Shape deformation analysis from the optimal control viewpoint

Sylvain Arguillère*   Emmanuel Trélat†   Alain Trouvé‡   Laurent Younes§

Abstract

A crucial problem in shape deformation analysis is to determine a deformation of a given shape into another one, which is optimal for a certain cost. It has a number of applications in particular in medical imaging.

In this article we provide a new general approach to shape deformation analysis, within the framework of optimal control theory, in which a deformation is represented as the flow of diffeomorphisms generated by time-dependent vector fields. Using reproducing kernel Hilbert spaces of vector fields, the general shape deformation analysis problem is specified as an infinite-dimensional optimal control problem with state and control constraints. In this problem, the states are diffeomorphisms and the controls are vector fields, both of them being subject to some constraints. The functional to be minimized is the sum of a first term defined as geometric norm of the control (kinetic energy of the deformation) and of a data-attachment term providing a geometric distance to the target shape.

This point of view has several advantages. First, it allows one to model general constrained shape analysis problems, which opens new issues in this field. Second, using an extension of the Pontryagin maximum principle, one can characterize the optimal solutions of the shape deformation problem in a very general way as the solutions of constrained geodesic equations. Finally, recasting general algorithms of optimal control into shape analysis yields new efficient numerical methods in shape deformation analysis. Overall, the optimal control point of view unifies and generalizes different theoretical and numerical approaches to shape deformation problems, and also allows us to design new approaches.

The optimal control problems that result from this construction are infinite dimensional and involve some constraints, and thus are nonstandard. In this article we also provide a rigorous and complete analysis of the infinite-dimensional shape space problem with constraints and of its finite-dimensional approximations.

Keywords: shape deformation analysis, optimal control, reproducing kernel Hilbert spaces, Pontryagin maximum principle, geodesic equations.

AMS classification: 58E99 49Q10 46E22 49J15 62H35 53C22 58D05

---

*Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France (`sylvain.arguillere@upmc.fr`).

†Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, Institut Universitaire de France, F-75005, Paris, France (`emmanuel.trelat@upmc.fr`).

‡Ecole Normale Supérieure de Cachan, Centre de Mathématiques et Leurs Applications, CMLA, 61 av. du Pdt Wilson, F-94235 Cachan Cedex, France (`trouve@cmla.ens-cachan.fr`).

§Johns Hopkins University, Center for Imaging Science, Department of Applied Mathematics and Statistics, Clark 324C, 3400 N. Charles st. Baltimore, MD 21218, USA (`laurent.younes@jhu.edu`).

Contents

1   Introduction

The mathematical analysis of shapes has become a subject of growing interest in the past few decades, and has motivated the development of efficient image acquisition and segmentation methods, with applications to many domains, including computational anatomy and object recognition.

The general purpose of shape analysis is to compare two (or more) shapes in a way that takes into account their geometric properties. Two shapes can be very similar from a human's point of view, like a circle and an ellipse, but very different from a computer's automated perspective. In *Shape Deformation Analysis*, one optimizes a deformation mapping one shape onto the other and bases the analysis on its properties. This of course implies that a cost has been assigned to every possible deformation of a shape, the design of this cost function being a crucial step in the method. This approach has been used extensively in the analysis of anatomical organs from medical images (see [22]).

In this framework, a powerful and convenient approach represents deformations as flows of diffeomorphisms generated by time-dependent vector fields [17, 36, 37]. Indeed, when considering the studied shapes as embedded in a real vector space $\mathbb{R}^d$, deformations of the whole space, like diffeomorphisms, induce deformations of the shape itself. The set of all possible deformations is then defined as the set of flows of time-dependent vector fields of a Hilbert space V, called space of "infinitesimal transformations", which is a subset of the space of all smooth bounded vector fields on $\mathbb{R}^d$.

This point of view has several interesting features, not the least of which being that the space of possible deformations is a well-defined subgroup of the group of diffeomorphisms, equipped with a structure similar to the one of a right-invariant sub-Riemannian metric [12, 31]. This framework has led to the development of a family of registration algorithms called *Large Deformation Diffeomorphic Metric Mapping* (LDDMM), in which the correspondence between two shapes comes from the minimization of an objective functional defined as a sum of two terms [9, 8, 10, 11, 13, 26, 29, 30, 34]. The first term takes into account the cost of the deformation, defined as the integral of the squared norm of the time-dependent vector field from which it arises. In a way, it is the total kinetic energy of the deformation. The second term is a data attachment penalizing the difference between the deformed shape and a target.

An appropriate class of Hilbert spaces of vector fields for V is the one of *reproducing kernel Hilbert spaces* (in short, RKHS) [7], because they provide very simple solutions to the spline interpolation problem when the shape is given by a set of landmarks [40, 42], which is an important special case since it includes most practical situations after discretization. This framework allows one to use tools from Riemannian geometry [40], along with classical results from the theory of Lie groups equipped with right-invariant metrics[5, 6, 24, 28, 42]. These existing approaches provide an account for some of the geometric information in the shape, like singularities for example. However, they do not consider other intrinsic properties of the studied shape, which can also depend on the nature of the object represented by the shape. For example, for landmarks representing articulations of a robotic arm, the deformation can be searched so as to preserve the distance between certain landmarks. For cardiac motions, it may be relevant to consider deformations of the shape assuming that the movement only comes from a force applied only along the fiber structure of the muscle. In other words, it may be interesting to constrain the possible deformations (by considering non-holonomic constraints) in order to better fit the model.

As a motivation, consider the following example from computational anatomy, in which one wants to study together different structures in the brain. In Figures 1(a) three shapes are delineated, and need to be matched to the corresponding shapes in the target, Figure 1(b).

Without adding constraints, there are two natural ways to model this problem. The first consists of finding a diffeomorphism that simultaneously maps the three initial shapes to the

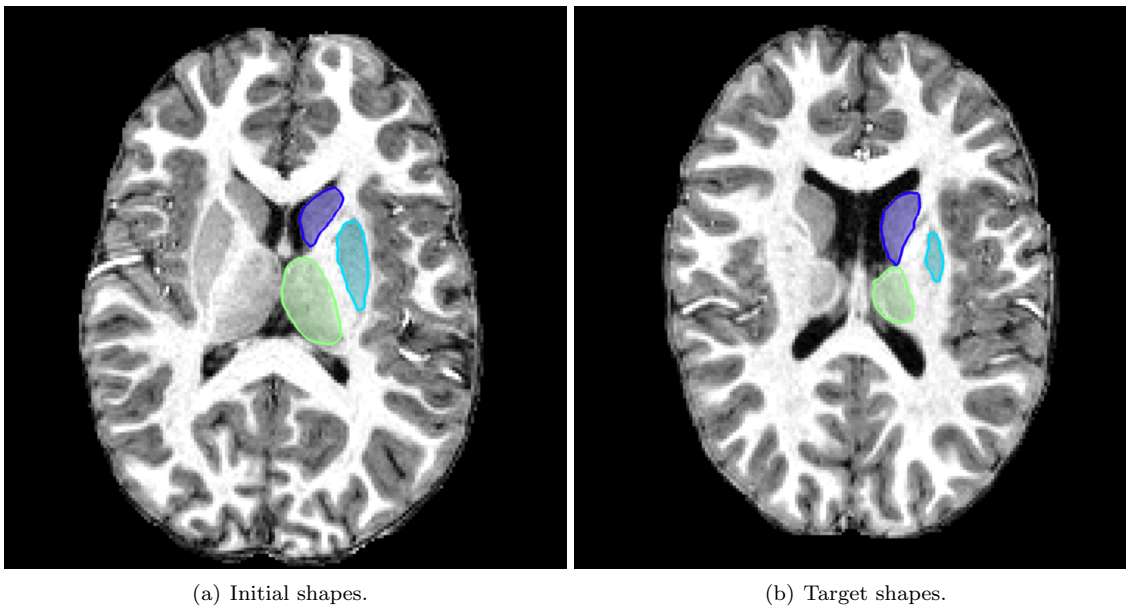(a) Initial shapes.   (b) Target shapes.

