0}g(t).p. \qquad (87)$$

The function ρ_p is a linear mapping from $\mathfrak{g}$ to T_pM, and as such it has a dual $\rho_p^* : T_p^*M \to \mathfrak{g}^*$ that maps cotangent vectors in M to the Lie coalgebra $\mathfrak{g}^*$. This dual mapping

we refer to as the cotangent lift momentum map and use the notation $\mathbf{J} : T^*M \to \mathfrak{g}^*$.

The most important property of $\mathbf{J}$ is that it is preserved under the coadjoint action:

$$\mathrm{Ad}^*_g \mathbf{J}m = \mathbf{J}g.m \quad \forall m \in T^*M. \tag{88}$$

The action of g on the cotangent bundle, which appears on the right-hand side above, maps a cotangent vector μ at point p to the vector $g.\mu \in T^*_{g.p}M$. Replacing squared norm with squared geodesic distance on the Riemannian manifold M, the first adjoint variable is then given by

$$\lambda_0(t) = \sum_{j, t_j > t} \mathbf{J}\gamma(t)\gamma(t_j)^{-1}.(\mathrm{Log}_{\gamma(t_j).p_0} J_j)^\flat. \tag{89}$$

Of particular interest is the case when the metric on G and the metric on the manifold M coincide, in the sense that for any vectors $\xi, \mu \in \mathfrak{g}$ and points $p \in M$:

$$\langle \xi, \mu \rangle_\mathfrak{g} = \langle \xi.p, \mu.p \rangle_{T_pM}. \tag{90}$$

Fixing a base point $p_0 \in M$, this means the mapping $g \to g.p_0$ is a Riemannian submersion. If, additionally, the metric on G is biinvariant, this implies that the covariant derivative satisfies [33]

$$\nabla_{\xi.p}\mu.p = (\overline{\nabla}_\xi \mu).p \tag{91}$$

so that geodesics and polynomials in M are generated by polynomials in G along with the action on the base point p_0.

7.1.1 Example: Rotations of the Sphere

Consider the sphere of radius one in $\mathbb{R}^3$, which is denoted $\mathbb{S}^2$. The group $\mathrm{SO}(3)$ acts naturally on the sphere. For this example, we will use the biinvariant metric on $\mathrm{SO}(3)$, which corresponds to using the identity for the A matrix in Sect. 6.2. Representing points on the sphere as unit vectors in $\mathbb{R}^3$, the group action is simply left multiplication by a matrix in $\mathrm{SO}(3)$:

$$\gamma.p = \gamma p \tag{92}$$

$$\xi.p = \xi p \tag{93}$$

for all $\gamma \in \mathrm{SO}(3), \xi \in \mathfrak{so}(3), p \in \mathbb{S}^2, v \in T_p\mathbb{S}^2$. The infinitesimal action is in fact a cross product, which is easily seen using the star map:

$$\xi.p = \xi p = (*\xi) \times p. \tag{94}$$

Representing elements in $\mathfrak{so}(3)^*$ as 3-vectors, we derive the cotangent lift momentum map as well; letting $a \in T_p\mathbb{S}^2$,

$$\mathbf{J}a = *(p \times a). \tag{95}$$

This can be interpreted as converting a linear momentum on the surface of the sphere into an angular momentum in $\mathfrak{so}(3)$ using the cross product with the moment arm p. The standard metric on the sphere corresponds to the standard biinvariant metric on $\mathrm{SO}(3)$ so that, as discussed previously, polynomials on $\mathbb{S}^2$ correspond to polynomials in $\mathrm{SO}(3)$ acting on points on the sphere.

The polynomial equations for the sphere are precisely those for $\mathrm{SO}(3)$, along with the action of $\gamma(t)$ on the base point $p_0 \in \mathbb{S}^2$. The derivative of $\gamma(t)$ is replaced by the equation

$$\frac{d}{dt}p(t) = \frac{d}{dt}(\gamma(t).p_0) = \xi_1(t).p(t). \tag{96}$$

The evolution of ξ_i is the same as that for $\mathrm{SO}(3)$. Figure 6 shows example polynomial curves in the rotation group and their action on a point on the sphere. Notice that the example polynomials on the sphere are precisely those shown in Fig. 1, although they were generated here using polynomials on $\mathrm{SO}(3)$ instead of integrating directly on the sphere.

In order to integrate the adjoint equations, the jump discontinuities must be computed using the log map on the sphere:

$$\mathrm{Log}_x y = \theta \left(\frac{y - \cos\theta x}{\sin\theta} \right), \quad \cos\theta = x^T y. \tag{97}$$

The flatting operation acts trivially on this vector, and the action of $\mathrm{SO}(3)$ on covectors corresponds to matrix-vector multiplication. Using this, along with the momentum map $\mathbf{J}$, we have the jump discontinuities for the first adjoint variable λ_0:

$$\lambda_0(t_j^-) - \lambda_0(t_j^+) = \gamma(t_j) \times (\mathrm{Log}_{\gamma(t_j)} J_j). \tag{98}$$

The higher adjoint variables satisfy the same ODEs as in Sect. 6.3.1.

