# Chapter 11
# Distances and Group Actions

## 11.1 General Principles

In this chapter we discuss metric comparisons between deformable objects and their relation to the registration methods that we have studied in the previous chapters. We start with a general discussion on the interplay between distances on a set and transformation groups acting on it.

### *11.1.1 Distance Induced by a Group Action*

Transformation groups acting on sets can help in defining or altering distances on these sets. We will first give a generic construction, based on a least action principle. We will then develop the related differential point of view, when a Lie group acts on a manifold.

A distance on a set $M$ is a mapping $d : M^2 \mapsto [0, +\infty)$ such that: for all $m, m', m'' \in M$,

$$\begin{aligned} &\text{D1.} \quad d(m, m') = 0 \Leftrightarrow m = m', \\ &\text{D2.} \quad d(m.m') = d(m', m), \\ &\text{D3.} \quad d(m, m'') \leq d(m, m') + d(m', m''). \end{aligned}$$

If D1 is not satisfied, but only the fact that $d(m, m) = 0$ for all $m$, one says (still assuming D2 and D3) that $d$ is a *pseudo-distance*.

If $G$ is a group acting on $M$, we will say that a distance $d$ on $M$ is $G$-equivariant if and only if for all $g \in G$, for all $m, m' \in M$, $d(g \cdot m, g \cdot m') = d(m, m')$. A mapping $d : M^2 \mapsto \mathbb{R}_+$ is a $G$-invariant distance if and only if it is a pseudo-distance such that $d(m, m') = 0 \Leftrightarrow \exists g \in G,\ g \cdot m = m'$. This is equivalent to stating that $d$ is a distance on the coset space $M/G$, composed of cosets, or orbits,

 
L. Younes, *Shapes and Diffeomorphisms*, Applied Mathematical Sciences 171,
https://doi.org/10.1007/978-3-662-58496-5_11

$$[m] = \{g \cdot m, g \in G\},$$

with the identification $d([m], [m']) = d(m, m')$. The next proposition shows how a $G$-equivariant distance can induce a $G$-invariant pseudo-distance.

**Proposition 11.1** *Let $d$ be equivariant under the left action of $G$ on $M$. The function $\tilde{d}$, defined by*

$$\tilde{d}([m], [m']) = \inf\{d(g \cdot m, g' \cdot m') : g, g' \in G\}$$

*is a pseudo-distance on $M/G$.*

*If, in addition, the orbits $[m]$ are closed subsets of $M$ (in the topology associated to $d$), then $\tilde{d}$ is a distance.*

Note that, because $d$ is $G$-equivariant, $\tilde{d}$ in the previous proposition is also given by

$$\tilde{d}([m], [m']) = \inf\{d(g \cdot m, m') : g \in G\}.$$

*Proof* The symmetry of $\tilde{d}$ is obvious, as is the fact that $\tilde{d}(([m], [m]) = 0$ for all $m$. For the triangle inequality, D3, it suffices to show that, for all $g_1, g_1', g_2', g_1'' \in G$, there exists $g_2, g_2'' \in G$ such that

$$d(g_2 \cdot m, g_2'' \cdot m'') \leq d(g_1 \cdot m, g_1' \cdot m') + d(g_2' \cdot m', g_1'' \cdot m''). \tag{11.1}$$

Indeed, if this is true, the minimum of the right-hand term in $g_1, g_1', g_2', g_1''$, which is $\tilde{d}([m], [m']) + \tilde{d}([m'], [m''])$, is larger than the minimum of the left-hand term in $g_2, g_2''$, which is $\tilde{d}([m], [m''])$.

To prove (11.1), write $d(g_2' \cdot m', g_1'' \cdot m'') = d(g_1' \cdot m', g_1'(g_2')^{-1} g_1'' \cdot m'')$, take $g_2 = g_1$ and $g_2'' = g_1'(g_2')^{-1} g_1''$; (11.1) is then a consequence of the triangle inequality for $d$.

We now make the additional assumption that the orbits are closed and prove that D1 is true. Take $m, m' \in M$ such that $\tilde{d}([m], [m']) = 0$. This implies that there exists a sequence $(g_n, n \geq 0)$ in $G$ such that $d(g_n \cdot m, m') \to 0$ when $n \to \infty$, so that $m'$ belongs to the closure of the orbit of $m$. Since the latter is assumed to be closed, this yields $m' \in [m]$, which is equivalent to $[m] = [m']$. □

The same statement can clearly be made with $G$ acting on the right on $m$, writing $m \mapsto m \cdot g$. We state it without proof.

**Proposition 11.2** *Let $d$ be equivariant under the right action of $G$ on $M$. The function $\tilde{d}$, defined by*

$$\tilde{d}([m], [m']) = \inf\{d(m \cdot g, m' \cdot g') : g, g' \in G\}$$

*is a pseudo-distance on $G \backslash M$.*

*If, in addition, the orbits $[m]$ are closed subsets of $M$ (in the topology associated to $d$), then $\tilde{d}$ is a distance.*

Here $G \backslash M$ denotes the coset space for the right action of $G$.

### *11.1.2 Distance Altered by a Group Action*

In this section, $G$ is still a group acting on the left on $M$, but we consider the product space $\mathcal{M} = G \times M$ and project on $M$ a distance defined on $\mathcal{M}$. The result of this analysis will be to allow a distance on $\mathcal{M}$ to incorporate a component that accounts for possible group transformations partially accounting for the difference between the compared objects.

The left action of $G$ on $M$ induces a right action of $G$ on $\mathcal{M}$, defined, for $k \in G$, $z = (h, m) \in \mathcal{M}$, by

$$z \cdot k = (hk, k^{-1} \cdot m).$$

For $z = (h, m) \in \mathcal{M}$, we define the projection $\pi(z) = h \cdot m$, taking values in $M$. This projection is constant on the orbits $z \cdot G$ for a given $z$, i.e., for all $k \in G$, $\pi(z \cdot k) = \pi(z)$.

Let $d_{\mathcal{M}}$ be a distance on $\mathcal{M}$. We let, for $m, m' \in M$

$$d(m, m') = \inf\{d_{\mathcal{M}}(z, z') : z, z' \in \mathcal{M}, \pi(z) = m, \pi(z') = m'\}. \tag{11.2}$$

We have the following proposition:

**Proposition 11.3** *If $d_{\mathcal{M}}$ is equivariant by the right action of $G$, then, the function $d$ defined by (11.2) is a pseudo-distance on $M$.*

*If, in addition, the orbits $[z] = \{z \cdot k, k \in G\}$ are closed in $\mathcal{M}$ in the topology associated to $d_{\mathcal{M}}$, then $d$ is a distance.*

This is in fact a corollary of Proposition 11.2. One only has to observe that the quotient space $G\backslash\mathcal{M}$ can be identified with $M$ via the projection $\pi$, and that the distance in (11.2) then becomes the projection distance introduced in Proposition 11.2.

### *11.1.3 Transitive Action*

#### Induced Distance

In this section, we assume that $\mathcal{G}$ is a group that acts transitively on $M$. The action being transitive means that for any $m, m'$ in $M$, there exists an element $z \in \mathcal{G}$ such that $m' = z \cdot m$.

We fix a reference element $m_0$ in $M$, and define the group $G$ by

$$G = \mathrm{Iso}_{m_0}(\mathcal{G}) = \{z \in \mathcal{G}, z \cdot m_0 = m_0\}.$$

This group is the isotropy group, or stabilizer, of $m_0$ in $\mathcal{O}$. We show that $\mathcal{G}$ can be identified with $\mathcal{M} := G \times M$, which will allow us to define a distance in $M$ by projecting a distance on $\mathcal{G}$ as in Sect. 11.1.2.

Assume that a function $\rho : M \to \mathcal{G}$ has been defined, such that for all $m \in M$, $m = \rho(m) \cdot m_0$. This is possible, because the action is transitive (using the axiom of choice). Define

$$\begin{aligned} \Psi : \quad & G \times M \to \mathcal{G} \\ & (h, m) \mapsto \rho(h \cdot m)h. \end{aligned}$$

$\Psi$ is a bijection: if $z \in \mathcal{G}$, we can compute a unique $(h, m)$ such that $z = \Psi(h, m)$; this $(h, m)$ must satisfy

$$z \cdot m_0 = \rho(h \cdot m)h \cdot m_0 = \rho(h \cdot m) \cdot m_0 = h \cdot m,$$

which implies that $\rho(h \cdot m) = \rho(z \cdot m_0)$ and therefore $h = \rho(z \cdot m_0)^{-1}z$, which is uniquely specified; but this also specifies $m = h^{-1}z \cdot m_0$. This proves that $\Psi$ is one-to-one and onto and provides the identification we were looking for.

The right action of $G$ on $\mathcal{M}$, which is $(h, m) \cdot k = (hk, k^{-1} \cdot m)$, translates to $\mathcal{G}$ via $\Psi$ with

$$\Psi((h, m) \cdot k) = \rho(hkk^{-1} \cdot m)hk = \Psi(h, m) \cdot k$$

so that the right actions (of $G$ on $\mathcal{M}$ and of $G$ on $\mathcal{G}$) "commute" with $\Psi$. Finally, the constraint $\pi(h, m_1) = m$ in Proposition 11.3 becomes $z \cdot m_0 = m$ via the identification. All this provides a new version of Proposition 11.3 for transitive actions, given by:

**Corollary 11.4** *Let $d_{\mathcal{G}}$ be a distance on $\mathcal{G}$ which is equivariant under the right action of the isotropy group of $m_0 \in M$. Define, for all $m, m' \in M$,*

$$d(m, m') = \inf\{d_{\mathcal{G}}(z, z') : z \cdot m_0 = m, z' \cdot m_0 = m'\}. \tag{11.3}$$

*Then $d$ is a pseudo-distance on $M$.*

Note that, if $d_{\mathcal{G}}$ is right equivariant under the action of $\mathrm{Iso}_{m_0}(\mathcal{G})$, the distance

$$\tilde{d}_{\mathcal{G}}(z, z') = d_{\mathcal{G}}(z^{-1}, (z')^{-1})$$

is left equivariant, which yields the symmetric version of the previous corollary.