Figure 1: Matching several shapes at once. Caudate (dark blue), Putamen (cyan) and Thalamus (green).

corresponding target. Such a model is however not much appropriate since, physically, the sub-organs are essentially disconnected and may move independently from each other. Some may be close to each other in an image an distant in another one, which creates very large deformations when the whole system is matched with a single diffeomorphism, creating artificial distortions and numerical challenges. The second possibility is then to consider all shapes independently, each shape being deformed by a diffeomorphism of the plane. This solves the previous problems, but raises a new one: indeed, in this second model, two shapes are allowed to overlap, which is not acceptable.

To solve these issues, we can use a fourth diffeomorphism that will model the background in which the structures are embedded. This "background deformation" will be applied to the complement of the three shapes on the plane. Then, one adds the constraints that the boundary of the shapes and that of the background should either move together (stitching constraints) or slide alongside (sliding constraints). Since the background is deformed by a diffeomorphism, this also prevents the three shapes from overlapping. This is a basic example of *multishapes*, which will be introduced in Section 2.3.

In order to take into account such (and more general) constraints in shape deformation problems, we propose to model these problems within the framework of *optimal control theory*, where the control system models the evolution of the deformation and the control is the time-dependent vector field (see preliminary ideas in [40]). The point of view of optimal control has already been adopted in shape analysis without constraints in [9, 10], in the particular case of landmarks and surfaces. It should be noted that this viewpoint is also closely related with Clebsch optimal control (see [19]), in which the control system is defined by the infinitesimal action of a group.

The purpose of this paper is to develop the general point of view of *optimal control for shape deformation analysis* as comprehensively as possible. We will show the relevance of this framework,

in particular because it can be used to model constrained shapes among many other applications.

Indeed, a lot of tools have been developed in control theory for solving optimal control problems with or without constraints. The well-known *Pontryagin maximum principle* (in short PMP, see [33]) provides first-order conditions for optimality in the form of Hamiltonian extremal equations with a maximization condition permitting the computation of the optimal control. It has been generalized in many ways, and a large number of variants or improvements have been made over the past decades, with particular efforts in order to be able to address optimal control problems involving general state/control constraints (see the survey article [23] and the many references therein). The analysis is, however, mainly done in finite dimension. Since shape analysis has a natural setting in infinite dimension (indeed, in 2D, the shape space is typically a space of smooth curves in $\mathbb{R}^2$), we need to derive an appropriate infinite-dimensional variant of the PMP for constrained problems. Such a variant is nontrivial and nonstandard, given that our constrained shape analysis problems generally involve an infinite number of equality constraints.

Such a PMP will allow us to derive in a rigorous geometric setting the (constrained) geodesic equations that must be satisfied by the optimal deformations.

Moreover, modeling shape deformation problems within the framework of optimal control theory can inherit from the many numerical methods that exist in this context and thus lead to new algorithms in shape analysis.

The paper is organized as follows.

Section 2 is devoted to modeling shape deformation problems with optimal control. We first briefly describe, in Section 2.1, the framework of diffeomorphic deformations arising from the integration of time-dependent vector fields belonging to a given RKHS, and recall some properties of RKHS's of vector fields. In Section 2.2 we introduce the action of diffeomorphisms on a shape space, and we model and define the optimal control problem on diffeomorphisms which is at the heart of the present study, where the control system stands for the evolving deformation and the minimization runs over all possible time-dependent vector fields attached to a given RKHS and satisfying some constraints. We prove that, under weak assumptions, this problem is well posed and has at least one solution (Theorem 1). Since the RKHS is in general only known through its kernel, we then provide a kernel formulation of the optimal control problem and we analyze the equivalence between both problems. In Section 2.3 we investigate in our framework two important variants of shape spaces, which are lifted shapes and multi-shapes. Section 2.4 is devoted to the study of finite-dimensional approximations of the optimal control problem. Section 2.5 contains a proof of Theorem 1.

Section 3 is dedicated to the derivation of the constrained geodesic equations in shape spaces, that must be satisfied by optimal deformations. We first establish in Section 3.1 an infinite-dimensional variant of the PMP which is adapted to our setting (Theorem 2). As an application, we derive in Section 3.2 the geodesic equations in shape spaces (Theorem 3), in a geometric setting, and show that they can be written as a Hamiltonian system.

In Section 4, we design some algorithms in order to solve the optimal control problem modeling the shape deformation problem. Problems without constraints are first analyzed in Section 4.1, and we recover some already known algorithms used in unconstrained shape spaces, however with a more general point of view. We are thus able to extend and generalize existing methods. Problems with constraints are investigated in Section 4.2 in view of solving constrained matching problems. We analyze in particular the augmented Lagrangian algorithm, and we also design a method based on shooting.

In Section 5 we provide numerical examples, investigating first a matching problem with constant total volume, and then a multishape matching problem. In addition, we will come back to the matching problem of Figure 1.

2 Modelling shape deformation problems with optimal control

The following notation will be used throughout the paper. Let $d \in \mathbb{N}$ fixed. A vector $a \in \mathbb{R}^d$ can be as well viewed as a column matrix of length d. The Euclidean norm of a is denoted by $|a|$. The inner product $a \cdot b$ between two vectors $a, b \in \mathbb{R}^d$ can as well be written, with matrix notations, as $a^T b$, where a^T is the transpose of a. In particular one has $|a|^2 = a \cdot a = a^T a$.

Let X be a Banach space. The norm on X is denoted by $\|\cdot\|_X$, and the inner product by $(\cdot, \cdot)_X$ whenever X is a Hilbert space. The topological dual X^* of X is defined as the set of all linear continuous mappings $p : X \to \mathbb{R}$. Endowed with the usual dual norm $\|p\|_{X^*} = \sup\{p(x) \mid x \in X,\ \|x\|_X = 1\}$, it is a Banach space. For $p \in X^*$, the natural pairing between p and $w \in X$ is $p(w) = \langle p, w \rangle_{X^*, X}$, with the duality bracket. If $X = \mathbb{R}^n$ then p can be identified with a column vector through the equality $p(w) = p^T w$.

Let M be an open subset of X, and let Y be another Banach space. The Fréchet derivative of a map $f : M \to Y$ at a point $q \in M$ is written as df_q. When it is applied to a vector w, it is denoted by $df_q.w$ or $df_q(w)$. When $Y = \mathbb{R}$, we may also write $\langle df_q, w \rangle_{X^*, X}$.

We denote by $W^{1,p}(0, 1; M)$ (resp. $H^1(0, 1; M)$) the usual Sobolev space of elements of $L^p(0, 1; M)$, with $1 \leqslant p \leqslant +\infty$ (resp., with $p = 2$) having a weak derivative in $L^p(0, 1; X)$. For $q_0 \in M$ we denote by $W^{1,p}_{q_0}(0, 1; M)$ (resp., by $H^1_{q_0}(0, 1; M)$) the space of all $q \in W^{1,p}(0, 1; M)$ (resp., $q \in H^1(0, 1; M)$) such that $q(0) = q_0$.

For every $\ell \in \mathbb{N}$, a mapping $\varphi : M \to M$ is called a $\mathcal{C}^\ell$ diffeomorphism if it is a bijective mapping of class $\mathcal{C}^\ell$ with an inverse of class $\mathcal{C}^\ell$. The space of all such diffeomorphisms is denoted by $\mathrm{Diff}^\ell(M)$. Note that $\mathrm{Diff}^0(M)$ is the space of all homeomorphisms of M.

For every mapping $f : \mathbb{R}^d \to X$ of class $\mathcal{C}^\ell$ with compact support, we define the usual semi-norm

$$\|f\|_\ell = \sup \left\{ \left\| \frac{\partial^{\ell_1 + \cdots + \ell_d} f(x)}{\partial x_1^{\ell_1} \ldots \partial x_d^{\ell_d}} \right\|_X \ \middle| \ x \in \mathbb{R}^d,\ (\ell_1, \ldots, \ell_d) \in \mathbb{N}^d,\ \ell_1 + \cdots + \ell_d \leqslant \ell \right\}.$$

We define the Banach space $\mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$ (endowed with the norm $\|\cdot\|_\ell$) as the completion of the space of vector fields of class $\mathcal{C}^\ell$ with compact support on $\mathbb{R}^d$ with respect to the norm $\|\cdot\|_\ell$. In other words, $\mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$ is the space of vector fields of class $\mathcal{C}^\ell$ on $\mathbb{R}^d$ whose derivatives of order less than or equal to ℓ converge to zero at infinity.