7.2 Lie Group Actions on Vector Spaces

We will assume in this section that the manifold is a vector space V and that G acts linearly on the left on V. Given a smooth linear group action, a vector ξ in the Lie algebra $\mathfrak{g}$ acts linearly on a vector $v \in V$ in the following way

$$\xi.v = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} g(\epsilon).v \tag{99}$$

where $g(\epsilon)$ is a curve in G satisfying $g(0) = e$ and $\frac{d}{d\epsilon}|_{\epsilon=0}g(\epsilon) = \xi$. Again we use the notation $\rho_v : \mathfrak{g} \to V$ to denote right-multiplication under this action:

$$\rho_v \xi := \xi.v \quad \forall v \in V, \xi \in \mathfrak{g}. \tag{100}$$

Geodesic ($k = 1$)   Quadratic ($k = 2$)   Cubic ($k = 3$)
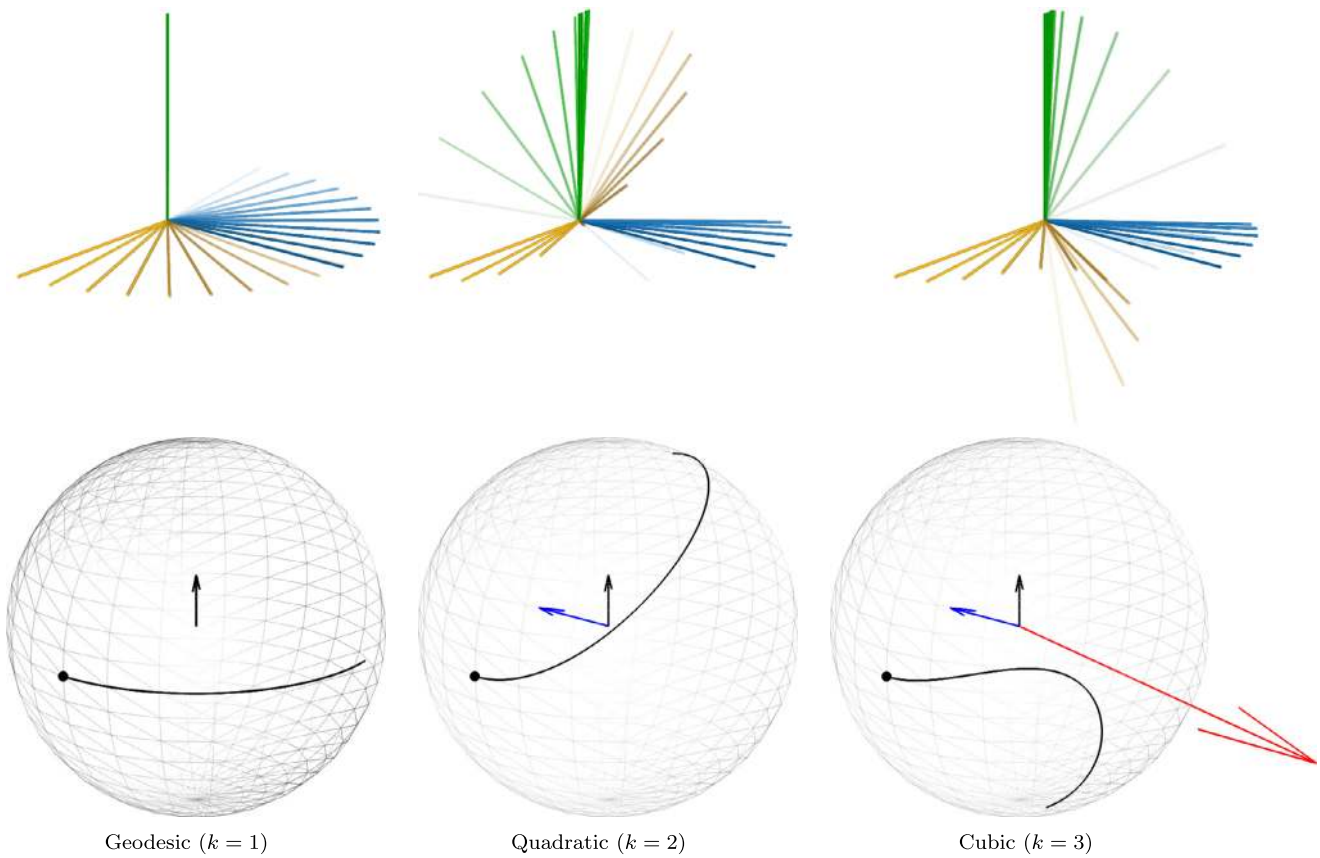
Fig. 6 Sample polynomial curves in SO(3) and their action on a base point $p_0 \in \mathbb{S}^2$ (*black dot*) on the sphere. In the *top row*, the rotating coordinate axes are shown for three polynomials. In the *bottom row*, the *arrows* show the vectors $\xi_1(0)$ (*black*), $\xi_2(0)$ (*blue*), and $\xi_3(0)$ (*red*), representing initial angular velocity, acceleration, and jerk. The action on the base point, $p(t) = \gamma(t).p_0 \in \mathbb{S}^2$, is represented as a *black curve* on the sphere. A geodesic corresponds to constant angular velocity, while the non-zero acceleration and jerk in the quadratic and cubic curves tilt the rotation axis

In the vector space setting, the cotangent lift momentum map (again defined as the dual of ρ_v), is written using the diamond notation introduced in [16]:

$$v \diamond a \in \mathfrak{g}^* \quad \forall v \in V, a \in V^*, \tag{101}$$

$$(v \diamond a, \xi)_{(\mathfrak{g}*,\mathfrak{g})} := (a, \rho_v \xi)_{(V^*,V)} \quad \forall \xi \in \mathfrak{g}. \tag{102}$$

The diamond map interacts with the coadjoint action Ad* in a convenient way:

$$\mathrm{Ad}^*_{g^{-1}}(v \diamond a) = (g.v) \diamond (g.a). \tag{103}$$

This relation is fundamental in that it shows that the diamond map is preserved under the coadjoint action. This is quite useful in our case, as we will soon see that diamond maps show up commonly in variational problems on inner product spaces.

