

Chapter 13

Analyzing Shape Datasets



We present in this chapter some “shape analysis” methods, among those that are mainly used in practice, where the goal is to provide a low-dimensional description and to perform statistical validations of hypotheses for datasets in which each object is a shape. Most recent applications of this framework have taken place in medical imaging, in which the shapes are provided by anatomical regions segmented by MRI or computer tomography scans. The analysis of the anatomy derived from such images is called *computational anatomy*, with a framework introduced in [137, 138], and has since generated a huge literature. Beside this important range of applications, shape analysis can also be used in computed vision, or in biology, which was, for example, the main focus of D’Arcy-Thompson’s seminal treatise [276] on Growth and Form. We here focus on methods that derive from the analysis of diffeomorphisms developed in the previous chapters, leading to “morphometric” [23, 68], or “diffeomorphometric” [197] analyses.

13.1 Reference-Based Representation

The diffeomorphic matching methods that were described in this book can be seen to have a dual purpose. As a first goal, they provide a comparison tool between shapes, generally based on a formal or rigorous Riemannian paradigm. They also, by nature, provide an algorithm that aligns a target shape along a reference shape, i.e., that estimates a diffeomorphism φ , such that $\varphi \cdot (\text{reference}) \simeq (\text{target})$. This correspondence, φ , can be seen as a representation of the relationship between the reference and the target in the diffeomorphism group, i.e., a parametrization of the target relative to the reference.

In more formal terms, registration algorithms provide, given a reference \bar{m} , a mapping $m \mapsto \Phi(m)$ from a shape space to the diffeomorphism group such that $\Phi(m) \cdot \bar{m} \simeq m$. From a dataset (m_1, \dots, m_N) of shapes, one can then obtain a dataset $(\varphi_1, \dots, \varphi_N)$ of diffeomorphisms, with $\varphi_k = \Phi(m_k)$. Even though diffeomorphisms may appear as more complex objects than many shapes, this representation actually

simplifies the analysis of the dataset. It is certainly natural to restrict this analysis to the restriction of the diffeomorphism (or its derivative) to the reference, which (e.g., when dealing with landmarks) can represent a huge reduction of dimension. Image or shape morphometry, as described, for example, in [23], determines features, or descriptors, of shapes in the dataset based on these diffeomorphisms, using point displacements or Jacobian matrices. These features can then be used in a statistical learning framework to draw conclusions on properties of interest about the dataset.

When dealing with shape spaces, M , that do not involve diffeomorphisms, such as Kendall's space (see Sect. 11.2), other reference-based representations can be used, the most natural, in the Riemannian case, being to use exponential charts for the metric, such that $\Phi(m) = v \in T_{\bar{m}}M$ with $m = \text{Exp}_{\bar{m}}(v)$. The statistical analysis can then be based on the vectors v_1, \dots, v_N that all belong to the same vector space. This point of view can actually be applied to the diffeomorphic representation, where one can use exponential charts, or, equivalently, the momentum representation, associated with right-invariant metrics on diffeomorphisms. Notice that the LDDMM algorithms described in Sect. 10.3 and later directly return such a representation while estimating an optimal correspondence.

Building such representations requires selecting a proper reference shape. While one can use any fixed shape for this purpose, it is, for many reasons, preferable to choose \bar{m} close to the studied dataset. It is understandably easier to analyze diffeomorphisms when the deformation they define are not too severe. Also, tangent space representations linearize the shape space, and one wants to reduce as much as possible the metric distortions they induce. This is why one typically computes \bar{m} as some kind of average of the dataset under study.

When using morphometric methods, one often estimates \bar{m} and computes its optimal correspondences with the dataset in a single algorithm, which is often called *groupwise registration* [26, 35, 36, 160, 180, 181, 288]. In its simplest form, when the registration between \bar{m} and m minimizes a cost function $U_{\bar{m},m}(\varphi)$, the associated groupwise registration minimizes $\sum_{k=1}^N U_{\bar{m},m_k}(\varphi_k)$ with respect to $\bar{m}, \varphi_1, \dots, \varphi_N$. Some additional regularization constraints may also be used for the reference. For example, if one uses the LDDMM algorithm, one can define a groupwise registration method for image matching via the minimization of (using the notation Sect. 10.3, V being an admissible space)

$$\frac{1}{2} \sum_{k=1}^N \int_0^1 \|v_k(t)\|_V^2 dt + \frac{1}{\sigma^2} \sum_{k=1}^N \|\bar{m} \circ \varphi_{10}^{v_k} - m_k\|_2^2 \quad (13.1)$$

with respect to v_1, \dots, v_N and \bar{m} . When \bar{m} is fixed, this provides N independent image registration problems, and when v_1, \dots, v_N are fixed, the optimal \bar{m} is given by

$$\bar{m} = \frac{1}{N} \sum_{k=1}^N m_k \circ \varphi_{01}^{v_k} \det(d\varphi_{01}^{v_k}).$$

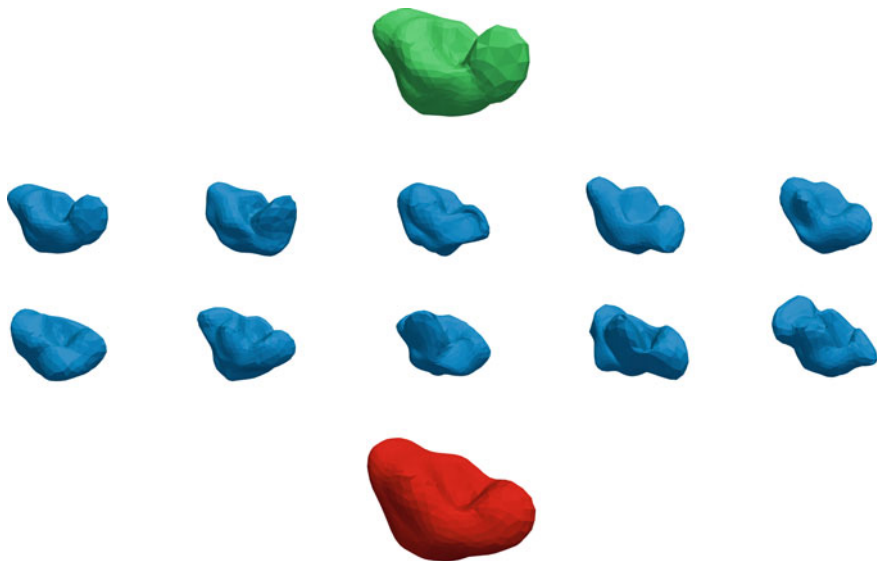


Fig. 13.1 Example of template estimation. Rows 2 and 3: ten randomly generated shapes. Row 1: hypertemplate, chosen as the first shape in the sequence. Row 4: estimated template