We define $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ as the set of all diffeomorphisms of class $\mathcal{C}^\ell$ that converge to identity at infinity. Clearly, $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ is the set of all $\varphi \in \mathrm{Diff}^\ell(\mathbb{R}^d)$ such that $\varphi - \mathrm{Id}_{\mathbb{R}^d} \in \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$. It is a group for the composition law $(\varphi, \psi) \mapsto \varphi \circ \psi$.

Note that, if $\ell \geqslant 1$, then $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ is an open subset of the affine Banach space $\mathrm{Id}_{\mathbb{R}^d} + \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$. This allows one to develop a differential calculus on $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$.

2.1 Preliminaries: deformations and RKHS of vector fields

Our approach to shape analysis is based on optimizing evolving deformations. A deformation is a one-parameter family of flows in $\mathbb{R}^d$ generated by time-dependent vector fields on $\mathbb{R}^d$. Let us define this concept more rigorously.

Diffeomorphic deformations. Let $\ell \in \mathbb{N}^*$. Let

$$
\begin{array}{rcl}
v : [0, 1] & \longrightarrow & \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d) \\
t & \longmapsto & v(t)
\end{array}
$$

be a time-dependent vector field such that the real-valued function $t \mapsto \|v(t)\|_\ell$ is integrable. In other words, we consider an element v of the space $L^1(0, 1; \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d))$.

According to the Cauchy-Lipshitz theorem, v generates a (unique) flow $\varphi : [0, 1] \to \mathrm{Diff}_0^\ell(\mathbb{R}^d)$ (see, e.g., [1, 42] or [35, Chapter 11]), that is a one-parameter family of diffeomorphisms such that

$$\frac{\partial \varphi}{\partial t}(t, x) = v(t, \varphi(t, x)),$$
$$\varphi(0, x) = x,$$

for almost every $t \in [0, 1]$ and every $x \in \mathbb{R}^d$. In other words, considering φ as a curve in the space $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$, the flow φ is the unique solution of

$$\dot{\varphi}(t) = v(t) \circ \varphi(t),$$
$$\varphi(0) = \mathrm{Id}_{\mathbb{R}^d}. \tag{1}$$

Such a flow φ is called a *deformation* of $\mathbb{R}^d$ of class $\mathcal{C}^\ell$.

Proposition 1. *The set of deformations of $\mathbb{R}^d$ of class $\mathcal{C}^\ell$ coincides with the set*

$$\left\{ \varphi \in W^{1,1}(0, 1; \mathrm{Diff}_0^\ell(\mathbb{R}^d)) \mid \varphi(0) = \mathrm{Id}_{\mathbb{R}^d} \right\}.$$

In other words, the deformations of $\mathbb{R}^d$ of class $\mathcal{C}^\ell$ are exactly the curves $t \mapsto \varphi(t)$ on $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ that are integrable on $(0, 1)$ as well as their derivative, such that $\varphi(0) = \mathrm{Id}_{\mathbb{R}^d}$.

Proof. Let us first prove that there exists a sequence of positive real numbers $(D_n)_{n \in \mathbb{N}}$ such that for every deformation φ of $\mathbb{R}^d$ of class $\mathcal{C}^\ell$, with $\ell \in \mathbb{N}^*$, induced by the time-dependent vector field $v \in L^1(0, 1; \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d))$, one has

$$\sup_{t \in [0,1]} \|\varphi(t) - \mathrm{Id}_{\mathbb{R}^d}\|_i \leqslant D_i \exp\left(D_i \int_0^1 \|v(t)\|_i \right), \tag{2}$$

for every $i \in \{0, \ldots, \ell\}$.

The case $i = 0$ is an immediate consequence of the integral formulation of (1). Combining the formula for computing derivatives of a composition of mappings with an induction argument shows that the derivatives of order i of $v \circ \varphi$ are polynomials in the derivatives of v and φ of order less than or equal to i. Moreover, these polynomials are of degree one with respect to the derivatives of v, and also of degree one with respect to the derivatives of φ of order i. Therefore we can write

$$\left| \frac{d}{dt} \partial_x^i \varphi(t, x) \right| \leqslant \|v(t)\|_i |\partial_x^i \varphi(t, x)| + \|v(t)\|_{i-1} P_i(|\partial_x^1 \varphi(t, x)|, \ldots, |\partial_x^{i-1} \varphi(t, x)|), \tag{3}$$

where P_i is a polynomial independent of v and φ, and the norms of the derivatives of the $\partial_x^j \varphi(t, x)$ are computed in the space of $\mathbb{R}^d$-valued multilinear maps. The result then follows from Gronwall estimates and from an induction argument on i.

That any deformation of $\mathbb{R}^d$ of class $\mathcal{C}^\ell$ is a curve of class $W^{1,1}$ in $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ is then a direct consequence of (2) and (3).

Conversely, for every curve φ on $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ of class $W^{1,1}$, we set $v(t) = \dot{\varphi}(t) \circ \varphi^{-1}(t)$, for every $t \in [0, 1]$. We have $\dot{\varphi}(t) = v(t) \circ \varphi(t)$ for almost every $t \in [0, 1]$, and hence it suffices to prove that $t \mapsto \|v(t)\|_\ell$ is integrable. The curve φ is continuous on $[0, 1]$ and therefore is bounded. This implies that $t \mapsto \varphi(t)^{-1}$ is bounded as well. The formula for computing derivatives of compositions of maps then shows that $\|v(t)\|_\ell$ is integrable whenever $t \mapsto \|\dot{\varphi}(t)\|_\ell$ is integrable, which completes the proof since φ is of class $W^{1,1}$. $\square$

Reproducing Kernel Hilbert Spaces of vector fields. Let us briefly recall the definition and a few properties of RKHS's (see [7, 40] for more details). Let $\ell \in \mathbb{N}^*$ be fixed.

Given a Hilbert space $(V, (\cdot, \cdot)_V)$, according to the Riesz representation theorem, the mapping $v \mapsto (v, \cdot)_V$ is a bijective isometry between V and V^*, whose inverse is denoted by K_V. Then for every $p \in V^*$ and every $v \in V$ one has $\langle p, v \rangle_{V^*, V} = (K_V p, v)_V$ and $\|p\|_{V^*}^2 = \|K_V p\|_V^2 = \langle p, K_V p \rangle_{V^*, V}$.

Definition 1. *A Reproducing Kernel Vector Space (RKHS) of vector fields of class $\mathcal{C}_0^\ell$ is a Hilbert space $(V, (\cdot, \cdot)_V)$ of vector fields on $\mathbb{R}^d$, such that $V \subset \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$ with continuous inclusion.*

Let V be an RKHS of vector fields of class $\mathcal{C}_0^\ell$. Then, for all $(b, y) \in (\mathbb{R}^d)^2$, by definition the linear form $b \otimes \delta_y$ on V, defined by $b \otimes \delta_y(v) = b^T v(y)$ for every $v \in V$, is continuous (actually this continuity property holds as well for every compactly supported vector-valued distribution of order at most ℓ on $\mathbb{R}^d$). By definition of K_V, there holds $\langle b \otimes \delta_y, v \rangle_{V^*, V} = (K_V(b \otimes \delta_y), v)_V$. The *reproducing kernel* K of V is then the mapping defined on $\mathbb{R}^d \times \mathbb{R}^d$, with values in the set of real square matrices of size d, defined by

$$K(x, y)b = K_V(b \otimes \delta_y)(x), \tag{4}$$

for all $(b, x, y) \in (\mathbb{R}^d)^3$. In other words, there holds $(K(\cdot, y)b, v)_V = b^T v(y)$, for all $(b, y) \in (\mathbb{R}^d)^2$ and every $v \in V$, and $K(\cdot, y)b = K_V(b \otimes \delta_y)$ is a vector field of class $\mathcal{C}^\ell$ in $\mathbb{R}^d$, element of V.