Commonly, data is provided in the form of points J_i in the vector space V. In that case, the inner product on V is

used to write the regression problem as a minimization of

$$E(\gamma, v_0) = \frac{1}{2} \sum_{j=1}^{N} \left\| \gamma(t_j).v_0 - J_j \right\|_V^2, \tag{104}$$

subject to the constraint that γ is a polynomial in G and $v_0 \in V$ is an evolving template vector. Without loss of generality, $\gamma(0)$ can also be constrained to be the identity so that v_0 is the template vector at time zero. Optimization of Eq. (104) with respect to v_0 requires the variation

$$\delta_{v_0} E = \sum_{j=1}^{N} \gamma(t_j)^{-1} \left(\gamma(t_j).v_0 - J_j \right)^\flat. \tag{105}$$

Here the musical flat symbol $\flat$ denotes lowering of indices using the metric on V, an operation mapping V to V^*. If the group G acts by isometries on V, then the group action commutes with flatting and the optimal base vector v_0 can

be computed in closed form

$$\hat{v}_0 = \frac{1}{N} \sum_{j=1}^{N} \gamma(t_j)^{-1}.J_j. \tag{106}$$

Even when G does not act by isometries, the optimal base vector can often be solved for in closed form.

The variation with respect to $\gamma(t_j)$ is more interesting:

$$\delta_{\gamma(t_j)} E = \big(\gamma(t_j).v_0\big) \diamond \big(\gamma(t_j).v_0 - J_j\big)^{\flat}. \tag{107}$$

Using this along with the relation between the coadjoint action and diamond map, we can write the first polynomial adjoint variable in closed form

$$\lambda_0(t) = \sum_{j,\,t_j > t} \big(\gamma(t).v_0\big) \diamond \big(\gamma(t)\gamma(t_j)^{-1}.\big(\gamma(t_j).v_0 - J_j\big)^{\flat}\big). \tag{108}$$

7.2.1 Example: Diffeomorphically Deforming Images

Right invariant Sobolev metrics on groups of diffeomorphisms are the main objects of study in computational anatomy [30]. Describing an image I as a square integrable function of a domain $\Omega \subset \mathbb{R}^d$, the left action of a diffeomorphism $\gamma \in \mathrm{Diff}(\Omega)$ is

$$\gamma.I = I \circ \gamma^{-1}. \tag{109}$$

The corresponding infinitesimal action of a velocity field ξ on an image is

$$\xi.I = -\xi^T \nabla I \tag{110}$$

and the diamond map is

$$(I \diamond \alpha)(y) = -\alpha(y)\nabla I(y). \tag{111}$$

Geodesic regression in this context, using an adjoint optimization method, has been previously studied [31]. Using their method, the initial momentum of a geodesic is constrained by horizontal: that is, $L\xi_1(0) = I_0 \diamond \alpha(0)$. As a result, changes in base image I_0 influence the behavior of the deformation itself.

Using our method, the base velocity vectors ξ_i are not constrained to be horizontal. Implementation of polynomial regression involves the expression above for the diamond map, along with the ad and ad* operators [28]

$$\mathrm{ad}_\xi X = D\xi X - DX\xi, \tag{112}$$

$$\mathrm{ad}_\xi^* m = Dm\xi + m\,\mathrm{div}\,\xi + (D\xi)^T m. \tag{113}$$

Inserting this into the right Euler-Poincaré equation yields the well-known EPDiff equation for geodesic evolution in the diffeomorphism group [16]:

$$\frac{d}{dt}m = -Dm\xi - m\,\mathrm{div}\,\xi - (D\xi)^T m. \tag{114}$$

For polynomials, momenta $m_i = L\xi_i$ are introduced and this EPDiff equation is generalized to

$$\frac{d}{dt}m_1 = -Dm_1\xi_1 - m_1\,\mathrm{div}\,\xi_1 - (D\xi_1)^T m_1 + m_2 \tag{115}$$

$$\begin{aligned}
\frac{d}{dt}m_i = m_{i+1} + \frac{1}{2}\big(L(D\xi_1\xi_i - D\xi_i\xi_1) \\
- Dm_i\xi_1 - (D\xi_1)^T m_i - m_i\,\mathrm{div}\,\xi_1 \\
- Dm_1\xi_i - (D\xi_i)^T m_1 - m_1\,\mathrm{div}\,\xi_i\big)
\end{aligned} \tag{116}$$

$$\begin{aligned}
\frac{d}{dt}m_k = \frac{1}{2}\big(L(D\xi_1\xi_i - D\xi_i\xi_1) \\
- Dm_k\xi_1 - (D\xi_1)^T m_k - m_k\,\mathrm{div}\,\xi_1 \\
- Dm_1\xi_k - (D\xi_i)^T m_1 - m_1\,\mathrm{div}\,\xi_k\big)
\end{aligned} \tag{117}$$

The estimation of the base image I_0 is simplified, as Eq. (105) is solved in closed form using

$$I_0(y) = \frac{\sum_j |D\gamma_j(y)| J_j \circ \gamma_j(y)}{\sum_j |D\gamma_j(y)|}. \tag{118}$$

As an example of image regression, synthetic data were generated and geodesic regression was performed using the adjoint method described above. Figure 7 shows the input images, as well as the estimated geodesic trend, which matches the input data well. Note that although the method presented in [31] is similar, using our abstraction, geodesic regression can be generalized to polynomials of any order, and to data which are not necessarily scalar-valued images.

8 Discussion

The Riemannian polynomial framework we have presented provides a general approach to regression for manifold-valued data. The greatest limitation to performing polynomial regression on a general Riemannian manifold is that it requires computation of the Riemannian curvature tensor, which is often tedious [29]. In a Lie group or homogeneous space, we have shown that the symmetries provided by the group allow for not only simple integration using parallel transport in the Lie algebra, but also simplified adjoint equations that do not require explicit curvature computation.