**Corollary 11.5** *Let $d_{\mathcal{G}}$ be a distance on $\mathcal{G}$ which is equivariant under the left action of the isotropy group of $m_0 \in M$. Define, for all $m, m' \in M$,*

$$d(m, m') = \inf\{d_{\mathcal{G}}(z, z') : z \cdot m = m_0, z' \cdot m' = m_0\}. \tag{11.4}$$

*Then $d$ is a pseudo-distance on $M$.*

From Propositions 11.1 and 11.2, $d$ in Corollaries 11.4 and 11.5 is a distance as soon as the orbits $g \cdot \mathrm{Iso}_{m_0}(\mathcal{G})$ (assuming, for example, a left action) are closed for $d_{\mathcal{G}}$. If the left translations $h \mapsto g \cdot h$ are continuous, this is true as soon as $\mathrm{Iso}_{m_0}(\mathcal{G})$

is closed. This last property is itself true as soon as the action $g \mapsto g \cdot m_0$ is continuous, from $G$ to $M$, given some topology on $M$.

Finally, if $d_{\mathcal{G}}$ is left- or right-invariant under the action of the whole group, $\mathcal{G}$, on itself, then the distances in (11.3) and (11.4) both reduce to

$$d(m, m') = \inf\{d_{\mathcal{G}}(\mathrm{id}, z) : z \cdot m = m'\}\,.$$

Indeed, assume right invariance (the left-invariant case is similar): then, if $z \cdot m_0 = m$ and $z' \cdot m_0 = m'$, then $z'z^{-1} \cdot m = m'$ and $d_{\mathcal{G}}(\mathrm{id}, z'z^{-1}) = d_{\mathcal{G}}(z, z')$. Conversely, assume that $\zeta \cdot m = m'$. Since the action is transitive, we know that there exists a $z$ such that $z \cdot m_0 = m$, in which case $\zeta z \cdot m_0 = m'$ and $d_{\mathcal{G}}(\mathrm{id}, \zeta) = d_{\mathcal{G}}(z, \zeta z)$. We summarize this in the following, in which we take $\mathcal{G} = G$:

**Corollary 11.6** *Assume that $G$ acts transitively on $M$. Let $d_G$ be a distance on $G$ that is left or right equivariant. Define, for all $m, m' \in M$,*

$$d(m, m') = \inf\{d_G(\mathrm{id}, g) : g \cdot m = m'\}\,. \tag{11.5}$$

*Then $d$ is a pseudo-distance on $M$.*

### Effort Functionals

As formalized in [135], one can build a distance on $M$ on which a group acts transitively using the notion of effort functionals. The definition we give here is slightly more general than in [135], to take into account a possible influence of the deformed object on the effort. We also make a connection with the previous, distance based, formulations.

We let $\mathcal{G}$ be a group acting transitively on $M$. Assume that a cost $\Gamma(z, m)$ is assigned to a transformation $m \to z \cdot m$. If $m$ and $m'$ are two objects, we define $d(m, m')$ as the minimal cost (effort) required to transform $m$ to $m'$, i.e.,

$$d(m, m') = \inf\{\Gamma(z, m) : z \in \mathcal{G}, z \cdot m = m'\}. \tag{11.6}$$

The proof of the following proposition is almost obvious.

**Proposition 11.7** *If $\Gamma$ satisfies:*

*C1.* $\Gamma(z, m) = 0 \Leftrightarrow z = \mathrm{id}_{\mathcal{G}}$,
*C2.* $\Gamma(z, m) = \Gamma(z^{-1}, z \cdot m)$,
*C3.* $\Gamma(zz', m) \leq \Gamma(z, m) + \Gamma(z', m)$,

*then $d$ defined by (11.6) is a pseudo-distance on $M$.*

In fact, this is equivalent to the construction of Corollary 11.5. To see this, let $G$ be the isotropy group of $m_0$ for the action of $\mathcal{G}$ on $M$. We have the following proposition.

**Proposition 11.8** *If $\Gamma$ satisfies C1, C2 and C3, then, for all $m_0 \in M$, the function $d_{\mathcal{G}}$ defined by*

$$d_{\mathcal{G}}(z, z') = \Gamma(z'z^{-1}, z \cdot m_0) \tag{11.7}$$

*is a distance on $\mathcal{G}$ which is equivariant under the right action of $G$. Conversely, given such a distance $d_{\mathcal{G}}$, one builds an effort functional $\Gamma$ satisfying C1, C2, C3 letting*

$$\Gamma(h, m) = d_{\mathcal{G}}(z, h \cdot z)$$

*where $z$ is any element of $\mathcal{G}$ with the property $z \cdot m = m_0$.*

The proof of this proposition is straightforward and left to the reader.

### *11.1.4 The Riemannian Viewpoint*

The previous sections have demonstrated the usefulness of building distances on a space $\mathcal{M}$ that are equivariant to the actions of a group $G$. Probably the easiest way to construct such a distance (at least when $\mathcal{M}$ is a differential manifold and $G$ is a Lie group) is to design a right-invariant Riemannian metric on $\mathcal{M}$ and use the associated geodesic distance. (See Appendix B.)

Recall that a Riemannian metric on $\mathcal{M}$ requires, for all $z \in \mathcal{M}$, an inner product $\left\langle \cdot\, , \,\cdot \right\rangle_z$ on the tangent space, $T_z\mathcal{M}$, to $\mathcal{M}$ at $z$, which depends smoothly on $z$. With such a metric, one defines the energy of a differentiable path $z(\cdot)$ in $\mathcal{M}$ by

$$E(z(\cdot)) = \int_0^1 \|\partial_t z\|^2_{z(t)}\, dt. \tag{11.8}$$

The associated Riemannian distance on $\mathcal{M}$ is

$$d_{\mathcal{M}}(z_0, z_1) = \inf\{\sqrt{E(z(\cdot))} : z(0) = z_0, z(1) = z_1\}. \tag{11.9}$$

To obtain a right-invariant distance, it suffices to ensure that the metric has this property. For $h \in G$, let $R_h$ denote the right action of $h$ on $\mathcal{M}$: $R_h : z \mapsto z \cdot h$. Let $dR_h(z) : T_z\mathcal{M} \to T_{z\cdot h}\mathcal{M}$ be its differential at $z \in \mathcal{M}$. The right invariance of the metric is expressed by the identity, true for all $z \in \mathcal{M}$, $A \in T_z\mathcal{M}$ and $h \in G$,

$$\|A\|_z = \|dR_h(z) \cdot A\|_{z\cdot h}\,. \tag{11.10}$$

When $\mathcal{M} = G \times M$, condition (11.10) implies that it suffices to define $\left\langle \cdot\, , \,\cdot \right\rangle_z$ at elements $z \in \mathcal{M}$ of the form $z = (\mathrm{id}, m)$ with $m \in M$. The metric at a generic point $(h, m)$ can then be computed, by right invariance, from the metric at $(h, m) \cdot h^{-1} = (\mathrm{id}, h^{-1} \cdot m)$. Because the metric at $(\mathrm{id}, m)$ can be interpreted as a way to attribute a cost to a deformation $(\mathrm{id}, h(t) \cdot m)$ with $h(0) = \mathrm{id}$ and small $t$, defining it corresponds to an analysis of the cost of an infinitesimal perturbation of $m$ by elements of $G$.

Of course, an identical construction could be made with left actions and left-invariant distances.

## 11.2 Invariant Distances Between Point Sets

### *11.2.1 Introduction*

The purpose of this section is to present the construction provided by Kendall [166] on distances between landmarks, taking the infinitesimal point of view that we have just outlined. Here, configurations of landmarks are considered up to similitude (translation, rotation, scaling). Since its introduction, this space has led to a rich literature that specially focuses on statistical data analysis on landmark data. The reader interested in further developments can refer to [91, 167, 266] and to the references therein.

We only consider the two-dimensional case, which is also the simplest. For a fixed integer $N > 0$ let $\mathcal{P}_N$ denote the set of configurations of $N$ points $(z^{(1)}, \ldots, z^{(N)}) \in (\mathbb{R}^2)^N$ such that $z^{(i)} \neq z^{(j)}$ for $i \neq j$. We assume that the order in which the points are listed matters, which means that we consider labeled landmarks. The set $\mathcal{P}_N$ can therefore be identified with an open subset of $\mathbb{R}^{2N}$.

Two configurations $(z^{(1)}, \ldots, z^{(N)})$ and $(\tilde{z}^{(1)}, \ldots, \tilde{z}^{(N)})$ will be identified if one can be deduced from the other by the composition, say $g$, of a translation and a plane similitude, i.e., $\tilde{z}^{(k)} = g \cdot z^{(k)}$ for $k = 1, \ldots, N$. The objects of interest are therefore equivalence classes of landmark configurations, which will be referred to as $N$-shapes.