It is easy to see that $(K(\cdot, x)a, K(\cdot, y)b)_V = a^T K(x, y)b$, for all $(a, b, x, y) \in (\mathbb{R}^d)^4$, and hence that $K(x, y)^T = K(y, x)$ and that $K(x, x)$ is positive semi-definite under the assumption that no nontrivial linear combination $a_1^T v(x_1) + \cdots + a_n^T v(x_n)$, with given distinct x_j's can vanish for every $v \in V$. Finally, writing $K_V(a \otimes \delta_y)(x) = K(x, y)a = \int_{\mathbb{R}^d} K(x, s)a \, d\delta_y(s)$, we have

$$K_V p(x) = \int_{\mathbb{R}^d} K(x, y) \, dp(y), \tag{5}$$

for every compactly supported vector-valued distribution p on $\mathbb{R}^d$ of order less than or equal to ℓ.[1]

As explained in [7, 42], one of the interests of such a structure is that we can define the kernel itself instead of defining the space V. Indeed a given kernel K yields a unique associated RKHS. It is usual to consider kernels of the form $K(x, y) = \gamma(|x - y|)Id_{\mathbb{R}^d}$ with $\gamma \in \mathcal{C}^\infty(\mathbb{R})$. Such a kernel yields a metric that is invariant under rotation and translation. The most common model is when γ is a Gaussian function but other families of kernels can be used as well [39, 42].

2.2 From shape space problems to optimal control

We define a *shape space* in $\mathbb{R}^d$ as an open subset M of a Banach space X on which the group of diffeomorphisms of $\mathbb{R}^d$ acts in a certain way. The elements of M, called *states of the shape*, are denoted by q. They are usually subsets or immersed submanifolds of $\mathbb{R}^d$, with a typical definition of the shape space as the set $M = \text{Emb}^1(S, \mathbb{R}^d)$ of all embeddings of class $\mathcal{C}^1$ of a given Riemannian manifold S into $\mathbb{R}^d$. For example, if S is the unit circle then M is the set of all parametrized $\mathcal{C}^1$ simple closed curves in $\mathbb{R}^d$. In practical applications or in numerical implementations, one has to consider finite-dimensional approximations, so that $\mathcal{S}$ usually just consists of a finite set of points, and then M is a space of landmarks (see [39, 42] and see examples further).

Let us first explain how the group of diffeomorphisms acts on the shape space M, and then in which sense this action induces a control system on M.

---

[1]Indeed, it suffices to note that

$$b^T K_V p(x) = (b \otimes \delta_x, K_V p)_{V^*, V} = (p, K_V b \otimes \delta_x) = \int_{\mathbb{R}^d} (K(y, x)b)^T \, dp(y) = b^T \int_{\mathbb{R}^d} K(x, y) dp(y).$$

The group structure of $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$. Let $\ell \in \mathbb{N}^*$. The set $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ is an open subspace of the affine Banach space $\mathrm{Id}_{\mathbb{R}^d} + \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$ and also a group for the composition law. However, we can be more precise.

First of all, the mappings $(\varphi, \psi) \mapsto \varphi \circ \psi$ and $\varphi \mapsto \varphi^{-1}$ are continuous (this follows from the formula for the computation of the derivatives of compositions of mappings).

Moreover, for every $\psi \in \mathrm{Diff}_0^\ell(\mathbb{R}^d)$, the right-multiplication mapping $\varphi \mapsto R_\psi(\varphi) = \varphi \circ \psi$ is Lipschitz and of class $\mathcal{C}^1$, as the restriction of the continuous affine map $(\mathrm{Id}_{\mathbb{R}^d} + v) \mapsto (\mathrm{Id}_{\mathbb{R}^d} + v) \circ \psi$. Its derivative $(dR_\psi)_{\mathrm{Id}_{\mathbb{R}^d}} : \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d) \to \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$ at $\mathrm{Id}_{\mathbb{R}^d}$ is then given by $v \mapsto v \circ \psi$. Moreover, $(v, \psi) \mapsto v \circ \psi$ is easily seen to be continuous.

Finally, the mapping

$$
\begin{array}{ccc}
\mathcal{C}_0^{\ell+1}(\mathbb{R}^d, \mathbb{R}^d) \times \mathrm{Diff}_0^\ell(\mathbb{R}^d) & \to & \mathcal{C}^\ell(\mathbb{R}^d, \mathbb{R}^d) \\
(v, \psi) & \mapsto & v \circ \psi
\end{array}
$$

is of class $\mathcal{C}^1$. Indeed we have $\|v \circ (\psi + \delta\psi) - v \circ \psi - dv_\psi . \delta\psi\|_\ell = o(\|\delta\psi\|_\ell)$, for every $\delta\psi \in \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$. Then, using the uniform continuity of any derivative $d^i v$ of order $i \leqslant \ell$, it follows that the mapping $\psi \mapsto dv_\psi$ is continuous. These properties are useful for the study of the Fréchet Lie group structure of $\mathrm{Diff}_0^\infty(\mathbb{R}^d)$ [32].

Group action on the shape space. In the sequel, we fix $\ell \in \mathbb{N}$, and we assume that the space $\mathrm{Diff}_0^{\max(1,\ell)}(\mathbb{R}^d)$ acts continuously on M (recall that M is an open subset of a Banach space X) according to a mapping

$$
\begin{array}{ccc}
\mathrm{Diff}_0^{\max(1,\ell)}(\mathbb{R}^d) \times M & \to & M \\
(\varphi, q) & \mapsto & \varphi \cdot q,
\end{array}
\tag{6}
$$

such that $\mathrm{Id}_{\mathbb{R}^d} \cdot q = q$ and $\varphi \cdot (\psi \cdot q) = (\varphi \circ \psi) \cdot q$ for every $q \in M$ and all $(\varphi, \psi) \in (\mathrm{Diff}^{\max(1,\ell)}(\mathbb{R}^d))^2$.

Definition 2. *M is a shape space of order $\ell \in \mathbb{N}$ if the action (6) is compatible with the properties of the group structure of $\mathrm{Diff}_0^{\max(1,\ell)}(\mathbb{R}^d)$ described above, that is:*

- *For every $q \in M$ fixed, the mapping $\varphi \mapsto \varphi \cdot q$ is Lipschitz with respect to the (weaker when $\ell = 0$) norm $\|\cdot\|_\ell$, i.e., there exists $\gamma > 0$ such that*

$$
\|\varphi_1 \cdot q - \varphi_2 \cdot q\|_X \leqslant \gamma \|\varphi_1 - \varphi_2\|_\ell
\tag{7}
$$

  for all $(\varphi_1, \varphi_2) \in (\mathrm{Diff}_0^{\max(1,\ell)}(\mathbb{R}^d))^2$.

- *The mapping $\varphi \mapsto \varphi \cdot q$ is differentiable at $\mathrm{Id}_{\mathbb{R}^d}$. This differential is denoted by ξ_q and is called the infinitesimal action of $\mathcal{C}_0^{\max(1,\ell)}(\mathbb{R}^d, \mathbb{R}^d)$. From (7) one has*

$$
\|\xi_q v\|_X \leqslant \gamma \|v\|_\ell,
$$

  for every $v \in \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$, and if $\ell = 0$ then ξ_q has a unique continuous extension to the whole space $\mathcal{C}_0^0(\mathbb{R}^d, \mathbb{R}^d)$.

- *The mapping*

$$
\begin{array}{ccc}
\xi : M \times \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d) & \longrightarrow & X \\
(q, v) & \longmapsto & \xi_q v
\end{array}
\tag{8}
$$

  is continuous, and its restriction to $M \times \mathcal{C}_0^{\ell+1}(\mathbb{R}^d, \mathbb{R}^d)$ is of class $\mathcal{C}^1$. In particular the mapping $q \mapsto \xi_q v$ is of class $\mathcal{C}^1$, for every bounded vector field v of class $\mathcal{C}^{\ell+1}$.

Example 1. For $\ell \geqslant 1$, the action of $\mathrm{Diff}_0^\ell(\mathbb{R}^d)$ on itself by left composition makes it a shape space of order ℓ in $\mathbb{R}^d$.