The theory of rolling maps on the sphere, introduced by Jupp & Kent [18], offer another perspective on Riemannian polynomials. On the sphere, this interesting interpretation is

Fig. 7 Image regression
example. Three synthetic
images where generated (*top
row*) at times 0,0.5,1. Geodesic
regression was performed,
resulting in the images shown in
the *second row*, corresponding
to the deformations in the *last
row*
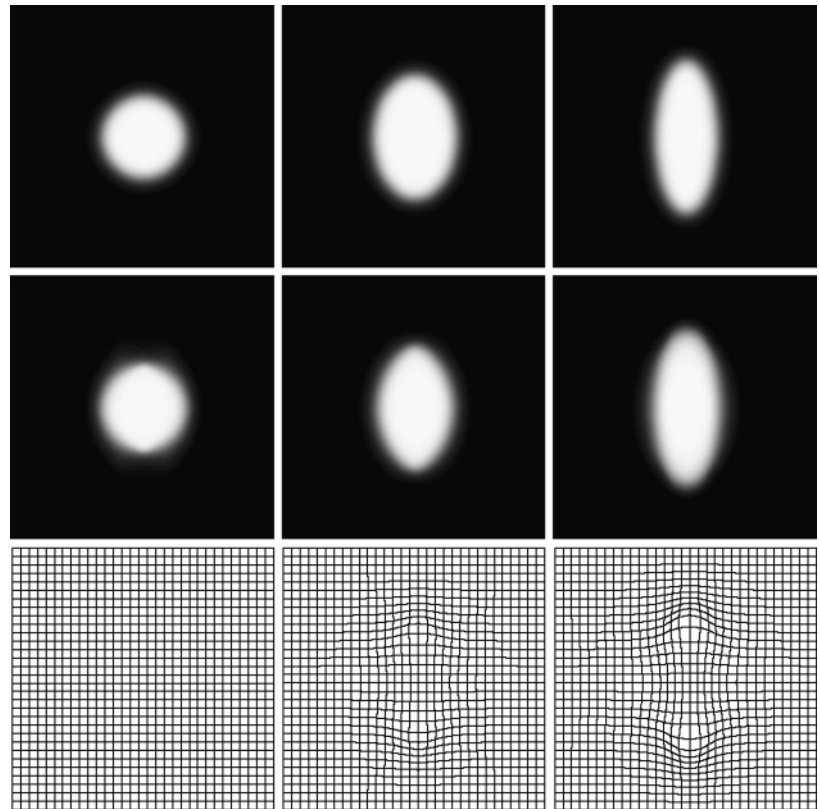
related to the group action described above. Given a curve
$\gamma : [0, 1] \to \mathbb{S}^2$, consider embedding both the sphere and a
plane in $\mathbb{R}^3$ such that the plane is tangent to the sphere at
the point $\gamma(0)$. Now roll the sphere along so that it remains
tangent at $\gamma(t)$ at every time, and such that no slipping or
twisting occurs. The resulting path, $\gamma_u : [0, 1] \to \mathbb{R}^2$, traced
out on the plane is called the unwrapped curve. Remarkably,
the property that γ is a k-order polynomial on $\mathbb{S}^2$ is equiva-
lent to the unwrapped curve γ_u being a k-order polynomial
in the conventional sense. For more information regarding
this connection to Jupp & Kent's rolling maps, as well as a
comparison to Noakes' cubic splines [32], the reader is re-
ferred to the literature of Leite & Krakowski [24].

Appendix A: Numerical Integration of the Polynomial Equations

By definition, in the limit $\Delta t \to 0$, the exponential map sat-
isfies $\dot{\gamma}(t) = v_1(t)$. To see that the forward integration algo-
rithm shown in Algorithm 1 approximates the polynomial
equations, let $w(t)$ be any vector field *parallel* along $\gamma(t)$.

That is,

$$\nabla_{\dot{\gamma}(t)} w(t) = 0. \tag{119}$$

Denote by $P_{\Delta t}(t) = \text{ParTrans}(p, \Delta t v, w)$ the parallel trans-
port of a vector $w \in T_p M$ along a geodesic from point p for
time Δt in the direction of vector $v \in T_p M$. Then

$$\frac{d}{dt}\langle w, v_i \rangle = \langle \nabla_{\dot{\gamma}} w, v_i \rangle + \langle w, \nabla_{\dot{\gamma}} v_i \rangle = \langle w, \nabla_{\dot{\gamma}} v_i \rangle \tag{120}$$

Now consider approximation of this inner product derivative
under our integration scheme:

$$\frac{d}{dt}\langle w, v_i \rangle \approx \lim_{\Delta t \to 0} \frac{1}{\Delta t}\Big(\big\langle P_{\Delta t} w(t), P_{\Delta t}\big(v_i(t) + \Delta t v_{i+1}(t)\big)\big\rangle$$
$$- \big\langle w(t), v_i(t)\big\rangle\Big). \tag{121}$$

The parallel transport operator is linear in the vectors being
transported, so

$$\frac{d}{dt}\langle w, v_i \rangle \approx \lim_{\Delta t \to 0} \frac{1}{\Delta t}\Big(\big\langle P_{\Delta t} w(t), P_{\Delta t} v_i(t)\big\rangle$$
$$+ \Delta t \big\langle P_{\Delta t} w(t), v_{i+1}(t)\big\rangle - \big\langle w(t), v_i(t)\big\rangle\Big)$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\Big(\big(\big\langle P_{\Delta t} w(t), P_{\Delta t} v_i(t)\big\rangle - \big\langle w(t), v_i(t)\big\rangle\big)$$
$$+ \lim_{\Delta t \to 0}\big\langle P_{\Delta t} w(t), v_{i+1}(t)\big\rangle\Big) \tag{122}$$