It will be convenient to identify the plane $\mathbb{R}^2$ with the set of complex numbers $\mathbb{C}$, a point $z = (x, y)$ being represented as $x + iy$. A plane similitude composed with a translation can then be written in the form $z \mapsto az + b$ with $a, b \in \mathbb{C}$, $a \neq 0$.

For $Z = (z^{(1)}, \ldots, z^{(N)}) \in \mathcal{P}_N$, we let $c(Z)$ be the center of inertia

$$c(Z) = (z^{(1)} + \cdots + z^{(N)})/N.$$

We also let $\|Z\|^2 = \sum_{k=1}^{N} |z^{(k)} - c(Z)|^2$.

### *11.2.2 The Space of Planar N-Shapes*

#### Construction of a Distance

Let $\Sigma_N$ be the quotient space of $\mathcal{P}_N$ by the equivalence relation: $Z \sim Z'$ if there exist $a, b \in \mathbb{C}$ such that $Z' = aZ + b$. We denote by $[Z]$ the equivalence class of $Z$ for this relation. We want to define a distance between two equivalence classes $[Z]$ and $[Z']$.

Following Sect. 11.1.4, we define a Riemannian metric on $\mathcal{P}_N$ which is invariant under the action. We therefore must define, for all $Z \in \mathcal{P}_N$, a norm $\|A\|_Z$ over all $A = (a_1, \ldots, a_N) \in \mathbb{C}^N$ such that for all $a, b \in \mathbb{C}$:

$$\|A\|_Z = \|a \cdot A\|_{aZ+b} ,$$

and it suffices to define such a norm for $Z$ such that $\|Z\| = 1$ and $c(Z) = 0$, since we have, for all $Z$,

$$\|A\|_Z = \left\| \frac{A}{\|Z\|} \right\|_{\frac{Z-c(Z)}{\|Z\|}} . \tag{11.11}$$

Once the metric has been chosen, the distance $D(W, Y)$ is defined by

$$D(W, Y)^2 = \inf \int_0^1 \|\partial_t Z\|^2_{Z(t)}\, dt, \tag{11.12}$$

the infimum being taken over all paths $Z(\cdot)$ such that $Z(0) = W$ and $Z(1) = Y$.

When $c(Z) = 0$ and $\|Z\| = 1$, we take

$$\|A\|_Z^2 = \sum_{k=1}^{N} |a^{(k)}|^2.$$

From (11.11) and (11.12), computing $D(W, Y)$ requires us to minimize, among all paths between $W$ and $Y$,

$$\int_0^1 \frac{\sum_{k=1}^N |\partial_t Z^{(k)}|^2}{\sum_{k=1}^N |Z^{(k)}(t) - c(Z(t))|^2} dt.$$

Let $\bar{z}(t) = c(Z(t))$, $v^{(k)}(t) = (Z^{(k)}(t) - C(Z(t)))/\|Z(t)\|$ and $\rho(t) = \|Z(t)\|$. The path $Z(\cdot)$ is uniquely characterized by $(v(\cdot), \rho(\cdot), \bar{z}(\cdot))$. Moreover, we have

$$\partial_t Z^{(k)} = \partial_t \bar{z} + \rho \partial_t v + v \partial_t \rho$$

so that we need to minimize

$$\int_0^1 \sum_{k=1}^{N} \left| \frac{\partial_t \bar{z}}{\rho} + \frac{\partial_t \rho}{\rho} \cdot v^{(k)} + \partial_t v^{(k)} \right|^2 dt.$$

This is equal (using $\sum_k v^{(k)} = 0$ and $\sum_k |v^{(k)}|^2 = 1$, together with the differentials of these expressions) to

$$N\int_0^1 \left(\frac{\partial_t \bar{z}}{\rho}\right)^2 dt + \int_0^1 \left(\frac{\partial_t \rho}{\rho}\right)^2 dt + \int_0^1 \sum_{k=1}^N \left|\partial_t v^{(k)}\right|^2 dt\,. \tag{11.13}$$

The end-point conditions are directly deduced from those initially given in terms of $W$ and $Y$.

The last term in (11.13), which only depends on $v$, can be minimized explicitly, under the constraints $\sum_k v^{(k)} = 0$ and $\sum_k |v^{(k)}|^2 = 1$, which imply that $v$ varies on a $(2N-3)$-dimensional real sphere. The geodesic distance is therefore given by the length of great circles, which yields the expression of the minimum: $\arccos(v(0)^* v(1))^2$, where the "$*$" exponent refers to the conjugate transpose.

Using this, we have

$$\begin{aligned} D(W,Y)^2 = \inf &\left(N\int_0^1 \left(\frac{\partial_t \bar{z}}{\rho}\right)^2 + \int_0^1 \left(\frac{\partial_t \rho}{\rho}\right)^2 dt\right) \\ &+ \arccos\left(\left(\frac{W-c(W)}{\|W\|}\right)^* \left(\frac{Y-c(Y)}{\|Y\|}\right)\right)^2, \end{aligned} \tag{11.14}$$

where the first infimum is over functions $t \mapsto (\bar{z}(t), \rho(t)) \in \mathbb{C} \times [0, +\infty[$, such that $\bar{z}(0) = c(W)$, $\bar{z}(1) = c(Y)$, $\rho(0) = \|W\|$, $\rho(1) = \|Y\|$.

The induced distance on $\Sigma_N$ is then given by

$$d([Y],[W]) = \inf\{D(Y, aW+b), a, b \in \mathbb{C}\}.$$

Writing $a = \lambda h$ with $\lambda = |a| > 0$ and $|h| = 1$. Then

$$\frac{aW+b-c(aW+b)}{\|aW+b\|} = h\frac{W-c(W)}{\|W\|}.$$

Moreover, given $h$, we can take $\lambda = \|Y\|/\|W\|$ and $b = c(Y) - \lambda h c(W)$ so that $\|Y\| = \|aW+b\|$ and $c(Y) = c(aW+b)$, for which the infimum in the right-hand side of (11.14) is zero. We therefore have

$$d([Y],[W]) = \inf_{h:|h|=1}\left(\arccos\left(\left(\frac{W-c(W)}{\|W\|}\right)^* \left(\frac{Y-c(Y)}{\|Y\|}\right)\right)\right).$$

Finally, optimizing this over the unit vector $h$, we get

$$d([Y],[W]) = \arccos\left|\left(\frac{W-c(W)}{\|W\|}\right)^* \left(\frac{Y-c(Y)}{\|Y\|}\right)\right|. \tag{11.15}$$

Denote by $S^{2N-3}$ the set of $v = (v_1, \dots, v_{N-1}) \in \mathbb{C}^{N-1}$ such that $\sum_i |v_i|^2 = 1$ (this can be identified with a real sphere of dimension $2N-3$). The complex projective space, denoted $\mathbb{C}P^{N-2}$, is defined as the space $S^{2N-3}$ quotiented by the equivalence relation: $v \mathcal{R} v'$ if and only if $\exists \nu \in \mathbb{C}$ such that $v' = \nu v$; in other words, $\mathbb{C}P^{N-2}$ contains all sets

$$S^1 \cdot v = \{\nu v, \nu \in \mathbb{C}, |\nu| = 1\}$$

when $v$ varies in $S^{2N-3}$. This set has the structure of an $(N-2)$-dimensional complex manifold, which means that it can be covered with an atlas of open sets that are in bijection with open subsets of $\mathbb{C}^{N-2}$ (with analytic changes of coordinates). Such an atlas is provided, for example, by the family $(\mathcal{O}_k, \Psi_k)$, $k = 1, \ldots, N$, where $\mathcal{O}_k$ is the set of all $S^1 \cdot v \in \mathbb{C}P^{N-2}$ with $v^{(k)} \neq 0$, and

$$\Psi_k(S^1 \cdot v) = (v^{(1)}/v^{(k)}, \ldots, v^{(k-1)}/v^{(k)}, v^{(k+1)}/v^{(k)}, \ldots, v^{(N-1)}/v^{(k)}) \in \mathbb{C}^{N-2}$$

for all $S^1 \cdot v \in \mathcal{O}_i$. In fact, $\Sigma_N$ is also an analytic complex manifold that can be identified with $\mathbb{C}P^{N-2}$.

Let us be more explicit with this identification [166]. Associate to $Z = (z^{(1)}, \ldots, z^{(N)})$ the family $(\zeta^{(1)}, \ldots, \zeta^{(N-1)})$ defined by

$$\zeta^{(k)} = (kz^{(k+1)} - (z^{(1)} + \cdots + z^{(k)}))/\sqrt{k^2 + k}.$$

One can verify that $\sum_{k=1}^{N-1} |\zeta^{(k)}|^2 = \|Z\|^2$ (similar decompositions are used, for example, for the analysis of large-dimensional systems of particles [305]). Denote by $F(Z)$ the element $S^1 \cdot (\zeta/\|Z\|)$ in $\mathbb{C}P^{N-2}$. One can check that $F(Z)$ only depends on $[Z]$ and that $[Z] \mapsto F(Z)$ is an isometry between $\Sigma_N$ and $\mathbb{C}P^{N-2}$.