Example 2. Let $\ell \in \mathbb{N}$ and let S be a $\mathcal{C}^\ell$ smooth compact Riemannian manifold. Consider the space $M = X = \mathcal{C}^\ell(S, \mathbb{R}^d)$ equipped with its usual Banach norm. Then M is a shape space of order ℓ, where the action of $\mathrm{Diff}_0^{\max(1,\ell)}(\mathbb{R}^d)$ is given by the composition $\varphi \cdot q = \varphi \circ q$. Indeed, it is continuous thanks to the rule for computing derivatives of a composition, and we also have

$$\|\varphi_1 \cdot q - \varphi_2 \cdot q\|_X \leqslant \gamma \|\varphi_1 - \varphi_2\|_\ell.$$

Moreover, given $q \in M$ and $v \in \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$, $\xi_q v$ is the vector field along q given by $\xi_q(v) = v \circ q \in \mathcal{C}^\ell(M, \mathbb{R}^d)$. Finally, the formula for computing derivatives of a composition yields

$$\|v \circ (q + \delta q) - v \circ q - dv_q . \delta q\|_X = o(\|\delta q\|_X),$$

for every $\delta q \in M$, and the last part of the definition follows. This framework describes most of shape spaces.

An interesting particular case of this general example is when $S = (s_1, \dots, s_n)$ is a finite set (zero-dimensional manifold), $X = (\mathbb{R}^d)^n$ and

$$M = \mathrm{Lmk}_d(n) = \{(x_1, \dots, x_n) \in (\mathbb{R}^d)^n \mid x_i \neq x_j \text{ if } i \neq j\}$$

is a (so-called) space of n landmarks in $\mathbb{R}^d$. For $q = (x_1, \dots, x_n)$, the smooth action of order 0 is $\varphi \cdot q = (\varphi(x_1), \dots, \varphi(x_n))$. For $v \in \mathcal{C}_0^0(\mathbb{R}^2, \mathbb{R}^2)$, the infinitesimal action of v at q is given by $\xi_q(v) = (v(x_1), \dots, v(x_n))$.

Remark 1. In most cases, and in all examples given throughout this paper, the mapping ξ restricted to $M \times \mathcal{C}_0^{\ell+k}(\mathbb{R}^d, \mathbb{R}^d)$ is of class $\mathcal{C}^k$, for every $k \in \mathbb{N}$.

Proposition 2. *For every $q \in M$, the mapping $\varphi \mapsto \varphi \cdot q$ is of class $\mathcal{C}^1$, and its differential at φ is given by $\xi_{\varphi \cdot q} dR_{\varphi^{-1}}$. In particular, given $q_0 \in M$ and given φ a deformation of $\mathbb{R}^d$ of class $\mathcal{C}_0^{\max 1, \ell}$, which is the flow of the time-dependent vector field v, the curve $t \mapsto q(t) = \varphi(t) \cdot q_0$ is of class $W^{1,1}$ and one has*

$$\dot{q}(t) = \xi_{\varphi(t) \cdot q_0} \dot{\varphi}(t) \circ \varphi(t)^{-1} = \xi_{q(t)} v(t), \tag{9}$$

for almost every $t \in [0, 1]$.

Proof. Let $q_0 \in M$, fix $\varphi \in \mathrm{Diff}_0^\ell(\mathbb{R}^d)$ and take $\delta\varphi \in T_\varphi \mathrm{Diff}_0^\ell(\mathbb{R}^d) = \mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$. Then $\varphi + \delta\varphi \in \mathrm{Diff}_0^\ell(\mathbb{R}^d)$ for $\|\delta\varphi\|_\ell$ small enough. We define $v = (dR_{\varphi^{-1}})_\varphi \delta\varphi = \delta\varphi \circ \varphi^{-1}$. We have

$$(\varphi + \delta\varphi) \cdot q = (\mathrm{Id}_{\mathbb{R}^d} + v) \cdot (\varphi \cdot q) = \varphi \cdot q + \xi_{\varphi \cdot q} v + o(v) = \varphi \cdot q + \xi_{\varphi \cdot q} \dot{\varphi} \circ \varphi^{-1} + o(\delta\varphi),$$

and therefore the mapping $\varphi \mapsto \varphi \cdot q$ is differentiable at φ, with continuous differential $\xi_{\varphi \cdot q} dR_{\varphi^{-1}}$. The result follows as the image of a curve of class $W^{1,1}$ through a $\mathcal{C}^1$ mapping is still of class $W^{1,1}$. The derivative is then computed thanks to the chain rule. $\square$

The result of this proposition shows that the shape $q(t) = \varphi(t) \cdot q_0$ is evolving in time according to the differential equation (9), where v is the time-dependent vector field associated with the deformation φ.

At this step we make a crucial connection between shape space analysis and control theory, by adopting another point of view. The differential equation (9) can be seen as a control system on M, where the time-dependent vector field v is seen as a control. In conclusion, the group of diffeomorphisms acts on the shape space M, and this action induces a control system on M.

As said in the introduction, in shape analysis problems, the shapes are usually assumed to evolve in time according to the minimization of some objective functional [40]. With the control theory viewpoint developed above, this leads us to model the shape evolution as an optimal control problem settled on M, that we define hereafter.

Induced optimal control problem on the shape space. We assume that the action of $\text{Diff}_0^{\max(\ell,1)}(\mathbb{R}^d)$ on M is smooth of order $\ell \in \mathbb{N}$. Let $(V, (\cdot, \cdot)_V)$ be an RKHS of vector fields of class $\mathcal{C}_0^\ell$ on $\mathbb{R}^d$. Let K denote its reproducing kernel (as defined in Section 2.1). Let Y be another Banach space. Most problems of shape analysis can be recast as follows.

Problem 1. *Let $q_0 \in M$, and let $C : M \times V \to Y$ be a mapping such that $v \mapsto C(q, v) = C_q v$ is linear for every $q \in M$. Let $g : M \to \mathbb{R}$ be a function. We consider the problem of minimizing the functional*

$$J_1(q, v) = \frac{1}{2} \int_0^1 \|v(t)\|_V^2 \, dt + g(q(1)) \tag{10}$$

over all $(q(\cdot), v(\cdot)) \in W_{q_0}^{1,1}(0, 1; M) \times L^2(0, 1; V)$ such that $\dot{q}(t) = \xi_{q(t)} v(t)$ and $C_{q(t)} v(t) = 0$ for almost every $t \in [0, 1]$.

In the problem above, q_0 stands for an initial shape, and C stands for continuous constraints. Recall that the infinitesimal action can be extended to the whole space $\mathcal{C}_0^\ell(\mathbb{R}^d, \mathbb{R}^d)$.

Note that if $t \mapsto v(t)$ is square-integrable then $t \mapsto \dot{q}(t)$ is square-integrable as well. Indeed this follows from the differential equation $\dot{q}(t) = \xi_{q(t)} v(t)$ and from Gronwall estimates. Therefore the minimization runs over the set of all $(q(\cdot), v(\cdot)) \in H_{q_0}^1(0, 1; M) \times L^2(0, 1; V)$.

Problem 1 is an infinite-dimensional optimal control problem settled on M, where the state $q(t)$ is a shape and the control $v(\cdot)$ is a time-dependent vector field. The constraints C can be of different kinds, as illustrated further. A particular but important case of constraints consists of *kinetic* constraints, i.e., constraints on the speed $\dot{q} = \xi_q v$ of the state, which are of the form $C_{q(t)} \xi_{q(t)} v(t) = C_{q(t)} \dot{q}(t) = 0$. Pure state constraints, of the form $C(q(t)) = 0$ with a differentiable map $C : M \to Y$, are in particular equivalent to the kinetic constraints $dC_{q(t)}.\xi_{q(t)} v(t) = 0$.

Remark 2. Problem 1 is similar to what is called in [19] a Clebsch optimal control problem, with the difference that the space of controls is only a subspace of the Lie algebra of the group.

To the best of our knowledge, except for very few studies (such as [43]), only unconstrained problems have been studied so far (i.e., with $C = 0$). In contrast, the framework that we provide here is very general and permits to model and solve far more general constrained shape deformation problems.