A modification of this method has been introduced in [180] where the reference is in the form $\bar{m} = m_0 \circ \psi_{10}^{v_0}$ where m_0 is fixed and v_0 is estimated jointly with the rest of the variables by minimizing

$$\frac{\lambda}{2} \int_0^1 \|v_0(t)\|_V^2 dt + \frac{1}{2} \sum_{k=1}^N \int_0^1 \|v_k(t)\|_V^2 dt + \frac{1}{\sigma^2} \sum_{k=1}^N \|m_0 \circ \varphi_{10}^{v_0} \circ \varphi_{10}^{v_k} - m_k\|^2$$

for some $\lambda > 0$. This constrains the topology of the estimated reference image to conform to that of the image m_0 . A similar approach has been introduced in [181] for surface matching, in which m_0 is referred to as a *hypertemplate*. One interesting feature of this approach is that it can be represented as a family of branching optimal control problems, each with its own maximum principle that can also be branched backwards in time to compute the gradient of the objective function: one first uses v_0 as a control leading from the hypertemplate to the template, then v_1, \dots, v_n as controls driving the template to the targets. An example of template estimation with this method is provided in Fig. 13.1.

When one uses a tangent representation on a shape manifold, \bar{m} is often estimated as a Fréchet mean, or Riemannian center of mass, of the collection m_1, \dots, m_N . Such a mean is defined as a minimizer of

$$F : \bar{m} \mapsto \sum_{k=1}^N d_M(\bar{m}, m_k)^2,$$

where d_M is the geodesic distance on M . It is important to point out that this minimization problem does not always have a unique solution, i.e., that some datasets may have more than one Fréchet mean, even though this fact is often ignored in practice (or, one considers that any Fréchet mean is a good candidate for the reference shape). In finite dimensions, sufficient conditions for the uniqueness of such means involve the curvature of M and Rauch's comparison theorems. They are beyond the scope of this book, but we refer to [4, 62, 164, 168, 229] for additional details.

The computation of the gradient of F can be based on the following lemma, which we state without proof.

Lemma 13.1 *Let M be a Riemannian manifold. For $m_0 \in M$ define $f(m) = d_M(m_0, m)^2$. Let Exp_{m_0} be defined in some neighborhood of 0, $\Omega \subset T_{m_0}M$, and be a diffeomorphism onto its image. Then, for all $v_0 \in \Omega$, for all $m \in \text{Exp}_{m_0}(\Omega)$ and $h \in T_m M$,*

$$df(m)h = 2\langle \dot{\gamma}(1), h \rangle_m = -2\langle \dot{\tilde{\gamma}}(0), h \rangle_m,$$

where $\gamma(t)$, $t \in [0, 1]$, is the geodesic joining m_0 and m and $\tilde{\gamma}(t) = \gamma(1 - t)$ is the geodesic between m and m_0 .

From this, we can deduce that if the dataset is fully included in a domain Ω that contains minimizing geodesics between any of its points (i.e., it is *geodesically convex*), is such that each of these geodesics is uniquely defined, and such that, for all $m \in \Omega$, Exp_m is a diffeomorphism from an open neighborhood of 0 in $T_m M$ onto Ω , then

$$dF(\bar{m})h = -2 \sum_{k=1}^N \dot{\gamma}_k(0),$$

where γ_k is the geodesic between \bar{m} and m_k , and the Fréchet mean must satisfy

$$\sum_{k=1}^N \dot{\gamma}_k(0) = 0.$$

This computation leads, in particular, to gradient descent algorithms designed to estimate the mean (see [177]).

It is also possible to define a reference shape through a stochastic shape model, in which \bar{m} is deformed via random diffeomorphisms, possibly with additional noise, to generate m_1, \dots, m_N . The estimation of \bar{m} can then be performed using maximum likelihood. While describing in detail the associated statistical model and the estimation algorithm would take us too far from this discussion (and we refer to [7, 8, 173] for such details), it is important to note that minimizing (13.1) in this case may lead (when the noise level is high enough) to biased estimates of \bar{m} , in the sense that, even if N tends to infinity, minimizers of (13.1) will differ from \bar{m} when the model is valid (which does not mean, however, that they cannot be used as reference shapes for subsequent morphometric analyses). See in particular [81] for a theoretical analysis of the issue.

13.2 Principal Component Analysis

Principal component analysis (PCA) is the simplest and most widely used method of performing dimension reduction for data analysis [92]. It is especially useful in shape analysis, which deals with virtually infinite-dimensional objects. The reader may refer, if needed, to the basic description of the method that is provided in Appendix E in the appendix for Hilbert spaces. In this section we focus on the specific adaptation of the approach to nonlinear shape spaces.

PCA is indeed a linear method, designed to be applied to vector spaces equipped with an inner product. On shape spaces, and more generally on Riemannian manifolds, a standard approach relies on a tangent-space linearization of the manifold using exponential charts (this is often referred to as tangent PCA). More precisely, given a dataset (m_1, \dots, m_N) and a reference element \bar{m} one computes normal coordinates $h_1, \dots, h_N \in T_{\bar{m}}M$ such that $m_k = \text{Exp}_{\bar{m}}(h_k)$ for $k = 1, \dots, N$, and performs PCA on the collection (h_1, \dots, h_N) using the Riemannian inner product $\langle \cdot, \cdot \rangle_{\bar{m}}$. The p first principal components then provide an orthonormal family (e_1, \dots, e_p) spanning a subspace of $T_{\bar{m}}M$, and the PCA representation is given by

$$\Phi : (\lambda_1, \dots, \lambda_p) \mapsto \text{Exp}_{\bar{m}} \left(\sum_{j=1}^p \lambda_j e_j \right) \in M.$$