**The Space of Triangles**

This construction, applied to the case $N = 3$ (which corresponds to triangles with labeled vertices), yields a quite interesting result. For a triangle $Z = (z^{(1)}, z^{(2)}, z^{(3)})$, the previous function $F(Z)$ can be written

$$\begin{aligned} F(Z) &= S^1. \left( \frac{\left[\frac{z^{(2)} - z^{(1)}}{\sqrt{2}}, \frac{2z^{(3)} - z^{(1)} - z^{(2)}}{\sqrt{6}}\right]}{\sqrt{|z^{(2)} - z^{(1)}|^2/2 + |2z^{(3)} - z^{(1)} - z^{(2)}|^2/6}} \right) \\ &= S^1 \cdot [v^{(1)}, v^{(2)}]. \end{aligned}$$

On the set $v^{(1)} \neq 0$ (i.e., the set $z^{(1)} \neq z^{(2)}$) we have the local chart

$$Z \mapsto v^{(2)}/v^{(1)} = \frac{1}{\sqrt{3}} \left( \frac{2z^{(3)} - z^{(2)} - z^{(1)}}{z^{(2)} - z^{(1)}} \right) \in \mathbb{C}.$$

If we let $v^{(2)}/v^{(1)} = \tan \frac{\theta}{2} e^{i\varphi}$, and $M(Z) = (\sin\theta \cos\varphi, \sin\theta \sin\psi, \cos\theta) \in \mathbb{R}^3$, we obtain a correspondence between the triangles and the unit sphere $S^2$.

This correspondence is isometric: the distance between two triangles $[Z]$ and $[\tilde{Z}]$, which has been defined above by

$$d([Z], [\tilde{Z}]) = \arccos \left| \frac{\sum_{k=1}^{3} z^{(k)} \left(\tilde{z}^{(k)}\right)^*}{\|Z\| \, \|\tilde{Z}\|} \right|,$$

gives, after passing to coordinates $\theta$ and $\varphi$, exactly the length of the great circle between the images $M(Z)$ and $M(\tilde{Z})$. We therefore obtain a representation of labeled triangular shapes as points on the sphere $S^2$, with the possibility of comparing them using the standard metric on $S^2$.

## 11.3 Parametrization-Invariant Distances Between Plane Curves

We now describe distances between two-dimensional shapes when they are defined as plane curves modulo changes of parameter. Such distances have been the subject of extensive and detailed mathematical studies, [193, 201], but here we only give an overview of the main ideas and results.

Simple parametrization-free distances can be defined directly from the ranges of the curves. For example, it is possible to use standard norms applied to arc-length parametrizations of the curves, like $L^p$ or Sobolev norms of the difference. With simple closed curves, one can measure the area of the symmetric difference between the interiors of the curves. A more advanced notion, the Hausdorff distance, is defined by

$$d(m, \tilde{m}) = \inf \{\varepsilon > 0, m \subset \tilde{m}^\varepsilon \text{ and } \tilde{m} \subset m^\varepsilon\},$$

where $m^\varepsilon$ is the set of points at a distance less than $\varepsilon$ from $m$ (and similarly for $\tilde{m}^\varepsilon$). The same distance can be used with the interiors for simple closed curves. In fact, the Hausdorff distance is a distance between closed sets as stated in the following proposition.

**Proposition 11.9** *For $\varepsilon > 0$ and a subset $A$ of $\mathbb{R}^d$, let $A^\varepsilon$ be the set of points $x \in \mathbb{R}^d$ such that there exists an $a \in A$ with $|a - x| < \varepsilon$. Let*

$$d_H(A, B) = \inf \{\varepsilon > 0 : A \subset B^\varepsilon \text{ and } B \subset A^\varepsilon\}.$$

*Then $d_H$ is a distance on the set of closed subsets of $\mathbb{R}^d$.*

*Proof* Symmetry is obvious, and we leave to the reader the proof of the triangular inequality, which is a direct consequence of the fact that $(A^\varepsilon)^{\varepsilon'} \subset A^{\varepsilon+\varepsilon'}$.

Assume that $d_H(A, B) = 0$. Then $A \subset B^\varepsilon$ for all $\varepsilon > 0$. But $\bigcap_\varepsilon B^\varepsilon = \bar{B}$, the closure of $B$. We therefore have

$$d_H(A, B) = 0 \Rightarrow A \subset \bar{B} \text{ and } B \subset \bar{A},$$

which implies that $A = B$ if both sets are closed.

One can also proceed similarly to Sect. 11.1.1. First define equivariant distances over parametrized curves, then optimize them with respect to changes of parameters. Let $\mathcal{C}$ be a set of parametrized curves, defined as functions $m : [0, 1] \mapsto \mathbb{R}^2$, subject to additional properties (such as smoothness, closedness, etc.), and $G$ a group of changes of parameter over [0, 1] (including changes of offset for closed curves). Consider the quotient space $\mathcal{S} = \mathcal{C}/G$ for the action $\varphi \cdot m = m \circ \varphi^{-1}$, which are curves modulo change of parameter (one may also want to quotient out rotation, scaling). Based on our discussion in Sect. 11.1.1, a pseudo-distance on $\mathcal{S}$ can be defined from a distance on $\mathcal{C}$ that is equivariant by changes of parameter.

$L^p$ norms between parametrized curves are not equivariant, unless $p = \infty$, with

$$d_\infty(m, \tilde{m}) = \sup_u |m(u) - \tilde{m}(u)|.$$

The distance obtained after reduction by diffeomorphism is called the Fréchet distance, defined by

$$d_F(m, \tilde{m}) = \inf_\varphi d_\infty(m \circ \varphi, \tilde{m}).$$

We note that if, for some diffeomorphism $\varphi$, $d_\infty(m \circ \varphi, \tilde{m}) \le \varepsilon$, then $m \subset \tilde{m}^\varepsilon$ and $\tilde{m} \subset m^\varepsilon$. So we get the relation

$$\varepsilon > d_F(m, \tilde{m}) \Rightarrow \varepsilon > d_H(m, \tilde{m}),$$

which implies $d_H \le d_F$. This and Proposition 11.9 prove that $d_F$ is a distance between curves.

We now consider equivariant distances on $\mathcal{C}$ based on Riemannian metrics derived from invariant norms on the tangent space. We only give an informal discussion, ignoring the complications that arise from the infinite dimension of the space of curves (see [199, 200] for a rigorous presentation). Tangent vectors to $\mathcal{C}$ are derivatives of paths in $\mathcal{C}$, which are time-dependent parametrized curves $t \mapsto m(t, \cdot)$. Tangent vectors therefore take the form $v = \partial_t m(t, \cdot)$, which are functions $v : [0, 1] \to \mathbb{R}^2$. Since a change of parameter in a time-dependent curve induces the same change of parameter on the time derivative, a norm on the tangent space to $\mathcal{C}$ is equivariant under the action of changes of parameter, if, for any $m, v, \varphi$,

$$\|v \circ \varphi^{-1}\|_{m\circ\varphi^{-1}} = \|v\|_m. \tag{11.16}$$

It is therefore sufficient to define $\|v\|_m$ for curves parametrized by arc length, since (11.16) then defines the metric for any parametrized curve.

We now want to define tangent vectors to "plane curves modulo change of parameters." We know that we can modify the tangential component of the time derivative of a time-dependent parametrized curve $t \mapsto m(t, \cdot)$ without changing the geometry of the evolving curve. It follows from this that tangent vectors to $\mathcal{S}$ at a curve $m$ are equivalent classes of vector fields along $m$ that share the same normal component to $m$, and can therefore be identified with this normal component itself, i.e., a scalar function along $m$. The induced metric on $\mathcal{S}$ is

$$\|a\|_m := \inf \left\{ \|v\|_m : v^T N = a \right\}.$$

The associated pseudo-distance on $\mathcal{S}$ is

$$d(m, \tilde{m})^2 = \inf \left\{ \int_0^1 \left\| \partial_t \mu^T N \right\|_{\mu(t)}^2 dt, \mu(0, \cdot) = m, \mu(1, \cdot) = \tilde{m} \right\}. \tag{11.17}$$

The fact that we only get a pseudo-distance in general is interestingly illustrated by the following simple example. Define

$$\|v\|_m^2 = \int_0^{L_m} |v(s)|^2 ds. \tag{11.18}$$

This is the $L^2$ norm in the arc-length parametrization. Then, as stated in the following theorem, we have $d(m, \tilde{m}) = 0$.

**Theorem 11.10** (Mumford–Michor) *The distance defined in* (11.17) *with the norm given by* (11.18) *vanishes between any smooth curves $m$ and $\tilde{m}$.*

A proof of this result can be found in [199, 200]. It relies on the observation that one can grow thin protrusions (“teeth”) on the curve at a cost which is negligible compared to the size of the tooth. It is an easy exercise to compute the geodesic length of a path that starts with a horizontal segment and progressively grows an isosceles triangle of width $\varepsilon$ and height $t$ (at time $t$) on the segment until $t = 1$. This length is $o(\varepsilon)$ (in fact, $O(\varepsilon^2 \ln \varepsilon)$). This implies that one can transform a curve into $O(1/\varepsilon)$ thin non-overlapping teeth at almost no cost. A repeated application of this concept is the basic idea in the construction made in [200] to create almost-zero-length paths between two arbitrary curves.

To prevent the distance from vanishing, one needs to penalize the curve length more than (11.18) does. For example, the distance associated with the metric

$$\|v\|_m^2 = L_m \int_0^{L_m} |v(s)|^2 ds, \tag{11.19}$$

introduced in [193, 259], does not give a degenerate distance on $\mathcal{S}$. The resulting distance is the area swept by the path relating the compared curves [259].