Remark 3. Assume V is an RKHS of class $\mathcal{C}_0^1$, and let $v(\cdot) \in L^2(0, 1; V)$. Then v induces a unique deformation $t \mapsto \varphi(t)$ on $\mathbb{R}^d$, and the curve $t \mapsto q_v(t) = \varphi(t) \cdot q_0$ satisfies $q(0) = q_0$ and $\dot{q}_v(t) = \xi_{q_v(t)} v(t)$ for almost every $t \in [0, 1]$. As above, it follows from Gronwall's lemma that $q \in H_{q_0}^1(0, 1; M)$. Moreover, according to the Cauchy-Lipschitz theorem, if $\ell \geqslant 1$ then $q(\cdot)$ is the unique such element of $H_{q_0}^1(0, 1; M)$. Therefore, if $\ell \geqslant 1$ then Problem 1 is equivalent to the problem of minimizing the functional $v \mapsto J_1(v, q_v)$ over all $v \in L^2(0, 1; V)$ such that $C_{q_v(t)} v(t) = 0$ for almost every $t \in [0, 1]$.

Remark 4. It is interesting to note that, since the control system $\dot{q}(t) = \xi_{q(t)} v(t)$ and the constraints $C_{q(t)} v(t) = 0$ are linear with respect to the control $v(\cdot)$, they define a *rank-varying sub-Riemannian structure* on M (as defined in [2]). Moreover, solutions of Problem 1 are obviously *minimizing geodesics* for this structure, since the integral part of the cost corresponds to the total energy of the horizontal curve.

Concerning the existence of an optimal solution of Problem 1, we need the following definition.

Definition 3. *A state q of a shape space M of order ℓ is said to have compact support if for some compact subset U of $\mathbb{R}^d$, for some $\gamma > 0$ and for all $(\varphi_1, \varphi_2) \in (Diff_0^{\max(\ell,1)}(\mathbb{R}^d))^2$, we have*

$$\|\varphi_1 \cdot q - \varphi_2 \cdot q\| \leqslant \gamma \|(\varphi_1 - \varphi_2)_{|U}\|_\ell,$$

where $(\varphi_1 - \varphi_2)_{|U}$ denotes the restriction of $\varphi_1 - \varphi_2$ to U.

Except for $Diff_0^{\max(\ell,1)}(\mathbb{R}^d)$ itself, every state of every shape space given so far in examples had compact support.

Theorem 1. *Assume that V is an RKHS of vector fields of class $\mathcal{C}^{\ell+1}$ on $\mathbb{R}^d$, that $q \mapsto C_q$ is continuous, and that g is bounded below and lower semi-continuous. If q_0 has compact support, then Problem 1 has at least one solution.*

In practice one does not usually have available a convenient, functional definition of the space V of vector fields. The RKHS V is in general only known through its kernel K, as already mentioned in Section 2.1 (and the kernel is often a Gaussian one). Hence Problem 1, formulated as such, is not easily tractable since one might not have a good knowledge (say, a parametrization) of the space V.

One can however derive, under a slight additional assumption, a different formulation of Problem 1 that may be more convenient and appropriate in view of practical issues. This is done in the next section, in which our aim is to obtain an optimal control problem only depending on the knowledge of the reproducing kernel K of the space V (and not directly on V itself), the solutions of which can be lifted back to the group of diffeomorphisms.

Kernel formulation of the optimal control problem. For a given $q \in M$, consider the transpose $\xi_q^* : X^* \to V^*$ of the continuous linear mapping $\xi_q : V \to X$. This means that for every $u \in X^*$ the element $\xi_q^* u \in V^*$ (sometimes called pullback) is defined by $\langle \xi_q^* u, v \rangle_{V^*,V} = \langle u, \xi_q(v) \rangle_{X^*,X}$, for every $v \in V$. Besides, by definition of K_V, there holds $\langle \xi_q^* u, v \rangle_{V^*,V} = (K_V \xi_q^* u, v)_V$. The mapping $(q, u) \in M \times X^* \mapsto \xi_q^* u \in V^*$ is often called a *momentum map* in control theory [28].

We start our discussion with the following remark. As seen in Example 2, we observe that, in general, given $q \in M$ the mapping ξ_q is far from being injective (i.e., one-to-one). Its null space $\mathrm{Null}(\xi_q)$ can indeed be quite large, with many possible time-dependent vector fields v generating the same solution of $q(0) = q_0$ and $\dot{q}(t) = \xi_{q(t)} v(t)$ for almost every $t \in [0,1]$.

A usual way to address this overdetermination consists of selecting, at every time t, a $v(t)$ that has minimal norm subject to $\xi_{q(t)} v(t) = \dot{q}(t)$ (resulting in a least-squares problem). This is the object of the following lemma.

Lemma 1. *Let $q \in M$. Assume that $\mathrm{Range}(\xi_q) = \xi_q(V)$ is closed. Then, for every $v \in V$ there exists $u \in X^*$ such that $\xi_q v = \xi_q K_V \xi_q^* u$. Moreover, the element $K_V \xi_q^* u \in V$ is the one with minimal norm over all elements $v' \in V$ such that $\xi_q v' = \xi_q v$.*

Proof. Let $\hat{v}$ denote the orthogonal projection of 0 on the space $A = \{v' : \xi_q v' = \xi_q v\}$, i.e., the element of A with minimal norm. Then $\hat{v}$ is characterized by $\xi_q \hat{v} = \xi_q v$ and $\hat{v} \in \mathrm{Null}(\xi_q)^\perp$. Using the Banach closed-range theorem, we have $(\mathrm{Null}(\xi_q))^\perp = K_V \mathrm{Range}(\xi_q^*)$, so that there exists $u \in X^*$ such that $\hat{v} = K_V \xi_q^* u$, and hence $\xi_q v = \xi_q K_V \xi_q^* u$. $\qquad\square$

Remark 5. Note that the latter assertion in the proof does not require $\mathrm{Range}(\xi_q)$ to be closed, since we always have $K_V(\mathrm{Range}(\xi_q)) \subset (\mathrm{Null}(\xi_q)^\perp)$.

Whether Range(ξ_q) = $\xi_q(V)$ is closed or not, this lemma and the previous discussion suggest replacing the control $v(t)$ in Problem 1 by $u(t) \in X^*$ such that $v(t) = K_V \xi_{q(t)}^* u(t)$. Plugging this expression into the system $\dot{q}(t) = \xi_{q(t)} v(t)$ leads to the new control system $\dot{q}(t) = K_{q(t)} u(t)$, where

$$K_q = \xi_q K_V \xi_q^*, \tag{11}$$

for every $q \in M$. The operator $K_q : X^* \to X$ is continuous and symmetric (i.e., $\langle u_2, K_q u_1 \rangle_{X^*,X} = \langle u_1, K_q u_2 \rangle_{X^*,X}$ for all $(u_1, u_2) \in (X^*)^2$), satisfies $\langle u, K_q u \rangle_{X^*,X} = \|K_V \xi_q^* u\|_V^2$ for every $u \in X$ and thus is positive semi-definite, and $(q, u) \mapsto K_q u$ is as regular as $(q, v) \mapsto \xi_q v$. Note that $K_q(X^*) = \xi_q(V)$ whenever $\xi_q(V)$ is closed.

This change of variable appears to be particularly relevant since the operator K_q is usually easy to compute from the reproducing kernel K_V of V, as shown in the following examples.