When working with shape spaces with a metric induced by a right-invariant Riemannian metric on diffeomorphisms through a Riemannian submersion, it is easier, and formally equivalent, to reformulate the problem in terms of Diff rather than M . One can see, in particular, that the LDDMM registration algorithm minimizes

$$\frac{1}{2} \int_0^1 \|v(t)\|_V^2 dt + U(\varphi_{01}^v \cdot \bar{m}, m_k)$$

for some data attachment term U , which is equivalent to minimizing

$$\frac{1}{2} \|h\|_V^2 + U(\text{Exp}_{\text{id}}(h) \cdot \bar{m}, m_k),$$

where Exp_{id} is the Riemannian exponential on the diffeomorphism group starting at $\varphi = \text{id}$. One can use the optimal h , say $h_k \in V$, as a representation of m_k on which PCA can be applied, using the V inner product. Using the notation introduced in Sect. 11.5.2, Definition 11.13, one can replace h by $\rho = \mathbb{L}h$ and solve the equivalent problem of minimizing

$$\frac{1}{2} \|\rho\|_{V^*}^2 + U(\text{Exp}_{\text{id}}^{\flat}(\rho) \cdot \bar{m}, m_k)$$

with optimal solution given by $\rho_k = \mathbb{L}h_k$. The advantage of doing so is the parsimony of the momentum representation, as discussed in Sect. 11.5.2. PCA is then performed on the dataset (ρ_1, \dots, ρ_N) using the V^* inner product. Once a PCA basis, say ξ_1, \dots, ξ_p , is computed, the representation is then

$$\Phi : (\lambda_1, \dots, \lambda_p) \mapsto \text{Exp}_{\text{id}}^b \left(\sum_{j=1}^p \lambda_j \rho_j \right) \cdot \bar{m} \in M.$$

Notice that this is a representation of the “deformed templates” ($\varphi_{01}^{v_k} \cdot \bar{m}, k = 1, \dots, N$) rather than of the original data (m_1, \dots, m_N) . This momentum PCA approach has been used multiple times in applications, starting with [290], in which it was introduced for landmark spaces. This formulation allows one to revisit the active shape model described in Sect. 6.2, in which shapes were represented by a decomposition in a linear basis, and develop a model based on the nonlinear representation associated with the function Φ above, which constrains the shape topology. Such approaches have been proposed in order to operate shape segmentation, or to regularize registrations in [285, 287].

Because tangent PCA is based on a linear representation of the manifold M , it necessarily suffers from the metric distortions that any linear representation must induce. The sum of residuals in the tangent space that is minimized by PCA may be quite different from the sum of squared distances of the actual shapes to their PCA representation provided by the mapping Φ . More precisely, one can formulate the search for p principal directions in tangent PCA as looking for a p -dimensional subspace $W \subset T_{\bar{m}}M$ such that

$$F(W) = \sum_{k=1}^p \min_{w \in W} \|h_k - w\|_{\bar{m}}^2 \quad (13.2)$$

is minimized, with $\text{Exp}_{\bar{m}}(h_k) = m_k$. However, in terms of approximating the dataset, one would probably be more interested in minimizing

$$F(W) = \sum_{k=1}^p \min_{w \in W} d_M(\text{Exp}_{\bar{m}}(w), m_k)^2, \quad (13.3)$$

which measures how far each shape is from its representation in the manifold. The two criteria may be quite different when the dataset is spread out away from \bar{m} and their solutions (the optimal W) may be quite different. Obviously, the first criterion is much easier to minimize than the second one, which represents a complex nonlinear optimization problem (with d_M usually non-explicit). One can make it slightly easier by building W one dimension at a time, starting with $p = 1$, in which one looks for the best geodesic approximating the data, progressively adding new directions without changing those that were found earlier. This procedure was introduced in [113] and called geodesic principal component analysis (GPCA). The non-incremental problem

requires a search within the space of all p -dimensional subspaces of $T_m M$, i.e., its Grassmann manifold of order p (cf. Sect. B.6.7).

Notice that one may opt for a simplified version of GPCA by replacing the Riemannian distance in M by some “extrinsic” discrepancy measure. For example, in the diffeomorphic framework, one can formulate the problem of finding a p -dimensional subspace W of V^* that minimizes

$$F(W) = \sum_{k=1}^p \min_{\rho \in W} U(\text{Exp}_{\text{id}}^b(\rho) \cdot \bar{m}, m_k), \quad (13.4)$$

in which U replaces the distance d_M and would be computationally more tractable. This problem can be rewritten in the form of finding ρ_1, \dots, ρ_N minimizing

$$G(\rho_1, \dots, \rho_N) = \sum_{k=1}^N U(\text{Exp}_{\text{id}}^b(\rho_k) \cdot \bar{m}, m_k)$$

subject to $\text{rank}(\rho_1, \dots, \rho_N) = p$. The problem in this form is tackled in [60], in which a gradient descent algorithm over p -dimensional subspaces of V^* is proposed.

13.3 Time Series

13.3.1 Single Trajectory

We now assume that the dataset (m_1, \dots, m_N) is a time series, so that it describes the evolution of a given shape captured at times, say, $\tau_1 < \dots < \tau_N$. We here study the regression problem of determining a function $\tau \mapsto m(\tau) \in M$ such that $m(\tau_k) \simeq m_k$.

Since geodesics are the Riemannian generalizations of straight lines in Euclidean spaces, one generalizes the standard linear regression model $m(\tau) = \bar{m} + \tau h$ to such spaces by looking for curves defined by $m(\tau) = \text{Exp}_{\bar{m}}(\tau h)$ for fixed $\bar{m} \in M$ and $v \in T_{\bar{m}} M$, which both need to be estimated from data. Notice that, in this case, \bar{m} is not an average of the considered dataset, but an “intercept”, representing the estimated position at $\tau = 0$. The resulting “geodesic regression” model [114] can then be associated with the generalization of least-square estimation, minimizing

$$\sum_{k=1}^N d_M(\text{Exp}_{\bar{m}}(\tau_k h), m_k)^2$$

with respect to \bar{m} and v . Notice that this problem is similar, but distinct from the search for a geodesic principal direction, which would first choose \bar{m} as a Fréchet mean, and then estimate h , with $\|h\|_{\bar{m}} = 1$, minimizing