The first line is zero, by definition of parallel transport. Also note that $\lim_{\Delta t \to 0} P_{\Delta t} w = w$, so that

$$\frac{d}{dt} \langle w, v_i \rangle = \langle w, \nabla_{\dot\gamma} v_i \rangle \approx \langle w, v_{i+1} \rangle. \tag{123}$$

As this holds for any parallel vector field w, this implies that our integration algorithm approximates the polynomial equation

$$\nabla_{\dot\gamma} v_i = v_{i+1}. \tag{124}$$

Appendix B: Derivation of Adjoint Equations in Riemannian Manifolds

In this appendix we derive the adjoint system for the polynomial regression problem. The approach to calculus of variations on Riemannian manifolds described here is very similar to that employed by Noakes et al. [32]. Consider a simplified objective function containing only a single data term, at time T:

$$E(\gamma, \{v_i\}, \{\lambda_i\}) = d(\gamma(T), y)^2 + \int_0^T \langle \lambda_0, \dot\gamma - v_1 \rangle dt$$
$$+ \sum_{i=1}^{k-1} \int_0^T \langle \lambda_i, \nabla_{\dot\gamma} v_i - v_{i+1} \rangle dt$$
$$+ \int_0^T \langle \lambda_k, \nabla_{\dot\gamma} v_k \rangle dt. \tag{125}$$

Now consider taking variations of E with respect to the vector fields v_i. For each i there are only two terms containing v_i, so if W is a test vector field along γ, then the variation of E with respect to v_i in the direction W satisfies

$$\int_0^T \langle \delta_{v_i} E, W \rangle dt = \int_0^T \langle \lambda_i, \nabla_{\dot\gamma} W \rangle dt - \int_0^T \langle \lambda_{i-1}, W \rangle dt. \tag{126}$$

The first term is integrated by parts to yield

$$\int_0^T \langle \delta_{v_i} E, W \rangle dt = \langle \lambda_i, W \rangle \big|_0^T - \int_0^T \langle \nabla_{\dot\gamma} \lambda_i, W \rangle dt$$
$$- \int_0^T \langle \lambda_{i-1}, W \rangle dt. \tag{127}$$

The variation with respect to v_i for $i = 1, \ldots, k$ is then given by

$$\delta_{v_i(t)} E = 0 = -\nabla_{\dot\gamma} \lambda_i - \lambda_{i-1}, \quad t \in (0, T) \tag{128}$$
$$\delta_{v_i(T)} E = 0 = \lambda_i(T) \tag{129}$$
$$\delta_{v_i(0)} E = -\lambda_i(t). \tag{130}$$

In order to determine the differential equation for λ_0, the variation with respect to γ must be computed. Let W again

denote a test vector field along γ. For some $\epsilon > 0$, let $\{\gamma_s : s \in (-\epsilon, \epsilon)\}$ be a differentiable family of curves satisfying

$$\gamma_0 = \gamma \tag{131}$$
$$\frac{d}{ds} \gamma_s \Big|_{s=0} = W. \tag{132}$$

If ϵ is chosen small enough, the vector field W can be extended to a neighborhood of γ such that $[W, \dot\gamma_s] = 0$, where a dot indicates the derivative in the $\frac{\partial}{\partial t}$ direction. The vanishing Lie bracket implies the following identities

$$\nabla_W \dot\gamma_s = \nabla_{\dot\gamma_s} W \tag{133}$$
$$\nabla_W \nabla_{\dot\gamma_s} = \nabla_{\dot\gamma_s} \nabla_W + R(W, \dot\gamma_s). \tag{134}$$

Finally, the vector fields v_i, λ_i are extended along γ_s via parallel translation, so that

$$\nabla_W v_i = 0 \tag{135}$$
$$\nabla_W \lambda_i = 0. \tag{136}$$

The variation of E with respect to γ satisfies

$$\int_0^T \langle \delta_\gamma E, W \rangle dt = \frac{d}{ds} E(\gamma_s, \{v_i\}, \{\lambda_i\}) \big|_{s=0}$$
$$= -\langle \mathrm{Log}_{\gamma(T)} y, W(T) \rangle$$
$$+ \frac{d}{ds} \int_0^T \langle \lambda_0, \dot\gamma_s - v_1 \rangle dt \Big|_{s=0}$$
$$+ \frac{d}{ds} \sum_{i=1}^{k-1} \int_0^T \langle \lambda_i, \nabla_{\dot\gamma_s} v_i - v_{i+1} \rangle dt \Big|_{s=0}$$
$$+ \frac{d}{ds} \int_0^T \langle \lambda_k, \nabla_{\dot\gamma_s} v_k \rangle dt \Big|_{s=0}. \tag{137}$$

As the λ_i are extended via parallel translation, their inner products satisfy

$$\frac{d}{ds} \langle \lambda_i, U \rangle |_{s=0} = \langle \nabla_W \lambda_i, U \rangle + \langle \lambda_i, \nabla_W U \rangle = \langle \lambda_i, \nabla_W U \rangle. \tag{138}$$

Then applying this to each term in the previous equation,

$$\int_0^T \langle \delta_\gamma E, W \rangle dt = -\langle \mathrm{Log}_{\gamma(T)} y, W(T) \rangle$$
$$+ \int_0^T \langle \lambda_0, \nabla_W \dot\gamma - \nabla_W v_1 \rangle dt$$
$$+ \sum_{i=1}^{k-1} \int_0^T \langle \lambda_i, \nabla_W \nabla_{\dot\gamma} v_i - \nabla_W v_{i+1} \rangle dt$$
$$+ \int_0^T \langle \lambda_k, \nabla_W \nabla_{\dot\gamma} v_k \rangle dt. \tag{139}$$