Example 3. Let $M = X = \mathcal{C}^0(S, \mathbb{R}^d)$ be the set of continuous mappings from a Riemannian manifold S to $\mathbb{R}^d$. The action of $\text{Diff}^0(\mathbb{R}^d)$ is smooth of order 0, with $\xi_q v = v \circ q$ (see Example 2). Let V be an RKHS of vector fields of class $\mathcal{C}_0^1$ on $\mathbb{R}^d$, with reproducing kernel K. Every $u \in X^*$ can be identified with a vector-valued Radon measure on S. Then

$$\langle \xi_q^* u, v \rangle_{V^*,V} = \langle u, v \circ q \rangle_{X^*,X} = \int_S v(q(s))^T du(s),$$

for every $q \in M$ and for every $v \in V$. In other words, one has $\xi_q^* u = \int_S du(s) \otimes \delta_{q(s)}$, and therefore, by definition of the kernel, we have $K_V \xi_q^* u = \int_S K_V(du(s) \otimes \delta_{q(s)}) = \int_S K(\cdot, q(s)) du(s)$. We finally infer that

$$K_q u(t) = \int_S K(q(t), q(s)) du(s).$$

Example 4. Let $X = (\mathbb{R}^d)^n$ and $M = \text{Lmk}_d(n)$ (as in Example 2). Then $\xi_q v = (v(x_1), \ldots, v(x_n))$, and every $u = (u_1, \ldots, u_n)$ is identified with a vector of X by $\langle u, w \rangle_{X^*,X} = \sum_{j=1}^n u_j^T w_j$. Therefore, we get $\xi_q^* u = \sum_{j=1}^n u_j \otimes \delta_{x_j}$, and $K_V \xi_q^* u = \sum_{j=1}^n K(x_j, \cdot) u_j$. It follows that

$$K_q u = \left(\sum_{j=1}^n K(x_1, x_j) u_j, \sum_{j=1}^n K(x_2, x_j) u_j, \ldots, \sum_{j=1}^n K(x_n, x_j) u_j \right).$$

In other words, K_q can be identified with matrix of total size $nd \times nd$ and made of square block matrices of size d, with the block (i, j) given by $K(x_i, x_j)$.

Following the discussion above and the change of control variable $v(t) = K_V \xi_{q(t)}^* u(t)$, we are led to consider the following optimal control problem.

Problem 2. *Let $q_0 \in M$, and let $C : M \times V \to Y$ be a mapping such that $v \mapsto C(q, v) = C_q v$ is linear for every $q \in M$. Let $g : M \to \mathbb{R}$ be a function. We consider the problem of minimizing the functional*

$$J_2(q, u) = \frac{1}{2} \int_0^1 \langle u(t), K_{q(t)} u(t) \rangle_{X^*,X} dt + g(q(1)) = E(q(t)) + g(q(1)) \tag{12}$$

over all couples $(q(\cdot), u(\cdot))$, where $u : [0, 1] \to X^*$ is a measurable function and $q(\cdot) \in W_{q_0}^{1,1}(0, 1; M)$ are such that $\dot{q}(t) = K_{q(t)} u(t)$ and $C_{q(t)} K_V \xi_{q(t)}^* u(t) = 0$ for almost every $t \in [0, 1]$.

Note that, setting $\ell(q, v) = \frac{1}{2} \langle v, v \rangle$, the relation $v = K_V \xi_q^* u$ can be recovered from the necessary condition $\frac{\delta \ell}{\delta v} = \xi_q^* u$ for minimizers in Clebsch optimal control (see [19]), obtained thanks to the Pontryagin maximum principle. This suggests that Problems 1 and 2 may be equivalent in certain cases. The precise relation between both problems is clarified in the following result.

Proposition 3. *Assume that* $\mathrm{Null}(\xi_q) \subset \mathrm{Null}(C_q)$ *and that* $\mathrm{Range}(\xi_q) = \xi_q(V)$ *is closed, for every* $q \in M$. *Then Problems 1 and 2 are equivalent in the sense that* $\inf J_1 = \inf J_2$ *over their respective sets of constraints.*

Moreover, if $(\bar{q}(\cdot), \bar{u}(\cdot))$ *is an optimal solution of Problem 2, then* $(\bar{q}(\cdot), \bar{v}(\cdot))$ *is an optimal solution of Problem 1, with* $\bar{v}(\cdot) = K_V \xi_{\bar{q}(\cdot)}^* \bar{u}(\cdot)$ *and* $\bar{q}(\cdot)$ *the corresponding curve defined by* $\bar{q}(0) = q_0$ *and* $\dot{\bar{q}}(t) = K_{\bar{q}(t)} \bar{u}(t)$ *for almost every* $t \in [0,1]$. *Conversely, if* $(\bar{q}(\cdot), \bar{v}(\cdot))$ *is an optimal solution of Problem 1 then there exists a measurable function* $\bar{u} : [0,1] \to X^*$ *such that* $\bar{v}(\cdot) = K_V \xi_{\bar{q}(\cdot)}^* \bar{u}(\cdot)$, *and* $\bar{u}(\cdot)$ *and* $(\bar{q}(\cdot), \bar{u}(\cdot))$ *is an optimal solution of Problem 2.*

Proof. First of all, if $J_2(q, u)$ is finite, then $v(\cdot)$ defined by $v(t) = K_V \xi_{q(t)} u(t)$ belongs to $L^2(0,1;V)$ and therefore, using the differential equation $\dot{q}(t) = \xi_{q(t)} v(t)$ for almost every t and Gronwall's lemma, we infer that $q \in H^1_{q_0}(0,1;M)$. The inequality $\inf J_1 \leqslant \inf J_2$ follows obviously.

Let us prove the converse. Let $\varepsilon > 0$ arbitrary, and let $v(\cdot) \in L^2(0,1;V)$ and $q \in H^1_{q_0}(0,1;M)$ be such that $J_1(q,v) \leqslant \inf J_1 + \varepsilon$, with $\dot{q}(t) = \xi_{q(t)} v(t)$ and $C_{q(t)} v(t) = 0$ for almost every $t \in [0,1]$. We can write $v(t) = v_1(t) + v_2(t)$ with $v_1(t) \in \mathrm{Null}(\xi_{q(t)})$ and $v_2(t) \in (\mathrm{Null}(\xi_{q(t)}))^\perp = \mathrm{Range}(K_V \xi_{q(t)}^*)$, for almost every $t \in [0,1]$, with $v_1(\cdot)$ and $v_2(\cdot)$ measurable functions, and obviously one has $\int_0^T \|v_2(t)\|_V^2 \, dt \leqslant \int_0^T \|v(t)\|_V^2 \, dt$. Then, choosing $u(\cdot)$ such that $v_2(\cdot) = K_V \xi_{q(\cdot)}^* u(\cdot)$, it follows that $J_2(u) = J_1(v_2) \leqslant J_1(v) \leqslant \inf J_1 + \varepsilon$. Therefore $\inf J_2 \leqslant \inf J_1$. The rest is obvious. $\qquad\square$

Remark 6. Under the assumptions of Proposition 3 and of Theorem 1, Problem 2 has at least one solution $\bar{u}(\cdot)$, there holds $\min J_1 = \min J_2$, and the minimizers of Problems 1 and 2 are in one-to-one correspondance according to the above statement.

Remark 7. The assumption $\mathrm{Null}(\xi_q) \subset \mathrm{Null}(C_q)$ is satisfied in the important case where the constraints are kinetic, and is natural to be considered since it means that, in the problem of overdetermination in v, the constraints can be passed to the quotient (see Lemma 1). Actually for kinetic constraints we have the following interesting result (proved further, see Remark 20), completing the discussion on the equivalence between both problems.

Proposition 4. *Assume that V is an RKHS of vector fields of class at least $\mathcal{C}_0^{\ell+1}$ on $\mathbb{R}^d$, that the constraints are kinetic, i.e., are of the form $C_q \xi_q v = 0$, and that the mapping $(q, w) \mapsto C_q w$ is of class $\mathcal{C}^1$. If $C_q \xi_q$ is surjective (onto) for every $q \in M$, then for every optimal solution $\bar{v}$ of Problem 1 there exists a measurable function $\bar{u} : [0,1] \to X^*$ such that $\bar{v} = K_V \xi_{\bar{q}}^* \bar{u}$, and $\bar{u}$ is an optimal solution of Problem 2.*

Note that this result does not require the assumption that $\mathrm{Range}(\xi_q) = \xi_q(V)$ be closed.

Remark 8. It may happen that Problems 1 and 2 do not coincide whenever $\mathrm{Range}(\xi_q)$ is not closed. Actually, if the assumption that $\mathrm{Range}(\xi_q)$ is closed is not satisfied then it may happen that the set of controls satisfying the constraints in Problem 2 are reduced to the zero control.