Another way to control degeneracy is to penalize high-curvature points, using for example

$$\|v\|_m^2 = \int_0^{L_m} (1 + a\kappa_m(s)^2)|v(s)|^2 ds. \tag{11.20}$$

This metric has been studied in [200], where it is shown (among other results) that the distance between distinct curves is positive. Finally, one can add derivatives of $v$ (with respect to arc length) in the definition of the metric; this provides *Sobolev metrics* [193, 201] that we have already described for curve evolution.

## 11.4 Invariant Metrics on Diffeomorphisms

We discuss here the construction of a right-invariant distance between diffeomorphisms. We will see, in particular, that it coincides with the direct construction made in Chap. 7.

Here also, we only make an informal (non-rigorous) discussion. We consider a group $G$ of diffeomorphisms of $\Omega$, and define (to fix our ideas) the tangent space to $G$ at $\varphi \in G$ by the set of $u : \Omega \to \mathbb{R}^d$ such that $\mathrm{id} + t\, u \circ \varphi^{-1} \in G$ for small enough $t$. Since the group product on $G$ is the composition, $\varphi\psi = \varphi \circ \psi$, the right translation $R_\varphi : \psi \mapsto \psi \circ \varphi$ is linear, and therefore "equal" to its differential: for $u \in T_\psi G$,

$$dR_\varphi(\psi)u = u \circ \varphi.$$

A metric on $G$ is right-invariant if, for all $\varphi, \psi \in G$ and for all $u \in T_\psi G$,

$$\|dR_\varphi(\psi)u\|_{\psi\circ\varphi} = \|u\|_\psi,$$

which yields, taking $\varphi = \psi^{-1}$:

$$\|u\|_\psi = \|u \circ \psi^{-1}\|_{\mathrm{id}}.$$

This implies that the energy of a path $(t \mapsto \varphi(t, \cdot))$ in $G$ must be defined by

$$E(\varphi(\cdot)) = \int_0^1 \left\|(\partial_t\varphi)(t, \varphi^{-1}(t, \cdot))\right\|_{\mathrm{id}}^2 dt.$$

If we let

$$v(t, x) = (\partial_t\varphi)(t, \varphi^{-1}(t, x)),$$

the energy can be written

$$E(\varphi) = \int_0^1 \|v(t, \cdot)\|_{\mathrm{id}}^2\, dt$$

with the identity

$$\partial_t\varphi(t, x) = v(t, \varphi(t, x)).$$

This implies that $\varphi$ is the flow associated to the velocity field $v(t,\cdot)$. We therefore retrieve the construction given in Chap. 7, with $\|\cdot\|_{\mathrm{id}} = \|\cdot\|_V$. Thus, $V$ in Chap. 7 has a role similar to that of the tangent space to $G$ at id here. Because of this, we let $V = T_{\mathrm{id}}G$ and $\|\cdot\|_{\mathrm{id}} = \|\cdot\|_V$ in the remaining discussion to homogenize the notation.

Assume that $V$ is admissible, according to Definition 7.14. The right-invariant distance on $G$ is

$$d(\varphi_0, \varphi_1) = \inf \sqrt{\int_0^1 \|v(t,\cdot)\|_V^2\, dt}, \tag{11.21}$$

where the minimum is taken over all $v$ such that, for all $x \in \Omega$, the solution of the ordinary differential equation

$$\partial_t y = v(t, y)$$

with initial conditions $y(0) = \varphi_0(x)$ is such that $y(1) = \varphi_1(x)$, consistently with Sect. 7.2.6.

We point out, however, that if $G$ is, say, $\mathrm{Diff}_0^{p,\infty}$, which has its own structure of infinite-dimensional differential manifold, then $V$ is a proper subspace of $T_{\mathrm{id}}G$, resulting in a "sub-Riemannian" metric.

### 11.4.1 The Geodesic Equation

The geodesic equation on $G$ is equivalent to the Euler–Lagrange equation associated to the variational problem (11.21). This is similar to what we have computed in Sect. 10.4, except that here we have a fixed end-point condition. One may address this with a method called the Euler–Poincaré reduction [150, 188], and the presentation we make here is related to it. The energy

$$E(v) = \frac{1}{2}\int_0^1 \|v(t)\|_V^2 dt$$

is minimized over all $v$ such that $\varphi_{01}^v = \varphi_1$ (without loss of generality, because the distance is right invariant, we can assume that $\varphi_0 = \mathrm{id}$).

Applying Theorem D.8 with

$$H_v(\varphi, p) = (p \mid v\circ\varphi) - \frac{\rho}{2}\|v\|_V^2$$

we obtain the fact that, if a trajectory is not "elusive," there exists $\rho \in \{0, 1\}$ and a co-state $p(\cdot)$ taking values in $C_0^p(\mathbb{R}^d, \mathbb{R}^d)^*$ such that

$$\begin{cases} \partial_t\varphi = v\circ\varphi \\ (\partial_t p \mid h) + (p \mid dv\circ\varphi\, h) = 0 \\ \rho v = \xi_\varphi^* p, \end{cases}$$

where $\xi_\varphi : v \mapsto v \circ \varphi$. The first two equations do not depend on whether $\rho = 0$ or 1 (i.e., whether the geodesic is normal or abnormal), and are identical to the momentum conservation equation (10.17) and EPDiff studied in the previous chapter. When $\rho = 1$, the third equation describes $v$, and provides the same necessary conditions for optimality as those found for the diffeomorphic problem in that chapter. Abnormal solutions are such that $\xi_\varphi^* p = 0$ along the trajectory.

Note that the solutions of "soft registration" problems, minimizing

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

for a differentiable function $U$, always provides normal geodesics.

### 11.4.2 A Simple Example

An explicit computation of the geodesic distance is generally impossible, but here is an exception, in one dimension. Take $\Omega = [0, 1]$ and

$$\|u\|_{\mathrm{id}}^2 = \int_0^1 |\partial_x u|^2\, dx.$$

Note that this norm is not admissible, because it cannot be used to control the supremum norm of $\partial_x u$. The associated energy of a path of diffeomorphisms $\varphi(t, \cdot)$ is

$$U(\varphi(\cdot)) = \int_0^1 \int_0^1 \left|\partial_x \left(\varphi_t \circ \varphi^{-1}(t, \cdot)\right)\right|^2 dx dt.$$

This gives, after expanding the derivative and making the change of variables $x = \varphi(t, y)$:

$$U(\varphi(\cdot)) = \int_0^1 \int_0^1 |\partial_t \partial_x \varphi|^2\, |\partial_x \varphi|^{-1}\, dy dt.$$

Define $q(t, y) = \sqrt{\partial_x \varphi(t, y)}$. We have

$$U(\varphi(\cdot)) = 4 \int_0^1 \int_0^1 |\partial_t q|^2\, dy dt,$$

which yields

$$U(\varphi(\cdot)) = 4 \int_0^1 \|\partial_t q(t, \cdot)\|_2^2\, dt.$$

If the problem were to minimize this energy under the constraints $q(0, \cdot) = \sqrt{\partial_x \varphi(0, \cdot)}$ and $q(1, \cdot) = \sqrt{\partial_x \varphi(1, \cdot)}$, the solution $q$ would be given by the line segment

$$q(t, x) = tq(1, x) + (1 - t)q(0, x).$$

There is, however, an additional constraint that comes from the fact that $q(t,\cdot)$ must provide a homeomorphism of [0, 1] for all $t$, which implies $\varphi(t,1)=1$, or, in terms of $q$

$$\|q(t,\cdot)\|_2^2 = \int_0^1 q(t,x)^2 dx = 1.$$

We therefore need to minimize the length of the path $q$ under the constraint that it remains on a Hilbert ($L^2$) sphere. Similar to the finite-dimensional case, geodesics on Hilbert spheres are great circles. This implies that the optimal $q$ is given by

$$q(t,\cdot) = \frac{1}{\sin\alpha}(\sin(\alpha(1-t))q_0 + \sin(\alpha t)q_1)$$

with $\alpha = \arccos\big\langle q_0\,,\ q_1\big\rangle_2$. The length of the geodesic is precisely given by $\alpha$, which provides a closed-form expression of the distance on $G$ [231]

$$d(\varphi,\tilde\varphi) = 2\arccos\int_0^1 \sqrt{\partial_x\varphi\,\partial_x\tilde\varphi}\,dx.$$

### *11.4.3 Gradient Descent*

Assume that a function $\varphi\mapsto U(\varphi)$ is defined over diffeomorphisms. Take $C^1$ $h$ and $\varepsilon_0$ small enough so that $\varphi+\varepsilon h$ is a diffeomorphism if $|\varepsilon|\le\varepsilon_0$, and assume that the Gâteaux derivative $\partial_\varepsilon U(\varphi+\varepsilon h)$ exists at $\varepsilon=0$, denoting it, as in Sect. 9.2, by

$$\partial_\varepsilon U(\varphi+\varepsilon h) = \Big(dU(\varphi)\,\Big|\,h\Big).$$

If a right-invariant metric is given, in the form

$$\big\langle h\,,\ h'\big\rangle_\varphi = \big\langle h\circ\varphi^{-1}\,,\ h'\circ\varphi^{-1}\big\rangle_V$$

as above, the gradient of $U$ at $\varphi$ is computed by identifying

$$\begin{aligned}\Big(dU(\varphi)\,\Big|\,h\Big) &= \big\langle\nabla U(\varphi)\,,\ h\big\rangle_\varphi\\ &= \big\langle\nabla U(\varphi)\circ\varphi^{-1}\,,\ h\circ\varphi^{-1}\big\rangle_V\\ &= \big(\mathbb{L}(\nabla U(\varphi)\circ\varphi^{-1})\ \big|\,h\circ\varphi^{-1}\big),\end{aligned}$$

where $\mathbb{L}=\mathbb{K}^{-1}$ is the duality operator on $V$. Since (with the notation of Sect. 9.2)

$$\Big(dU(\varphi)\,\Big|\,h\Big) = \big(\bar\partial U(\varphi)\ \big|\,h\circ\varphi^{-1}\big),$$

we see that, using $\overline{\nabla}^V U = \mathbb{K}\bar{\partial} U$,

$$\nabla U(\varphi) = \overline{\nabla}^V U(\varphi) \circ \varphi$$

and the evolution equation introduced in (9.7) is nothing but a Riemannian gradient descent for $U$ for the considered metric.