Then by construction, since $\nabla_W v_i = 0$,

$$
\int_0^T \langle \delta_\gamma E, W \rangle dt = -\langle \mathrm{Log}_{\gamma(T)} y, W(T) \rangle
$$
$$
+ \int_0^T \langle \lambda_0, \nabla_W \dot\gamma \rangle dt
$$
$$
+ \sum_{i=1}^k \int_0^T \langle \lambda_i, \nabla_W \nabla_{\dot\gamma} v_i \rangle dt. \qquad (140)
$$

Then using the Lie bracket and curvature identities, this is written as

$$
\int_0^T \langle \delta_\gamma E, W \rangle dt = -\langle \mathrm{Log}_{\gamma(T)} y, W(T) \rangle
$$
$$
+ \int_0^T \langle \lambda_0, \nabla_{\dot\gamma} W \rangle dt
$$
$$
+ \sum_{i=1}^k \int_0^T \langle \lambda_i, \nabla_{\dot\gamma} \nabla_W v_i + R(W, \dot\gamma) v_i \rangle dt,
$$
$$
(141)
$$

which is further simplified, again using the identity $\nabla_W v_i = 0$:

$$
\int_0^T \langle \delta_\gamma E, W \rangle dt = -\langle \mathrm{Log}_{\gamma(T)} y, W(T) \rangle
$$
$$
+ \int_0^T \langle \lambda_0, \nabla_{\dot\gamma} W \rangle dt
$$
$$
+ \sum_{i=1}^k \int_0^T \langle \lambda_i, R(W, \dot\gamma) v_i \rangle dt, \qquad (142)
$$

Using the Bianchi identities, it can be demonstrated that the curvature tensor satisfies the identity [9]:

$$
\langle A, R(B, C) D \rangle = -\langle B, R(D, A) C \rangle, \qquad (143)
$$

for any vectors A, B, C, D. The covariant derivative along γ is also integrated by parts to arrive at

$$
\int_0^T \langle \delta_\gamma E, W \rangle dt = -\langle \mathrm{Log}_{\gamma(T)} y, W(T) \rangle
$$
$$
+ \langle \lambda_0, W \rangle |_0^T - \int_0^T \langle \nabla_{\dot\gamma} \lambda_0, W \rangle dt
$$
$$
- \sum_{i=1}^k \int_0^T \langle R(v_i, \lambda_i) \dot\gamma, W \rangle dt. \qquad (144)
$$

Finally, gathering terms, the adjoint equation for λ_0 and its gradients are obtained:

$$
\delta_{\gamma(t)} E = 0 = -\nabla_{\dot\gamma} \lambda_0 - \sum_{i=1}^k R(v_i, \lambda_i) \dot\gamma, \quad t \in (0, T) \quad (145)
$$

$$
\delta_{\gamma(T)} E = 0 = -\mathrm{Log}_{\gamma(T)} y + \lambda_0 \qquad (146)
$$
$$
\delta_{\gamma(0)} E = -\lambda_0. \qquad (147)
$$

Along with the variations with respect to v_i, this constitutes the full adjoint system. Extension to the case of multiple data at multiple time points is trivial, and results in the adjoint system presented in Sect. 4.

Appendix C: Derivation of Adjoint Equations in Lie Groups

Let G be a Lie group with Lie algebra $\mathfrak{g}$, equipped with a right invariant metric. Let $\gamma : [0, 1] \to G$ be a polynomial in G of order k with right-trivialized velocities $\xi_i : [0, 1] \to \mathfrak{g}$. Recall the equations for a perturbation $Z, \delta\xi_i$ of this polynomial:

$$
\frac{d}{dt} Z = \delta\xi_1 - \mathrm{ad}_{\xi_1} Z \qquad (148)
$$
$$
\frac{d}{dt} \delta\xi_i = \overline{\nabla}_{\delta\xi_1} \xi_i + \overline{\nabla}_{\xi_1} \delta\xi_i + \delta\xi_{i+1}. \qquad (149)
$$

The second equation can be rewritten

$$
\frac{d}{dt} \delta\xi_i = \frac{1}{2} \mathrm{ad}_{\delta\xi_1} \xi_i + \frac{1}{2} \mathrm{sym}_{\delta\xi_1} \xi_i + \overline{\nabla}_{\xi_1} \delta\xi_i + \delta\xi_{i+1} \quad (150)
$$
$$
= -\frac{1}{2} \mathrm{ad}_{\xi_i} \delta\xi_1 + \frac{1}{2} \mathrm{sym}_{\xi_i} \delta\xi_1 + \overline{\nabla}_{\xi_1} \delta\xi_i + \delta\xi_{i+1}
$$
$$
(151)
$$
$$
= (-\overline{\nabla}_{\xi_i} + \mathrm{sym}_{\xi_i}) \delta\xi_1 + \overline{\nabla}_{\xi_1} \delta\xi_i + \delta\xi_{i+1}. \qquad (152)
$$

This suggests the following matrix form ODE:

$$
\frac{d}{dt}
\begin{pmatrix} Z \\ \delta\xi_1 \\ \vdots \\ \delta\xi_k \end{pmatrix}
$$
$$
=
\begin{pmatrix}
\mathrm{ad}_{\xi_1} & I & \cdots & 0 & 0 & 0 \\
0 & \mathrm{sym}_{\xi_1} & I & \cdots & 0 & \\
0 & -\overline{\nabla}_{\xi_2} + \mathrm{sym}_{\xi_2} & \overline{\nabla}_{\xi_1} & I & \cdots & 0 \\
\vdots & & & & & \vdots \\
0 & -\overline{\nabla}_{\xi_k} + \mathrm{sym}_{\xi_k} & 0 & \cdots & & \overline{\nabla}_{\xi_1}
\end{pmatrix}
$$
$$
\times
\begin{pmatrix} Z \\ \delta\xi_1 \\ \vdots \\ \delta\xi_k \end{pmatrix}. \qquad (153)
$$