$$\sum_{k=1}^N \min_{\tau} d_M(\text{Exp}_{\bar{m}}(\tau h), m_k)^2.$$

As discussed with PCA, the intrinsic error criterion using the Riemannian distance seldom leads to tractable optimization algorithms, and it is often replaced with a discrepancy measure that is more amenable to computation, minimizing

$$\sum_{k=1}^N U(\text{Exp}_{\bar{m}}(\tau_k h), m_k).$$

The exponential function being associated with a second-order differential equation, the derivative with respect to \bar{m} and h of each term in the sum above can be computed using the formulas derived in Sect. C.4 for the variation of solutions of ODEs with respect to their initial conditions. Some regularization may be added to the objective function, to control, for example, the topology of the intercept, \bar{m} , which may be chosen in the form $\bar{m} = \text{Exp}_{\bar{m}}(h_0) \cdot m_0$ for some fixed shape m_0 . Adding some penalty on the norms of h_0 and h , one can minimize

$$(h_0, h) \mapsto \lambda_0 \|h_0\|_{m_0}^2 + \lambda \|h\|_{\bar{m}}^2 + \sum_{k=1}^N U(\text{Exp}_{\bar{m}}(\tau_k h), m_k),$$

with $\bar{m} = \text{Exp}_{\text{id}}(h_0) \cdot m_0$. This model was implemented in [155] on spaces of surfaces with a metric induced by diffeomorphisms, with a similar approach developed in [110]. Still in the diffeomorphic framework, a geodesic regression algorithm for images has been proposed in [219], and an approach using image metamorphosis has been proposed in [153].

Notice also that one can spare the estimation of \bar{m} by assuming that $\tau_0 = 0$, considering the first observation as a baseline. This creates, however, an asymmetry in the data, in which the noise or variation from the geodesic is neglected for the baseline, which may sometimes be artificial [24, 242, 243].

It is not difficult to modify the previous framework to correct for the property that geodesics evolve at constant speed by making a time reparametrization of the trajectory. This corresponds to the model $m(\tau) = \text{Exp}_{\bar{m}}(f(\tau)h)$, where f is an increasing function from $[0, 1]$ to $[0, 1]$ that also needs to be estimated. This time reparametrization can be estimated using a method akin to LDDMM, modeling f as the result of a diffeomorphic flow [94], but simpler methods can be used, too, such as, for example, optimizing

$$(\bar{m}, h, \tilde{\tau}_1, \dots, \tilde{\tau}_N) \mapsto \sum_{k=1}^N d_M(\text{Exp}_{\bar{m}}(\tilde{\tau}_k h), m_k)^2$$

subject to $0 = \tilde{\tau}_1 < \tilde{\tau}_2 < \dots < \tilde{\tau}_{N-1} < \tilde{\tau}_N = 1$, which corresponds to monotonic regression with respect to time [155]. The derivative of $\text{Exp}_{\bar{m}}(\tilde{\tau}h)$ with respect to

$\tilde{\tau}$ is straightforward to compute, since it is given by the speed of the geodesic, and is readily obtained from the differential equation that is integrated to compute this geodesic.

Aligning all shapes along a single geodesic may sometimes be too restrictive when the time series exhibits several modes of variation, and more flexible methods can be derived. At the extreme end of this range of methods, one can use a piecewise geodesic approach, which consists in estimating h_1, \dots, h_{N-1} such that $\text{Exp}_{m_k}(h_k) = m_{k+1}$ with $h_k \in T_{m_k}M$. One can then define

$$m(\tau) = \text{Exp}_{m_k} \left(\frac{\tau - \tau_k}{\tau_{k+1} - \tau_k} h_k \right)$$

for $\tau \in [\tau_k, \tau_{k+1}]$. This is just the Riemannian generalization of a piecewise linear curve interpolating the observed trajectory. One can modify this formulation by allowing for some error in the interpolation, thus taking into account possible measurement noise in the observed m_k 's, minimizing, for example,

$$\frac{\lambda_0}{2} \|h_0\|_{\bar{m}_0}^2 + \frac{\lambda}{2} \sum_{k=0}^{N-1} \|h_k\|_{\bar{m}_k}^2 + \sum_{k=1}^N U(\bar{m}_k, m_k),$$

where \bar{m}_k is defined recursively by $\bar{m}_{k+1} = \text{Exp}_{\bar{m}_k}(h_k)$, $h_k \in T_{\bar{m}_k}M$ and \bar{m}_0 is a fixed shape. One obtains an equivalent formulation with the following time-continuous problem of minimizing

$$\frac{\lambda_0}{2} \|h_0\|_{\bar{m}_0}^2 + \frac{\lambda}{2} \int_0^1 \|\dot{\gamma}(t)\|_{\dot{\gamma}(t)}^2 dt + \sum_{k=1}^N U(\gamma(\tau_k), m_k) \quad (13.5)$$

subject to $\gamma(0) = \text{Exp}_{\bar{m}_0}(h_0)$. Indeed, the solution γ must be a minimizing geodesic between $\bar{m}_k := \gamma(\tau_k)$ and \bar{m}_{k+1} , with constant speed, which directly leads to a piecewise geodesic solution. In the LDDMM framework, this equivalent formulation reduces to minimizing

$$\frac{\lambda_0}{2} \|v_0\|_V^2 + \frac{\lambda}{2} \int_0^1 \|v(t)\|_V^2 dt + \sum_{k=1}^N U(\varphi_{0\tau_k}^v \circ \text{Exp}_{\text{id}}(v_0) \cdot \bar{m}_0, m_k)$$

with respect to $v_0 \in V$ and $t \mapsto v(t) \in L^2([0, 1], V)$ (see [196, 197, 204]).

Piecewise geodesic interpolation is continuous in time, but not differentiable, and will certainly be too sensitive to noise, even when using inexact interpolation. Time differentiability of the solution can be obtained by controlling the second derivative of γ instead of the first derivative in (13.5), leading to a Riemannian generalization of interpolating splines. As discussed in Sect. B.6.4, curve acceleration in Riemannian manifolds involves the covariant derivative, and a formulation of the Riemannian spline problem can be obtained by replacing $\dot{\gamma}(t)$ by $\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t)$ in (13.5) [220].