### 11.4.4 Diffeomorphic Active Contours

As a new example of an application of this formalism, we provide a Riemannian version of the active contours algorithm discussed in Sect. 5.4. Let, for a curve $m$,

$$E(m) = \int_m F(p) d\sigma_m + \int_{\Omega_m} \tilde{F}(x) dx. \tag{11.22}$$

We can, fixing a template curve $m_0$, define the functional

$$U(\varphi) = E(\varphi(m_0)).$$

Letting $m = \varphi(m_0)$, a straightforward computation gives

$$\left(\bar{\partial} U(\varphi) \mid v\right) = -\int_m (\kappa F - F^T N + \tilde{F}) v^T N d\sigma_m,$$

from which we deduce

$$\nabla U(\varphi)(x) = -\int_m (\kappa F - F^T N + \tilde{F}) K(\varphi(x), \cdot) N d\sigma_m.$$

This defines the continuous time gradient descent algorithm,

$$\partial_t \varphi(t, x) = \int_{m(t)} (\kappa F - F^T N + \tilde{F}) K(\varphi(t, x), \cdot) N d\sigma_{m(t)}$$

with $m(t) = \varphi(t, \cdot) \circ m_0$.

This algorithm also be expressed as an evolution equation in terms of $m(t)$ only, yielding the *diffeomorphic active contours evolution equation* [21, 310]

$$\partial_t m(t, u) = \int_{m(t)} (\kappa F - F^T N + \tilde{F}) K(m(t, u), \cdot) N d\sigma_{m(t)}. \tag{11.23}$$

A similar discussion can be made for surfaces instead of curves.

Examples of segmentations using this equation are provided in Fig. 11.1.

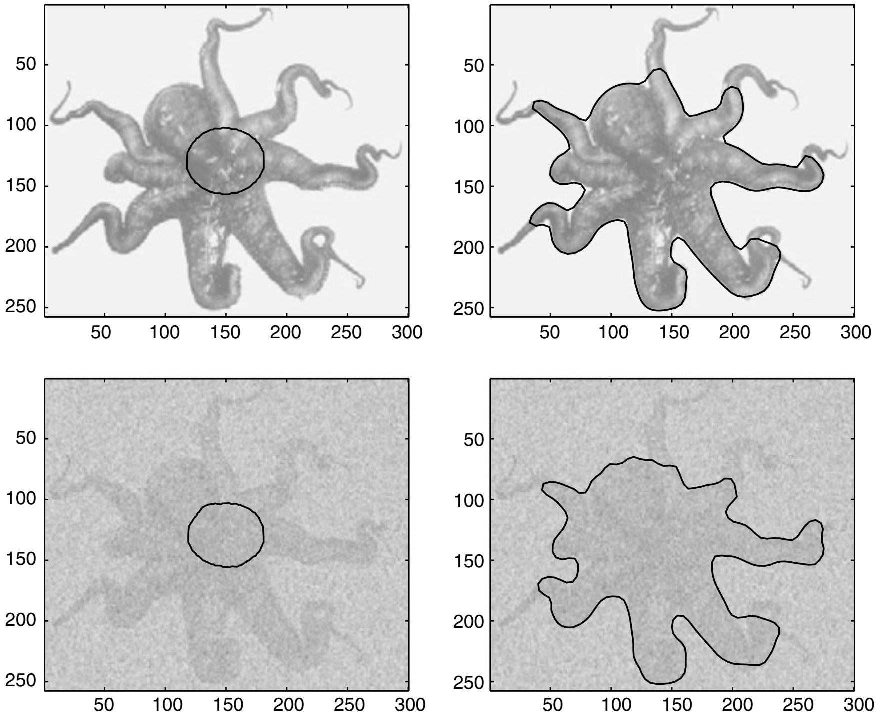


**Fig. 11.1** Diffeomorphic active contours (compare with Fig. 5.4). On each row, the left image is the initial contour, and the right one is the solution obtained with diffeomorphic active contours. The first row presents a clean image and the second a noisy one

## 11.5 Group Actions and Riemannian Submersion

### *11.5.1 Riemannian Submersion*

We temporarily switch to general (but finite-dimensional) Lie groups before returning to diffeomorphisms. Let $\mathcal{G}$ be a Lie group acting transitively (and smoothly) on a manifold, $M$. Fixing a reference element $m_0 \in M$, Corollary 11.5 and Eq. (11.4) show how a distance that is left-equivariant under the action of $G = \mathrm{Iso}_{m_0}(\mathcal{G})$ can be projected to a pseudo-distance on $M$. We now provide the infinitesimal version of this result, which involves the notion of Riemannian submersion discussed in Sect. B.6.7.

Define, as done in Sect. 11.1.3,

$$\begin{aligned} \pi : \quad & \mathcal{G} \to M \\ & g \mapsto g \cdot m_0 \,. \end{aligned}$$

This map is onto because the action is transitive and one can show [142] that it is a submersion, i.e., that $d\pi(g)$ has full rank (the dimension of $M$) for $g \in \mathcal{G}$.

For this submersion, the fiber over $m \in M$ is the set $\pi^{-1}(m) = \{g \in \mathcal{G} : g \cdot m_0 = m\}$. Fix $g \in \pi^{-1}(m)$. Another group element $\tilde{g}$ belongs to $\pi^{-1}(m)$ if and only if $g^{-1}\tilde{g} \in G$, so that $\pi^{-1}(m) = gG$ is a coset. Furthermore [142], the mapping

$$\begin{aligned} [\pi]: \quad \mathcal{G}/G &\to M \\ [g] &\mapsto \pi(g) \end{aligned}$$

is an isomorphism.

Assume that $\mathcal{G}$ is equipped with a Riemannian metric that is right-invariant for the action of $G$, so that, for any $g \in \mathcal{G}$, $h \in G$ and $w \in T_g\mathcal{G}$,

$$\|w\|_g = \|dR_h(g)w\|_{gh}.$$

In other terms $dR_h(g)$ is an isometry between $T_g\mathcal{G}$ and $T_{gh}\mathcal{G}$ (or $R_h$ is a *Riemannian isometry*). Then one can build a Riemannian metric on $M$ such that $\pi$ is a Riemannian submersion. Indeed, if $g, \tilde{g} \in \pi^{-1}(m)$, then there exists an $h \in G$ such that $\tilde{g} = gh$ so that $T_{\tilde{g}}\mathcal{G} = dR_h T_g\mathcal{G}$. Moreover, $\pi(g'h) = \pi(g')$ for all $g' \in \mathcal{G}$ implies that $d\pi(gh)dR_h(g) = d\pi(g)$. This shows that $\mathcal{V}_{gh} = dR_h(g)\mathcal{V}_g$, where

$$\mathcal{V}_g = \left\{w \in T_g\mathcal{G} : d\pi(g)w = 0\right\}$$

is the vertical space at $g$. Letting $\mathcal{H}_g = \mathcal{V}_g^{\perp}$ be the horizontal space at $g$, this and the fact that $dR_h(g)$ is an isometry implies that $dR_h(g)\mathcal{H}_g = \mathcal{H}_{gh}$, so that the restriction of $dR_h(g)$ to the horizontal space provides an isometry between these spaces. This allows us to define, for any $m \in M$ and tangent vector $\xi \in T_mM$:

$$\|\xi\|_m = \|w\|_g$$

for any $g \in \pi^{-1}(m)$, where $w$ is uniquely defined by $d\pi(g)w = \xi$ and $w \in \mathcal{H}_g$. Using the minimizing property of the orthogonal projection, an equivalent definition is that

$$\|\xi\|_m = \min\left\{\|w\|_g : w \in T_g\mathcal{G}, d\pi(g)w = \xi\right\}. \tag{11.24}$$

This is the infinitesimal counterpart of Eq. 11.3.