In order to derive the adjoint Jacobi field, one simply computes the negative adjoint of the matrix in the above equa-

tion. The adjoint of the above matrix is

$$
\begin{pmatrix}
-\mathrm{ad}^\dagger_{\xi_1} & 0 & \cdots & 0 & 0 \\
-I & -\mathrm{sym}^\dagger_{\xi_1} & \overline{\nabla}^\dagger_{\xi_2}-\mathrm{sym}^\dagger_{\xi_2} & \cdots & \overline{\nabla}^\dagger_{\xi_k}-\mathrm{sym}^\dagger_{\xi_k} \\
0 & -I & -\overline{\nabla}_{\xi_1} & 0 & \cdots \\
\vdots & & & & 0 \\
0 & 0 & \cdots & -I & -\overline{\nabla}^\dagger_{\xi_1}
\end{pmatrix}.
\tag{154}
$$

Now note that the adjoint of the $\overline{\nabla}_\xi$ operator is $-\overline{\nabla}_\xi$, since (using Eq. (52))

$$
2\overline{\nabla}^\dagger_X Y = \mathrm{ad}^\dagger_X Y + \mathrm{sym}^\dagger_X Y
\tag{155}
$$

$$
= \mathrm{ad}^\dagger_X Y - \mathrm{ad}_X Y + \mathrm{ad}^\dagger_Y X
\tag{156}
$$

$$
= -\mathrm{ad}_X Y - \mathrm{sym}_X Y
\tag{157}
$$

$$
= -2\overline{\nabla}_X Y.
\tag{158}
$$

Now let $\lambda_0, \dots, \lambda_k : [0,1] \to \mathfrak{g}$ be adjoint variables representing gradients with respect to position γ and velocities $\xi_1, \dots, \xi_k$. Using the equations above, we write the reduced polynomial adjoint equations as

$$
\frac{d}{dt}\lambda_0 = -\mathrm{ad}^\dagger_{\xi_1}\lambda_0
\tag{159}
$$

$$
\frac{d}{dt}\lambda_1 = -\lambda_0 - \mathrm{sym}^\dagger_{\xi_1}\lambda_1 + \sum_{i=2}^{k}\left(-\overline{\nabla}_{\xi_k}-\mathrm{sym}^\dagger_{\xi_k}\right)\lambda_k
\tag{160}
$$

$$
\frac{d}{dt}\lambda_i = -\lambda_{i-1} + \overline{\nabla}_{\xi_1}\lambda_i \quad i = 2, \dots, k.
\tag{161}
$$

The first adjoint variable, λ_0, takes on jump discontinuities when passing data points, which are derived identically to the geodesic case. Also note that this derivation is for right invariant metrics using right trivialized vectors, but the equivalent derivation in the case of left invariance is essentially identical.

References

1. Bandulasiri, A., Gunathilaka, A., Patrangenaru, V., Ruymgaart, F., Thompson, H.: Nonparametric shape analysis methods in glaucoma detection. I. J. Stat. Sci. **9**, 135–149 (2009)
2. Bookstein, F.L.: Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge Univ. Press, Cambridge (1991)
3. Bou-Rabee, N.: Hamilton-Pontryagin integrators on Lie groups. Ph.D. thesis, California Institute of Technology (2007)
4. Bruveris, M., Gay-Balmaz, F., Holm, D., Ratiu, T.: The momentum map representation of images. J. Nonlinear Sci. **21**(1), 115–150 (2011)
5. Burnham, K., Anderson, D.: Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York (2002)
6. Cates, J., Fletcher, P.T., Styner, M., Shenton, M., Whitaker, R.: Shape modeling and analysis with entropy-based particle systems. In: Proceedings of Information Processing in Medical Imaging (IPMI) (2007)
7. Cheeger, J., Ebin, D.G.: Comparison Theorems in Riemannian Geometry, vol. 365. AMS Bookstore, Providence (1975)
8. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.C.: Population shape regression from random design data. Int. J. Comput. Vis. **90**(2), 255–266 (2010)
9. do Carmo, M.P.: Riemannian Geometry, 1st edn. Birkhäuser, Boston (1992)
10. Driesen, N, Raz, N: The influence of sex, age, and handedness on corpus callosum morphology: A meta-analysis. Psychobiology (1995). doi:10.3758/BF03332028
11. Dryden, I.L., Kume, A., Le, H., Wood, A.T.: A multi-dimensional scaling approach to shape analysis. Biometrika **95**(4), 779–798 (2008)
12. Fletcher, PT: Geodesic regression and the theory of least squares on Riemannian manifolds. Int. J. Comput. Vis. (2012). doi:10.1007/s11263-012-0591-y
13. Fletcher, P.T., Liu, C., Pizer, S.M., Joshi, S.C.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging **23**(8), 995–1005 (2004)
14. Giambò, R., Giannoni, F., Piccione, P.: An analytical theory for Riemannian cubic polynomials. IMA J. Math. Control Inf. **19**(4), 445–460 (2002)
15. Hinkle, J., Muralidharan, P., Fletcher, P.T., Joshi, S.C.: Polynomial regression on Riemannian manifolds. In: ECCV, Florence, Italy, vol. 3, pp. 1–14 (2012)
16. Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler-Poincaré equations and semidirect products with applications to continuum theories. Adv. Math. **137**, 1–81 (1998)
17. Huckemann, S., Hotz, T., Munk, A.: Intrinsic shape analysis: geodesic principal component analysis for Riemannian manifolds modulo Lie group actions. Discussion paper with rejoinder. Stat. Sin. **20**, 1–100 (2010)
18. Jupp, P.E., Kent, J.T.: Fitting smooth paths to spherical data. Appl. Stat. **36**(1), 34–46 (1987)
19. Kendall, D.G.: Shape manifolds, procrustean metrics, and complex projective spaces. Bull. Lond. Math. Soc. **16**(2), 81–121 (1984)
20. Kendall, D.G.: A survey of the statistical theory of shape. Stat. Sci. **4**(2), 87–99 (1989)
21. Kenobi, K., Dryden, I.L., Le, H.: Shape curves and geodesic modelling. Biometrika **97**(3), 567–584 (2010)
22. Kume, A., Dryden, I.L., Le, H.: Shape-space smoothing splines for planar landmark data. Biometrika **94**(3), 513–528 (2007). doi:10.1093/biomet/asm047
23. Le, H., Kendall, D.G.: The Riemannian structure of Euclidean shape spaces: a novel environment for statistics. Ann. Stat. **21**(3), 1225–1271 (1993)
24. Silva Leite, F, Krakowski, K: Covariant differentiation under rolling maps. Departamento de Matemática, Universidade de Coimbra, Portugal (2008), No. 08-22, 1–8
25. Lewis, A., Murray, R.: Configuration controllability of simple mechanical control systems. SIAM J. Control Optim. **35**(3), 766–790 (1997)
26. Machado, L., Leite, F.S., Krakowski, K.: Higher-order smoothing splines versus least squares problems on Riemannian manifolds. J. Dyn. Control Syst. **16**(1), 121–148 (2010)
27. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. **19**(9), 1498–1507 (2007)
28. Marsden, J., Ratiu, T.: Introduction to Mechanics and Symmetry: a Basic Exposition of Classical Mechanical Systems, vol. 17. Springer, Berlin (1999)