The analysis of the new variational problem becomes more involved when studied intrinsically on the manifold, and here we restrict to situations in which one can work in a local chart, and take advantage of the Hamiltonian formulation of geodesics described in Sect. B.6.6. Using the notation of this section we let $S(m)$ be the representation of the metric in the chart, so that $\|h\|_m^2 = h^T S(m)h$, with the notation abuse of using h in order to represent a vector in $T_m M$ and its expression in a chart. Letting $H(m, a) = a^T S(m)^{-1}a/2$, with $a \in T_m M^*$, the geodesic equations in Hamiltonian form are

$$\begin{cases} \partial_t m = S(m)^{-1}a \\ \partial_t a + \frac{1}{2}\partial_m(a^T S(m)^{-1}a) = 0. \end{cases}$$

Moreover, given a curve γ on M (not necessarily a geodesic), one has, letting $a(t) = S(\gamma(t))\dot{\gamma}(t)$,

$$\nabla_{\dot{\gamma}}\dot{\gamma} = S(\gamma)^{-1} \left(\partial_t a + \frac{1}{2}\partial_{\gamma}(a^T S(\gamma)^{-1}a) \right).$$

One can then reformulate the Riemannian spline problem as an optimal control problem, with state (γ, a) and control u , minimizing

$$\frac{\lambda_0}{2}\|h_0\|_{\bar{m}_0}^2 + \frac{\lambda}{2}\int_0^1 u(t)^T S(\gamma(t))^{-1}u(t) dt + \sum_{k=1}^N U(\gamma(\tau_k), m_k)$$

subject to the state equation

$$\begin{cases} \partial_t \gamma = S(\gamma)^{-1}a \\ \partial_t a + \frac{1}{2}\partial_{\gamma}(a^T S(\gamma)^{-1}a) = u \end{cases}$$

with initial condition $\gamma(0) = \text{Exp}_{\bar{m}_0}(h_0)$ and free condition for $a(0)$. One can consider higher-order Riemannian splines by iterating covariant derivatives (see, e.g., [122, 182]). This approach to the spline problem was introduced for shape spaces with a Riemannian metric induced by diffeomorphisms in [284], with further developments in [265]. In this case, one can take advantage of the right-invariance of the metric in the group to reformulate the problem as minimizing

$$\frac{\lambda_0}{2}\|v_0\|_V^2 + \frac{\lambda}{2}\int_0^1 \|u(t)\|^2 dt + \sum_{k=1}^N U(\varphi_{0\tau_k}^v \circ \text{Exp}_{\text{id}}(v_0) \cdot \bar{m}_0, m_k)$$

subject to

$$\begin{cases} \partial_t \gamma = (\mathbb{K}\rho) \cdot \gamma \\ \partial_t \rho + \text{ad}_{\mathbb{K}\rho}^* \rho = u \end{cases}$$



Fig. 13.2 Time series with four surfaces: left to right: $t = 1, 2, 3, 4$

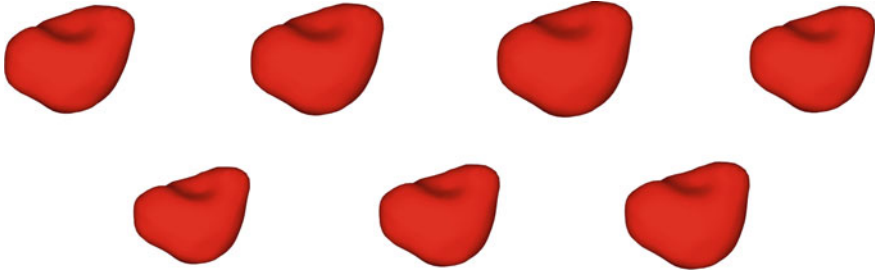


Fig. 13.3 Piecewise geodesic interpolation of sequence in Fig. 13.2 with seven time points (left to right and top to bottom): $t = 1, 1.5, 2, 2.5, 3, 3.5, 4$

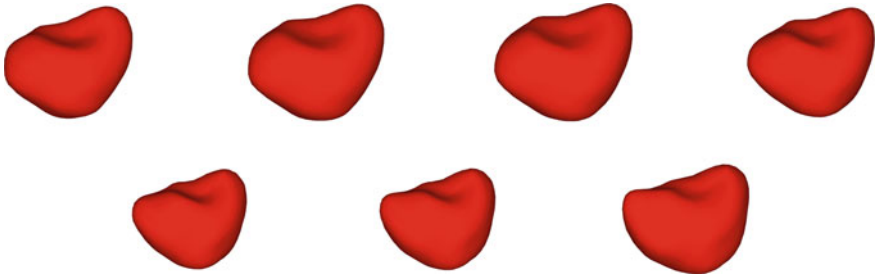


Fig. 13.4 Spline interpolation of sequence in Fig. 13.2 with seven time points (left to right and top to bottom): $t = 1, 1.5, 2, 2.5, 3, 3.5, 4$

with $\rho, u \in V^*$, \mathbb{K} the inverse duality operator of V , and where the second equation in the system is the EPDiff equation. Notice that the norm on u in the integral is left unspecified, and there is much flexibility in choosing it, because u now belongs to a fixed space, V^* . One can take, in particular, any metric on a space W^* that is continuously embedded in V^* (so that V is embedded in W), bringing more regularity constraints to the control u . This includes, in particular, the L^2 norm, which significantly simplifies the implementation of the problem. Figures 13.2, 13.3, 13.4, 13.5 and 13.6 compares the interpolation schemes on a sequence of four target surfaces. The piecewise geodesic and spline methods interpolate the target almost exactly, with some small differences at intermediate points. The geodesic interpolation is more regular, but makes large errors interpolating the sequence. The difference between the methods is especially apparent when plotting the volumes of the interpolated surfaces over time (Fig. 13.6).