Using the Lie group structure, this construction can also be analyzed solely on the group's Lie algebra, $\mathfrak{g} = T_{\mathrm{id}}\mathcal{G}$. Notice that, if $\pi(g) = m$ and $\tilde{g}(t)$ is a curve on $\mathcal{G}$ such that $\tilde{g}(0) = g$ and $\partial_t\tilde{g}(0) = w$, then, taking derivatives at $t = 0$,

$$d\pi(g)w = \partial_t(\tilde{g}(t) \cdot m_0) = \partial_t(\tilde{g}(t)g^{-1}) \cdot m = v \cdot m,$$

where $v = dR_{g^{-1}}(g)w$ and $v \cdot m$ refers to the infinitesimal action (cf. Sect. B.5.3). From this, we deduce that, letting

$$V_m = \{v \in \mathfrak{g} : v \cdot m = 0\},$$

one has $\mathcal{V}_g = dR_g(\mathrm{id})V_m$. Moreover, we can rewrite (11.24) as

$$\|\xi\|_m = \min\left\{\|dR_g(\mathrm{id})v\|_g : v \in \mathfrak{g}, v \cdot m = \xi\right\}, \tag{11.25}$$

for any $g \in \pi^{-1}(m)$, with $\mathfrak{g} = T_{\mathrm{id}}\mathcal{G}$. The mapping $v \mapsto \|dR_g(\mathrm{id})v\|_g$ provides a Euclidean norm on $\mathfrak{g}$ that does not depend on which $g$ is chosen in $\pi^{-1}(m)$, therefore only depending on $m$. Denoting this by $\|\cdot\|_m$, $dR_g(\mathrm{id})$ is, by construction, an isometry from $\mathfrak{g}$ to $T_g\mathcal{G}$ that maps $V_m$ onto $\mathcal{V}_g$, and therefore maps $H_m$ onto $\mathcal{H}_g$, where $H_m$ is the space perpendicular to $V_m$ with respect to the dot product associated with $\|\cdot\|_m$ (which we shall denote by $V_m^{\perp_m}$). For $\xi \in T_mM$, there is therefore a unique vector, $v^\xi \in H_m$, such that $v^\xi \cdot m = \xi$ and (11.24) and (11.25) simply become

$$\|\xi\|_m = \|v^\xi\|_{\mathrm{id}} = \min\{\|v\|_m : v \in \mathfrak{g}, v \cdot m = \xi\}. \tag{11.26}$$

Under the stronger assumption that the metric on $\mathcal{G}$ is right-invariant, so that $R_h$ is a Riemannian isometry for all $h \in \mathcal{G}$, we have $\|dR_g(\mathrm{id})v\| = \|v\|_{\mathrm{id}}$ for all $g \in \mathcal{G}$ and $v \in \mathfrak{g}$, so that $\|v\|_m = \|v\|_{\mathrm{id}}$ for all $m \in M$ and one has:

$$\|\xi\|_m = \min\{\|v\|_{\mathrm{id}} : v \cdot m = \xi\}. \tag{11.27}$$

One also defines *horizontal linear forms*, or *horizontal covectors*, which are linear forms $z \in T_{\mathrm{id}}G^*$ such that $(z \mid v) = 0$ for all $v \in \mathcal{V}_m$.

If $\xi \in T_mM$, we have defined $v^\xi$ as the vector in $T_{\mathrm{id}}G$ that minimizes $\|v\|_{\mathrm{id}}$ among all $v$ such that $v \cdot m = \xi$, i.e., the orthogonal projection on $H_m$ of any $v_0^\xi$ such that $v_0^\xi \cdot m = \xi$. This leads to the following definitions, in which we let $v^\xi = h_m(\xi)$.

**Definition 11.11** Let $\mathcal{G}$ be a Lie group acting transitively on a manifold $M$. If $m \in M$ and $\xi \in T_mM$, the horizontal lift of $\xi$ is the vector $h_m(\xi) \in H_m = V_m^{\perp_m}$ such that $h_m(\xi) \cdot m = \xi$.

If $v \in T_{\mathrm{id}}G$, we call $\pi_{\mathcal{H}_m}(v)$ the horizontal part of $v$ at $m$ and $v - \pi_{\mathcal{H}_m}(v)$ its vertical part at $m$, where $\pi_{H_m}$ is the orthogonal projection for $\|\cdot\|_m$, so that

$$\pi_{H_m}(v) = h_m(v \cdot m). \tag{11.28}$$

The projection on $M$ of the Riemannian metric on $G$ is defined by

$$\langle \xi , \eta \rangle_m = \langle h_m(\xi) , h_m(\eta) \rangle_{\mathrm{id}}. \tag{11.29}$$

In the full right-invariant case, geodesics for the projected metric are immediately deduced from those on $\mathcal{G}$, as stated in the following proposition.

**Proposition 11.12** *Assume that the metric on $\mathcal{G}$ is right-invariant. Then the geodesic on $M$ starting at $m$ in the direction $\xi$ is deduced from horizontal geodesics on $\mathcal{G}$ by*

$$\mathrm{Exp}_m(t\xi) = \mathrm{Exp}_{\mathrm{id}}(th_m(\xi)) \cdot m. \tag{11.30}$$

*Proof* This is a direct consequence of Proposition B.27. Let $\mu^*(t) = \mathrm{Exp}_m(t\xi)$ and $\hat{\mu}(t)$ be its horizontal lift starting at some $g \in \pi^{-1}(m)$. Then $\hat{\mu}g^{-1}$ is also a geodesic on $\mathcal{G}$ and necessarily takes the form given in (11.30). □

## 11.5.2 The Momentum Representation

We now apply these general results to groups of diffeomorphisms. While our previous discussion was limited to finite dimensions, some of the concepts introduced there can be generalized to the infinite-dimensional case. As before, we let $V$ be a Hilbert space of vector fields continuously embedded in $C_0^p(\mathbb{R}^d, \mathbb{R}^d)$. To simplify the definition of derivatives, we assume that the shape space $M$ is an open subset of a Banach space $Q$, and that the mapping $\mathcal{A} : \varphi \mapsto \varphi \cdot m$ is differentiable from $\mathrm{Diff}_0^p \times M$ to $Q$. Note that this imposes some restrictions on $M$ and $Q$. For example, if $M$ is the space of $C^q$ embeddings from the unit circle to $\mathbb{R}^2$, then one needs $q \le p$ for the action $\varphi \cdot q = \varphi \circ q$ to take values in $M$, and $q \le p - 1$ to ensure its differentiability.

We will consider the action of $\mathrm{Diff}_V$, the group of attainable diffeomorphisms (Definition 7.15), on $M$. One of our basic assumptions in finite dimensions was the transitivity of the action, which will not hold in general. We will however fix a reference shape $\bar{m}$, and define the space of attainable shapes as the orbit $M_V = \mathrm{Diff}_V \cdot \bar{m}$ of $\bar{m}$ through the action of $\mathrm{Diff}_V$. For $m \in M_V$, we define

$$Q_m = \{v \cdot m : v \in V\} \subset Q$$

and the norm

$$\|\xi\|_m = \min\{\|v\|_V : \xi = v \cdot m\}$$

for $\xi \in Q_m$.

Notice that the infinitesimal action $v \cdot m = \partial_1\mathcal{A}(\mathrm{id}, m)v$ is a bounded linear map from $C_0^p(\mathbb{R}^d, \mathbb{R}^d)$ to $Q$, and so is its restriction to $V$ for $(V, \|\cdot\|_m)$, the Hilbert space topology. This implies that the space

$$V_m = \{v \in V : v \cdot m = 0\}$$

is closed in $V$, and we still denote $V_m^\perp$ by $H_m$. In particular, we have

$$\|\xi\|_m = \|v^\xi\|_V,$$

where $v^\xi = \pi_{H_m}(v)$, for any vector field $v$ satisfying $v \cdot m = \xi$. This implies that the mapping $v \mapsto v \cdot m$ is an isometry between $H_m$ and $Q_m$, which incidentally proves that the latter is a Hilbert space. Given this, we can define the variational problem of minimizing

$$\int_0^1 \|\partial_t m(t)\|^2_{m(t)}\, dt$$

subject to $m(0) = m_0$ and $m(1) = m_1$, $m_0, m_1 \in M_V$ and see that this problem is equivalent to minimizing

$$\int_0^1 \|v(t)\|^2_V\, dt$$

subject to $m(0) = 0, m(1) = m_1$ and $\partial_t m = v \cdot m$. It is also equivalent to minimizing the same energy subject to the constraint $\varphi^v_{01} \cdot m_0 = m_1$.

In finite dimensions, we showed that, if the acting group $\mathcal{G}$ is equipped with a right-invariant Riemannian metric and $M$ with the associated projected metric, then solutions of this problem (which are geodesics in $M$) can be identified with geodesics in $\mathcal{G}$ that start horizontally. So, fixing $m_0 \in M$, the representation $w \mapsto \mathrm{Exp}_{\mathrm{id}}(w) \cdot m_0$ provided a local chart of $M$ around $m_0$ when defined over a neighborhood of 0 in $H_{m_0}$.

With diffeomorphisms, we know that we can find elusive, abnormal and normal geodesics, with only the latter associated with an equation that can be solved given initial conditions $(m_0, w)$. We therefore restrict the "local chart" representation to only those solutions, and express it in terms of co-tangent vectors instead of tangent vectors, which is equivalent in theory, but, as we will see, is much more parsimonious in practice.

As we have seen, this geodesic equation is characterized by momentum conservation, namely

$$\mathbb{L}v(t) = Ad_{\varphi(t)^{-1}}(\mathbb{L}v(0))$$

with $\partial_t \varphi = v \circ \varphi$ and $\mathbb{L}$ is the duality operator of $V$. Given $m \in M_V$, we define the space of horizontal momenta at $m$ simply by $\mathbb{L}H_m \subset V^*$.

**Definition 11.13** Let $D_m$ be the subset of $\mathbb{L}H_m$ consisting of initial momenta for which the conditions of Theorem 10.13 on the existence of solutions of the geodesic equation hold. The momentum representation of a deformable template $\bar{m}$ is the map

$$\begin{aligned} \mathrm{Exp}^\flat_{\bar{m}} : \quad D_m &\to \mathrm{Diff}_V \cdot \bar{m} \\ \rho &\mapsto \mathrm{Exp}_{\bar{m}}(\mathbb{K}\rho) \cdot \bar{m} \end{aligned} \tag{11.31}$$

which associates to a horizontal momentum $\rho$ the position at time 1 of the geodesic initialized at $(\bar{m}, (\mathbb{K}\rho) \cdot \bar{m})$ in $M$.