29. Micheli, M., Michor, P., Mumford, D.: Sectional curvature in terms of the cometric, with applications to the Riemannian manifolds of landmarks. SIAM J. Imaging Sci. **5**(1), 394–433 (2012)

30. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. J. Math. Imaging Vis. **24**(2), 209–228 (2006). doi:10.1007/s10851-005-3624-0

31. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI) (2011)

32. Noakes, L., Heinzinger, G., Paden, B.: Cubic splines on curved surfaces. IMA J. Math. Control Inf. **6**, 465–473 (1989)

33. O'Neill, B.: The fundamental equations of a submersion. Mich. Math. J. **13**(4), 459–469 (1966)

34. Shi, X., Styner, M., Lieberman, J., Ibrahim, J.G., Lin, W., Zhu, H.: Intrinsic regression models for manifold-valued data. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2009, pp. 192–199. Springer, Berlin (2009)

35. Singh, N., Wang, A., Sankaranarayanan, P., Fletcher, P., Joshi, S.: Genetic, structural and functional imaging biomarkers for early detection of conversion from MCI to AD. In: Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 132–140. Springer, Berlin (2012)

36. Turaga, P., Veeraraghavan, A., Srivastava, A., Chellappa, R.: Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(11), 2273–2286 (2011)

37. Vaillant, M., Glaunes, J.: Surface matching via currents. In: Information Processing in Medical Imaging, pp. 1–5. Springer, Berlin (2005)

38. Younes, L., Qiu, A., Winslow, R., Miller, M.: Transport of relational structures in groups of diffeomorphisms. J. Math. Imaging Vis. **32**(1), 41–56 (2008)

39. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage **31**(3), 1116–1128 (2006)

Jacob Hinkle earned B.S. degrees in mathematics and physics from Miami University in May 2006. In 2007, he joined the Scientific Computing and Imaging (SCI) Institute at the University of Utah. He received his Ph.D. in Bioengineering from the University of Utah in July 2013. His current research interests include medical image analysis, computational anatomy, and manifold-based statistics.

P. Thomas Fletcher received his B.A. degree in Mathematics at the University of Virginia in 1999. He received an M.S. in Computer Science in 2002 followed by a Ph.D. in Computer Science in 2004 from the University of North Carolina at Chapel Hill. He is currently an Assistant Professor in the School of Computing at the University of Utah. His research interests include manifold statistics, shape analysis, medical image analysis

Sarang Joshi joined SCI as an Associate Professor of the Department of Bio Engineering in 2006. Before coming to Utah, Dr. Joshi was an Assistant Professor of Radiation Oncology and an Adjunct Assistant Professor of Computer Science at the University of North Carolina in Chapel Hill. Prior to joining Chapel Hill Dr. Joshi was Director of Technology Development at IntellX, a Medical Imaging start-up company which was later acquired by Medtronic. Sarang's research interests are in the field of Computational Anatomy. The principal aim of computational anatomy is the development of specialized mathematical and computational tools for the precise study of anatomical variability and the application of these tools for improved medical treatment, diagnosis and understanding of disease. In 2005 he spent a year on sabbatical at DKFZ (German Cancer Research Center) in Heidelberg, Germany, as a visiting scientist in the Department of Medical Physics where he focused on developing four dimensional radiation therapy approaches for improved treatment of Prostate and Lung Cancer. He was also one of the founding partner of Morphormics, Inc. which was recently acquired by Accuray. He has won numerous awards including the 2007 David Marr Best Paper Award, The international journal Signal Processing 2010 Most cited paper Award, and MICCAI 2010 Best of the Journal Issue Award. He holds numerous patents in the area of image registration and segmentation