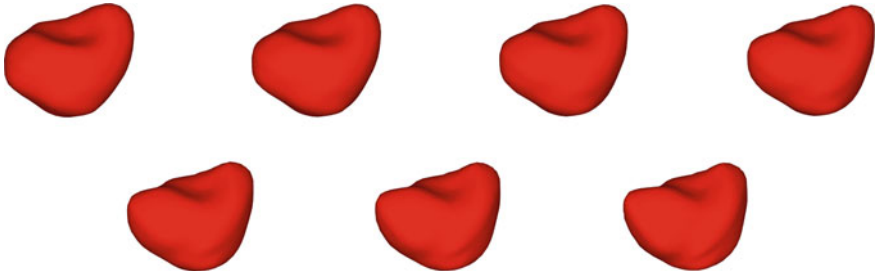


Fig. 13.5 Geodesic interpolation of sequence in Fig. 13.2 with seven time points (left to right and top to bottom): $t = 1, 1.5, 2, 2.5, 3, 3.5, 4$

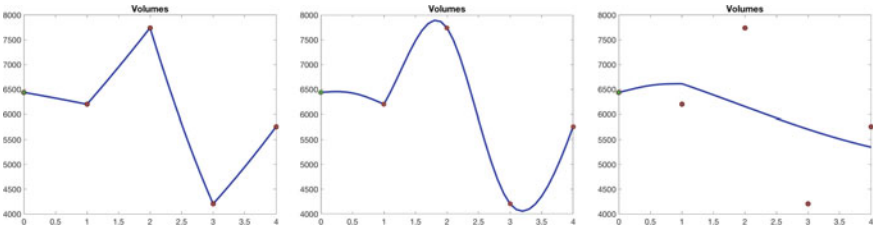


Fig. 13.6 Evolution of the volumes of the interpolated surfaces in Fig. 13.2. The first leg of the trajectory (from time 0 to 1) is the adjustment of the baseline starting with the volume of the “hypertemplate”. The dots that follow are the volumes of the target surfaces. The time interpolation step is $\delta t = 0.1$. From left to right: piecewise geodesic, spline and geodesic

13.3.2 Multiple Trajectories

We now consider the situation in which several time series are observed, and start with the problem of computing an average trajectory from them. Assume, to begin with, that one observes full trajectories in the form of functions $m_k : [0, 1] \rightarrow M$ (assuming that the time interval has been rescaled to $[0, 1]$), for $k = 1, \dots, n$. The goal is to compute an average trajectory \bar{m} .

The simplest and most direct approach is to apply one of the averaging methods that were discussed in Sect. 13.1 to each time coordinate separately. For example, one can define $\bar{m}(\tau)$ as a Fréchet mean, minimizing

$$\sum_{k=1}^n d_M(\bar{m}(\tau), m_k(\tau))^2$$

for each τ . This requires however that the observed trajectories are correctly aligned with each other, which may be valid in some contexts (e.g., for cardiac motion, which can be parametrized using well defined epochs in the cardiac cycle) but not always. In the general case, averaging has to be combined with some time realignment.

As a possible approach, let us consider this problem within the metamorphosis framework that was discussed in the previous chapter, which can be used to place a Riemannian metric on the space of trajectories, with respect to which Fréchet means can be computed while allowing for changes of parametrization. So, consider a metamorphosis metric in which the acting group is the space of diffeomorphisms of $\Omega = [0, 1]$ acting on curves via $g \cdot m = m \circ g^{-1}$ (reparametrization). Consider an RKHS H on functions defined over $[0, 1]$ that vanish at 0 and at 1. We can associate a metamorphosis to the function

$$F(\xi, m, z) = \|\xi\|_H^2 + \frac{1}{\sigma^2} \|z\|_2^2$$

defined for $\xi \in H, m \in C^k(\Omega, M)$ and z a vector field along m (so that $z(\tau) \in T_{m(\tau)}$ for all $\tau \in \Omega$), with

$$\|z\|_2^2 = \int_0^1 \|z(\tau)\|_{m(\tau)}^2 d\tau.$$

The squared distance between \bar{m} and m_k itself can be computed by minimizing

$$\int_0^1 \|\xi(t, \cdot)\|_H^2 dt + \frac{1}{\sigma^2} \int_0^1 \|z(t, \cdot)\|_2^2 dt$$

subject to $m(0, \cdot) = \bar{m}, m(1, \cdot) = m_k$ and $\partial_t m + \xi \partial_\tau m = z$. We use here the convention of denoting by $t \in [0, 1]$ the (numerical) metamorphosis time and $\tau \in \Omega (= [0, 1])$ the “real” time associated with observed trajectories. Defining g as the flow of the equation $\partial_t g(t, \tau) = \xi(t, g(t, \tau))$, and letting $\alpha(t, \tau) = m(t, g(t, \tau))$, this objective function can be rewritten as

$$\int_0^1 \|\xi(t, \cdot)\|_H^2 dt + \frac{1}{\sigma^2} \int_0^1 \int_\Omega \|\partial_t \alpha(t, \tau)\|_{\alpha(t, \tau)}^2 \partial_\tau g d\tau dt,$$

which, after a change of variable in time, takes the form

$$\int_0^1 \|\xi(t, \cdot)\|_H^2 dt + \frac{1}{\sigma^2} \int_\Omega \left(\frac{1}{c_g(\tau)} \int_0^1 \|\partial_t \tilde{\alpha}(t, \tau)\|_{\tilde{\alpha}(t, \tau)}^2 dt \right) d\tau,$$

where

$$c_g(\tau) = \int_0^1 \partial_\tau g^{-1} dt$$

and $\tilde{\alpha}(t, \tau) = \alpha(\lambda(t, \tau), \tau)$ for some invertible time change $\lambda(\cdot, \tau)$ from $[0,1]$ onto itself. (See the computation following Eq. (12.28).) This has to be minimized in ξ and α (or $\tilde{\alpha}$) with the constraints $\alpha(0) = \bar{m}$ and $\alpha(1) = m_k \circ g(1), \partial_t g = \xi \circ g$. Using the fact that $\tilde{\alpha}(\cdot, \tau)$ minimizes the geodesic energy on M between \bar{m} and $m_k \circ g(1)$, we finally find that computing the distance can be done by minimizing

$$\int_0^1 \|\xi(t, \cdot)\|_H^2 dt + \frac{1}{\sigma^2} \int_{\Omega} \frac{d_M(\bar{m}(\tau), m_k(g(1, \tau)))^2}{c_g(\tau)} d\tau$$

with respect to ξ . A Fréchet mean between m_1, \dots, m_n for the metamorphosis metric should therefore minimize, with respect to \bar{m} and ξ_1, \dots, ξ_n ,

$$\sum_{k=1}^n \int_0^1 \|\xi_k(t, \cdot)\|_H^2 dt + \frac{1}{\sigma^2} \sum_{k=1}^n \int_{\Omega} \frac{d_M(\bar{m}(\tau), m_k(g_k(1, \tau)))^2}{c_{g_k}(\tau)} d\tau$$