In finite dimensions, we have proved that horizontality is preserved along geodesics. We retrieve this fact directly in this infinite-dimensional case, as a consequence of the conservation of momentum.

**Proposition 11.14** *Let $\bar{m}$ be a deformable object and $\rho_0 \in D_{\bar{m}}$. Let $(\rho(t), \varphi(t))$ be the evolving momentum and diffeomorphism provided by EPDiff initialized with*

*$\rho(0) = \rho_0$. Let $m(t) = \varphi(t) \cdot \bar{m}$ be the evolution of the deformable object. Then, at all times $t$, $\rho(t) \in \mathbb{L}H_{m(t)}$.*

*Proof* We will prove that if $w \in V_m$ and $\varphi$ a diffeomorphism, then $\mathrm{Ad}_\varphi w \in V_{\varphi \cdot m}$.

Before proving this fact, we verify that it implies the proposition, which requires us to check that $(w \in V_{m(t)}) \Rightarrow ((\rho(t) \mid w) = 0)$. From the conservation of momentum we have

$$(\rho(t) \mid w) = \left(\rho_0 \,\middle|\, \mathrm{Ad}_{\varphi(t)^{-1}} w\right)$$

and $\mathrm{Ad}_{\varphi(t)^{-1}} w \in V_{\bar{m}}$ if $w \in \mathcal{V}_{m(t)}$, which implies that $(\rho(t) \mid w) = 0$.

We now prove our claim. Let $\psi(\varepsilon)$ be such that $\partial_\varepsilon \psi(0) = w$, and $\partial_\varepsilon(\psi \cdot m)(0) = 0$ at $\varepsilon = 0$. By definition, $\mathrm{Ad}_\varphi w = \partial_\varepsilon(\varphi \circ \psi \circ \varphi^{-1})(0)$. But we have

$$\partial_\varepsilon(\varphi \circ \psi \circ \varphi^{-1}) \cdot (\varphi \cdot m) = dA_\varphi(\psi \cdot m)\, \partial_\epsilon(\psi \cdot m) = 0,$$

which implies that $\mathrm{Ad}_\varphi w \in V_{\varphi \cdot m}$. (Here, $A_\varphi$ denotes the action $A_\varphi : m \mapsto \varphi \cdot m$.) □

We now describe horizontal momenta in a few special cases. First assume that deformable objects are point sets, so that

$$M = \left\{(x_1, \ldots, x_N) \in (\mathbb{R}^d)^N, x_i \neq x_j \text{ for } i \neq j\right\}$$

and $Q = (\mathbb{R}^d)^N$.

If $m = (x_1, \ldots, x_N)$, we have

$$V_m = \{v \in V : v(x_1) = \cdots = v(x_N) = 0\}.$$

Letting $e_1, \ldots, e_d$ be the canonical basis of $\mathbb{R}^d$, $V_m$ is therefore defined as the set of $v$'s such that $\left(e_j \delta_{x_k} \,\middle|\, v\right) = 0$ for all $j = 1, \ldots, d$ and $k = 1, \ldots N$. So $V_m = W^\perp$, where $W$ is the vector space generated by the $d \times N$ vector fields $K(\cdot, x_k)e_j$. Because $W$ is finite-dimensional, it is closed and $H_m = V_m^\perp = (W^\perp)^\perp = W$. Switching to momenta, we obtain the fact that, for point sets $m = (x_1, \ldots, x_N)$

$$\mathbb{L}H_m = \left\{\sum_{k=1}^N z_k \delta_{x_k}, z_1, \ldots, z_N \in \mathbb{R}^d\right\}.$$

In particular, we see that the momentum representation is parametrized by the $Nd$-dimensional set $(z_1, \ldots, z_N)$ and therefore has the same dimension as the considered objects. Finally, we note that, in this finite-dimensional shape space, one has $M_V = M$ and $\| \cdot \|_m$ provides a Riemannian metric on this space.

The description of $V_m$ is still valid when $m$ is a general parametrized subset of $\mathbb{R}^d$: $m : u \mapsto m(u) = x_u \in \mathbb{R}^d$, defined for $u$ in a, so far, arbitrary set $U$. Then

$$V_m = \{v \in V : v(x_u) = 0, u \in U\} \tag{11.32}$$

and we still have $V_m = W^\perp$, where $W$ is the vector space generated by the vector fields $K(\cdot, x_u)e_j$, $j = 1, \ldots, d$, $u \in U$. The difference, now, is that $W$ is not finite-dimensional if $m$ is infinite, and not necessarily a closed subspace of $V$, so that

$$H_m = (W^\perp)^\perp = \bar{W},$$

the closure of $W$ in $V$. Turning to the momenta, this says that

$$\mathbb{L}H_m = \overline{\left\{\sum_{k=1}^{n} z_k \delta_{x_{u_k}}, n \geq 0, z_1, \ldots, z_n \in \mathbb{R}^d, u_1, \ldots, u_n \in U\right\}},$$

where the closure is now in $V^*$.

This argument applies to parametrized curves and surfaces, but must be adapted for geometric objects, that is, curves and surfaces seen modulo a change of parametrization. In this case, deformable objects are equivalent classes of parametrized manifolds. One way to address this is to use infinite-dimensional local charts that describe the equivalence classes in a neighborhood of a given object $m$. We will not detail this rigorously here, but the interested reader can refer to [201] for such a construction with plane curves.

Intuitively, however, the resulting description of $V_m$ is clear. In contrast to the parametrized case, for which vector fields in $V_m$ were not allowed to move any point in $m$, it is now possible to do so, provided the motion happens within $m$, i.e., the vector fields are tangent to $m$. This leads to the following set:

$$V_m = \{v \in V : v(x) \text{ is tangent to } m \text{ for all } x \in m\}.$$

Since $v(x)$ being tangent to $m$ is equivalent to $N^T v(x) = 0$ for all $N$ normal to $m$ at $x$, we see that $V_m = W^\perp$, where $W$ is the vector space generated by vector fields $K(\cdot, x)N$, with $x \in m$ and $N$ normal to $m$ at $x$. Again, this implies that $H_m = \bar{W}$ and that

$$\mathbb{L}H_m = \overline{\left\{\sum_{k=1}^{n} z_k \delta_{x_k}, n \geq 0, x_1, \ldots, x_n \in m, z_1, \ldots, z_n \in N_{x_k} m\right\}},$$

where $N_x m$ is the set of vectors that are normal to $m$ at $x$.

Now, consider the example of smooth scalar functions (or images): $m : \mathbb{R}^d \to \mathbb{R}$. In this case, the action being $\varphi \cdot m = m \circ \varphi^{-1}$, the set $V_m$ is

$$V_m = \left\{v \in V : \nabla m^T v = 0\right\},$$

which directly implies that $V_m = W^\perp$, where $W$ is the vector space generated by $K(\cdot, x)\nabla m(x)$ for $x \in \mathbb{R}^d$. Horizontal momenta therefore span the set

$$\mathbb{L}H_m = \overline{\left\{ \sum_{k=1}^{n} \nabla m(x_k)\delta_{x_k}, n \geq 0, x_1, \ldots, x_n \in \mathbb{R}^d \right\}}.$$

We conclude with the action of diffeomorphisms on measures, for which:

$$(\varphi \cdot m \mid f) = (m \mid f \circ \varphi),$$

so that $v \in V_m$ if and only if $\left(m \;\middle|\; \nabla f^T v\right) = 0$ for all smooth $f$. So $V_m = W^{\perp}$, where

$$W = \left\{\mathbb{K}(\nabla f\, m),\, f \in C^1(\Omega, \mathbb{R})\right\}$$

so that

$$\mathbb{L}H_m = \overline{\left\{\nabla f\, m,\, f \in C^1(\Omega, \mathbb{R})\right\}}.$$

We point out that the horizontal spaces may be much larger than one would expect by formally extending the case of point sets. For example, if $V$ is a Gaussian RKHS and $m$ is a set with non-empty interior, then $V_m = \{0\}$ in (11.32) and $H_m = V$! This is because Gaussian RKHSs only contain analytic functions [207, 268]. For the same reason, if $m$ is a curve that contain a line segment, all vector fields in $V_m$ must vanish on the whole line containing the segment.

This behavior cannot happen when $V$ is a space containing all compactly-supported smooth functions, such as Sobolev spaces. In this case, if $m$ is a closed subset of $\mathbb{R}^d$, then any smooth vector field with support in $\mathbb{R}^d \setminus m$ must belong to $V_m$ and any $\rho \in \mathbb{L}H_m$ must therefore vanish on such functions, which shows that $\rho$ (as a generalized function) is supported by $m$.

However, even in such contexts, an explicit description of horizontal momenta is generally beyond reach, and one generally restrict the momentum representation to more "manageable" subsets of $H_m$, using, for example, measure momenta supported by $m$, as considered in Sect. 10.5.6. As we have seen, such measure momenta cover most of the cases of interest for diffeomorphic matching with a differentiable end-point cost, even when using Gaussian kernels.

The momentum representation provides a diffeomorphic version of the deformable template approach described for polygons in Sect. 6.3. As we have seen, it can be applied to a wide class of deformable objects. Applications to datasets of three-dimensional medical images can be found in [143, 236, 290, 300].