with $\partial_t g_k(t, \tau) = \xi_k(t, g_k(t, \tau))$. One can use an alternating minimization scheme to solve this problem since, with fixed \bar{m} , ξ_1, \dots, ξ_n are solutions of independent “ordinary” metamorphosis problems on M , and for fixed ξ_1, \dots, ξ_n , the average $\bar{m}(\tau)$ can be obtained, for each τ , as a weighted Fréchet average minimizing

$$\sum_{k=1}^n d_M(\bar{m}(\tau), m_k(g_k(1, \tau)))^2 / c_{g_k}(\tau).$$

Some modifications to this formulation are still needed in the case of shape spaces acted upon by diffeomorphisms, because the geodesic distance in this case is generally not computable exactly, but must be approximated through algorithms such as LDDMM. For example, on spaces of surfaces, one can replace $d_M(\bar{m}(\tau), m_k(g_k(1, \tau)))^2$ by the minimizer of

$$\int_0^1 \|v_k^\tau\|_V^2 dt + D(\varphi_{01}^{v_k^\tau} \cdot \bar{m}, m_k \circ g_k(1, \tau))$$

for some discrepancy measure D , such as those described in Sect. 9.7.3. The minimization then needs to be done with respect to $\xi_1, \dots, \xi_n, v_1, \dots, v_n$ and \bar{m} . When ξ_1, \dots, ξ_n is fixed, this is the same problem as the one considered in (13.1) and below, and can be solved separately for each τ .

With fixed \bar{m}, v_1, \dots, v_n the problem splits into n independent problems, each of them requiring the minimization of a function taking the form

$$\int_0^1 \|\xi_k\|_H^2 dt + \int_{\Omega} \frac{\Phi_k(\tau, m_k \circ g_k(1, \tau))}{c_{g_k}(\tau)} d\tau,$$

with

$$\Phi(\tau, \tilde{m}) = \int_0^1 \|v_k^\tau\|_V^2 dt + D(\varphi_{01}^{v_k^\tau} \cdot \tilde{m}, \tilde{m}).$$

The gradient of this objective function with respect to ξ_k can be obtained using the formulas developed in Sect. C.5 for the differentiation of solutions of ordinary differential equations (we skip the details). One can obtain a simpler method

by disregarding the weights c_g that come from the metamorphosis metric, and just minimize

$$\sum_{k=1}^n \int_0^1 \|\xi_k\|_H^2 dt + \sum_{k=1}^n \int_0^1 \int_{\Omega} \|v_k^\tau\|_V^2 d\tau dt + \sum_{k=1}^n \int_{\Omega} D(\varphi_{01}^{v_k^\tau} \cdot \bar{m}, m_k \circ g_k(1, \tau)) d\tau,$$

leading to a formulation similar to that developed in [95].

When dealing with sparse observations, i.e., when the k th trajectory is observed at a small number of time points $\tau_{k,1}, \dots, \tau_{k,j_k}$, one can simply replace integrals over Ω by discrete sums, so that the last two terms in the previous expression become

$$\sum_{k=1}^n \sum_{i=1}^{j_k} \int_0^1 \|v_k^i\|_V^2 dt + \sum_{k=1}^n \sum_{i=1}^{j_k} D(\varphi_{01}^{v_k^i} \cdot \bar{m}, m_k \circ g_k^i).$$

Some regularization must then also be added to this objective function to ensure that \bar{m} is smooth as a function of τ . One can, for example, ensure that $\tau \mapsto \bar{m}(\tau)$ is a geodesic on M , or a Riemannian spline as described in the previous section.

13.3.3 Reference-Centered Representations of Time Series

We now focus on methods that place the observed trajectories in a single coordinate system, allowing for the use of a statistical methods designed for linear spaces. This was not done in the previous discussion, which addressed the computation of an average curve.

We first point out that the reference-based representation discussed in Sect. 13.1 is still an option here, in the sense that, given a reference $\bar{m}_0 \in M$, one can still consider a representation of a family of curves $m_1(\cdot), \dots, m_n(\cdot)$ as $v_1(\cdot), \dots, v_n(\cdot)$, where v_1, \dots, v_n are curves in $T_{\bar{m}_0}M$ such that $m_k(\tau) = \text{Exp}_{\bar{m}_0}(v_k(\tau))$ for all τ . This approach, or its registration counterpart in which one computes a collection of diffeomorphisms φ_k^τ for $k = 1, \dots, n$ such that $\varphi_k^\tau \cdot \bar{m}_0 = m_k$, is probably the most commonly used in applications.

However, when using this approach, it is difficult to untangle the part of $v_k(\cdot)$ that describes the evolution within the trajectory from that describing the translation from the reference to that trajectory. Because of this, several methods have been designed that move trajectories as a whole rather than each point individually. More precisely, assume that each trajectory, m_k , has a representation with respect to its own reference, or baseline, \bar{m}_k in the form

$$m_k(\tau) = \text{Exp}_{\bar{m}_k}(v_k(\tau)),$$

with $v_k(\tau) \in T_{\bar{m}_k}M$ for all τ . (If one uses, for example, geodesic regression, then $v_k(\tau) = \tau v_k(1)$.) Given a global reference, \bar{m}_0 , one builds a reference-centered rep-

resentation by “translating” each $v_k(\tau)$ from $T_{\bar{m}_k}M$ to $T_{\bar{m}_0}M$. Notice that, in Euclidean spaces, this operation is trivial, because $v_k(\tau) = m_k(\tau) - \bar{m}_k$ and its translation is just itself!

In Riemannian manifolds, the natural operation for translating tangent vectors is parallel transport, which is described in Sect. B.6.5. This operation must be done along a curve in M connecting the original basis point of the vector that needs to be translated to its target, and the result depends on the chosen curve. When no such curve is specified, it is also natural to choose a minimizing geodesic.

Let, therefore, γ_k be a geodesic such that $\gamma_k(1) = \bar{m}_k$ and $\gamma_k(0) = \bar{m}_0$. The representation of the trajectory m_k in $T_{\bar{m}}M$ is then given by $w_k(\cdot)$ such that $w_k(\tau)$ is the parallel transport of $v_k(\tau)$ along γ_k . After applying this to all curves, we indeed end up with a description of the dataset given by w_1, \dots, w_n , which are all curves in $T_{\bar{m}_0}M$. Parallel transport was introduced for the analysis of manifold data in [163, 176] and for groups of diffeomorphisms and the associated shape spaces in [309], followed by [235, 303, 312].

Using this construction, one therefore represents each trajectory in the form

$$m_k(\tau) = \text{Exp}_{\gamma_k(1)}(\mathcal{T}_{\gamma_k,0,1}w_k(\tau)),$$

where $\mathcal{T}_{\gamma,0,\tau}$ denotes the parallel transport along γ from time 0 to τ . One can also use an alternative approach, proposed in [94] (to which we refer for more details), in which the construction is done in the reverse order. In addition to w_1, \dots, w_n in $T_{\bar{m}}M$, this approach also requires an average curve $\bar{m}(\cdot)$ with $\bar{m}(0) = \bar{m}_0$, and the observed trajectories are represented in the form

$$m_k(\tau) = \text{Exp}_{\bar{m}(\tau)}(\mathcal{T}_{\bar{m},0,\tau}(w_k(\tau))).$$

We conclude this chapter with a description of the parallel transport equations in the diffeomorphism groups and their image via Riemannian submersions. Let, as usual, V be an admissible Hilbert space, and consider the right-invariant metric on Diff defined by $\|\delta\varphi\|_\varphi = \|\delta\varphi \circ \varphi^{-1}\|_V$. One can check (after a lengthy application of Eq. (B.9)) that the Levi-Civita connection on this space is given by

$$\nabla_X Y(\varphi) = \left(\frac{1}{2}(\mathbb{K} \text{ad}_v^* \mathbb{L}w + \mathbb{K} \text{ad}_w^* \mathbb{L}v - \text{ad}_v w) + Xw \right) \circ \varphi,$$

where $v(\varphi) = X(\varphi) \circ \varphi^{-1}$ and $w(\varphi) = Y(\varphi) \circ \varphi^{-1}$ are functions defined on Diff and taking values in V , and $\text{ad}_v w = dvw - dwv$. In particular, if φ depends on time with $\partial_t \varphi = v \circ \varphi$ and $Y(t) = w(t) \circ \varphi(t)$ is a vector field along this curve, then

$$\frac{DY}{Dt} = \left(\frac{1}{2}(\mathbb{K} \text{ad}_v^* \mathbb{L}w + \mathbb{K} \text{ad}_w^* \mathbb{L}v - \text{ad}_v w) + \partial_t w \right) \circ \varphi$$

and parallel transport is equivalent to

$$\partial_t w + \frac{1}{2}(\mathbb{K} \operatorname{ad}_v^* \mathbb{L} w + \mathbb{K} \operatorname{ad}_w^* \mathbb{L} v - \operatorname{ad}_v w) = 0.$$

Taking $v = w$, one retrieves the geodesic equation (EPDiff) given by $\partial_t v + \mathbb{K} \operatorname{ad}_v^* \mathbb{L} v = 0$.

Now consider a shape space M on which Diff acts, such that $\pi(\varphi) = \varphi \cdot m_0$ is a Riemannian submersion (for a fixed $m_0 \in M$). The vertical space at φ is the space $V_m \circ \varphi$, where $V_m = \{v : v \cdot m = 0\}$, and the horizontal space is $H_m \circ \varphi$ with $H_m = V_m^\perp$. The horizontal lift of $\xi \in T_m M$ is the unique vector $v^\xi \in H_m$ such that $v^\xi \cdot m = \xi$ (cf. Sect. 11.5).

If $m(\cdot)$ is a curve on M and $\eta_0 \in T_{m(0)} M$, its parallel transport $\eta(\cdot)$ along m is characterized by

$$\left(\partial_t v^\eta + \frac{1}{2}(\mathbb{K} \operatorname{ad}_{v^\eta}^* \mathbb{L} v^\eta + \mathbb{K} \operatorname{ad}_{v^\eta}^* \mathbb{L} v^\eta - \operatorname{ad}_{v^\eta} v^\eta) \right) \cdot m(t) = 0$$

at all times, with $\xi = \partial_t m$ (this results from Eq. (B.15)). Assume, to simplify the discussion, that M is an open subset of a Banach space Q (otherwise, consider the following computation as valid in a local chart). Writing $v^\eta \cdot m = \eta$, we have $\partial_t \eta = (\partial_t v^\eta) \cdot m + dA_{v^\eta}(m)\xi$, where we have denoted by A_w the mapping $m \mapsto w \cdot m$. Using this, we obtain the parallel transport equation along m

$$\partial_t \eta - dA_{v^\eta}(m)\xi + \left(\frac{1}{2}(\mathbb{K} \operatorname{ad}_{v^\eta}^* \mathbb{L} v^\eta + \mathbb{K} \operatorname{ad}_{v^\eta}^* \mathbb{L} v^\eta - \operatorname{ad}_{v^\eta} v^\eta) \right) \cdot m(t) = 0. \quad (13.6)$$

On shape spaces of point sets, i.e., $m = (x_1, \dots, x_N)$, the infinitesimal action is just $v \cdot m = (v(x_1), \dots, v(x_N))$ and the horizontal lift is such that

$$\mathbb{L} v^\xi = \sum_{k=1}^N \alpha_k^\xi \delta_{x_k},$$

where $(\alpha_1^\xi, \dots, \alpha_N^\xi)$ are obtained by solving the equations $\sum_{j=1}^N K(x_k, x_j) \alpha_j^\xi = \xi_k$. Moreover, we have

$$dA_{v^\eta} \xi = \sum_{k=1}^N (\partial_1 K(x_k, x_j) \xi_k) \alpha_j^\eta.$$

This makes all terms in (13.6) explicit.

The situation is not as simple on spaces of images, in which $\varphi \cdot m = m \circ \varphi^{-1}$ and $v \cdot m = -\nabla m^T v$. One has, in this case, $dA_{v^\eta} \xi = -\nabla \xi^T v$, which is simple, but the horizontal lift of ξ consists of minimizing $\|v\|_V^2$ subject to $\xi = -\nabla m^T v$. While this problem has a unique minimizer, the characterization of this minimizer using Lagrange multipliers requires finding a Banach space W such that $h \mapsto -\nabla m^T h$, from V to W , is bounded and has closed range (see Theorem D.4). This problem is, to our knowledge, still open in the general case.