Let us provide a situation where this occurs. Let $v(q) \in \overline{\mathrm{Range}(K_V \xi_q^*)} \setminus \mathrm{Range}(K_V \xi_q^*)$ with $\|v(q)\|_V = 1$. In particular $v(q) \in (\mathrm{Null}(\xi_q))^\perp$. Assume that C_q is defined as the orthogonal projection onto $(\mathbb{R}v(q) \oplus \mathrm{Null}(\xi_q))^\perp = v(q)^\perp \cap (\mathrm{Null}(\xi_q))^\perp$. Then $\mathrm{Null}(C_q) = \mathbb{R}v(q) \oplus \mathrm{Null}(\xi_q)$. We claim that $\mathrm{Null}(C_q K_V \xi_q^*) = \{0\}$. Indeed, let $u \in X^*$ be such that $C_q K_V \xi_q^* u = 0$. Then on the one part $K_V \xi_q^* u \in \mathrm{Null}(C_q)$, and on the other part, $K_V \xi_q^* u \in \mathrm{Range}(K_V \xi_q^*) \subset (\mathrm{Null}(\xi_q))^\perp$. Therefore $K_V \xi_q^* u \in \mathrm{Null}(C_q) \cap (\mathrm{Null}(\xi_q))^\perp = \mathbb{R}v(q)$, but since $v(q) \notin \mathrm{Range}(K_V \xi_q^*)$, necessarily $u = 0$.

2.3 Further comments: lifted shape spaces and multishapes

In this section we provide one last way to study shape spaces and describe two interesting and important variants of shape spaces, namely lifted shape spaces and multishapes. We show that

a slightly different optimal control problem can model the shape deformation problem in these spaces.

Lifted shape spaces. Lifted shapes can be used to keep track of additional parameters when studying the deformation of a shape. For example, when studying n landmarks $(x_1, \ldots, x_n)$ in $\mathbb{R}^d$, it can be interesting to keep track of how another point x is moved by the deformation.

Let M and $\hat{M}$ be two shape spaces, open subsets of two Banach spaces X and $\hat{X}$ respectively, on which the group of diffeomorphisms of $\mathbb{R}^d$ acts smoothly with respective orders ℓ and $\hat{\ell}$. Let V be an RKHS of vector fields in $\mathbb{R}^d$ of class $\mathcal{C}_0^{\max(\ell,\hat{\ell})}$. We denote by ξ_q (respectively $\xi_{\hat{q}}$) the infinitesimal action of V on M (respectively $\hat{M}$). We assume that there exists a $\mathcal{C}^1$ equivariant submersion $P : \hat{M} \to M$.

By equivariant, we mean that $P(\varphi \cdot \hat{q}) = \varphi \cdot P(\hat{q})$, for every diffeomorphism $\varphi \in \mathrm{Diff}^{\max(\ell,\hat{\ell})+1}$ and every $\hat{q} \in \hat{M}$. Note that this implies that $dP_{\hat{q}}.\xi_{\hat{q}} = \xi_q$ and $\xi_{\hat{q}}^* dP_{\hat{q}}^* = \xi_q^*$.

For example, for $n < \hat{n}$, the projection $P : \mathrm{Lmk}_d(\hat{n}) \to \mathrm{Lmk}_d(n)$ defined by $P(x_1, \ldots, x_{\hat{n}}) = (x_1, \ldots, x_n)$ is a $\mathcal{C}^1$ equivariant submersion. More generally, for a compact Riemannian manifold $\hat{S}$ and a submanifold $S \subset \hat{S}$, the restriction mapping $P : \mathrm{Emb}(\hat{S}, \mathbb{R}^d) \to \mathrm{Emb}(S, \mathbb{R}^d)$ defined by $P(q) = q_{|S}$ is a $\mathcal{C}^1$ equivariant submersion for the action by composition of $\mathrm{Diff}^1(\mathbb{R}^d)$.

The constructions and results of Section 2.2 can be applied to this setting, and in particular the deformation evolution induces a control system on $\hat{M}$, as investigated previously.

Remark 9. Let V be an RKHS of bounded vector fields of class $C_0^{\max(\ell,\hat{\ell})+1}$. Let g be a data attachment function on M and let C be a mapping of constraints. We set $\hat{g} = g \circ P$ and $\hat{C}_{\hat{q}} = C_{P(\hat{q})}$. Then a time-dependent vector field v in V is a solution of Problem 1 for M with constraints C and data attachment g if and only if it is also a solution of Problem 1 for $\hat{M}$ with constraints $\hat{C}$ and data attachment $\hat{g}$. This remark will be used for finite-dimensional approximations in Section 2.4.

One can however define a control system of a different form, by lifting the control applied on the smaller shape space M to the bigger shape space $\hat{M}$.

The method goes as follows. Let $q_0 \in M$ and $\hat{q}_0 \in P^{-1}(q_0)$. Consider a measurable map $u : [0,1] \to X^*$ and the corresponding curve $q(\cdot)$ defined by $q(0) = q_0$ and $\dot{q}(t) = K_{q(t)} u(t)$ for almost every $t \in [0,1]$, where $K_q = \xi_q K_V \xi_q^*$. This curve is the same as the one induced by the time-dependent vector field $v(\cdot) = K_V \xi_{q(\cdot)}^* u(\cdot)$. The deformation φ corresponding to the flow of v defines on $\hat{M}$ a new curve $\hat{q}(t) = \varphi(t) \cdot \hat{q}_0$ with speed

$$\dot{\hat{q}}(t) = \xi_{\hat{q}} K_V \xi_{q(t)}^* u(t) = K_{\hat{q}(t)} dP_{\hat{q}(t)}^* u(t),$$

with $K_{\hat{q}} = \xi_{\hat{q}} K_V \xi_{\hat{q}}^*$. Note that $P(\hat{q}(t)) = q(t)$ for every $t \in [0,1]$. We have thus obtained a new class of control problems.

Problem 3. *Let $\hat{q}_0 \in \hat{M}$, and let $C : \hat{M} \times V \to Y$ be continuous and linear with respect to the second variable, with Y a Banach space. Let $g : \hat{M} \to \mathbb{R}$ be a real function on $\hat{M}$. We consider the problem of minimizing the functional*

$$J_3(\hat{q}, u) = \frac{1}{2} \int_0^1 \langle u(t), K_{P(\hat{q}(t))} u(t) \rangle_{X^*,X} dt + g(\hat{q}(1))$$

over all $(\hat{q}(\cdot), u(\cdot))$, where $u : [0,1] \to X^*$ is a measurable function and $\hat{q}(\cdot) \in W_{\hat{q}_0}^{1,1}(0,1;\hat{M})$ are such that $\dot{\hat{q}}(t) = K_{\hat{q}(t)} dP_{\hat{q}(t)}^* u(t)$ and $C_{\hat{q}(t)} K_V \xi_{P(\hat{q}(t))}^* u(t) = 0$ for almost every $t \in [0,1]$.

Note that, if g and C only depend on $P(\hat{q})$ then the solutions $u(t)$ of Problem 3 coincide with the ones of Problem 2 on M.

Problem 3 can be reformulated back into an optimal control problem on V and on $\hat{M}$, similar to Problem 1, by adding the constraints $D_{\hat{q}}v = 0$ where $D_{\hat{q}}v$ is the orthogonal projection of v on $\text{Null}(\xi_{P(\hat{q})})$.

Some examples of lifted shape spaces can be found in [43], and more recently in [18], where controls are used from a small number of landmarks to match a large number of landmarks, with additional state variables defining Gaussian volume elements. Another application of lifted shape spaces will be mentioned in Section 2.4, where they will be used to approximate infinite-dimensional shape spaces by finite-dimensional ones.

Multishapes. As mentioned in the introduction, shape analysis problems sometimes involve collections of shapes that must be studied together, each of them with specific properties associated with a different space of vector fields. These situations can be modeled as follows.

Consider some shape spaces $M_1, \ldots, M_k$, open subsets of Banach spaces $X_1, \ldots, X_k$, respectively, on which diffeomorphisms of $\mathbb{R}^d$ acts smoothly on each shape space M_i with order ℓ_i. Let $k_i \geqslant 1$, and consider $V_1, \ldots, V_k$, RKHS's of vector fields of $\mathbb{R}^d$ respectively of class $\mathcal{C}_0^{\ell_i + k_i}$ with kernels $K_1, \ldots, K_k$, as defined in Section 2.1. In such a model we thus get k control systems, of the form $\dot{q}_i(\cdot) = \xi_{i,q_i(\cdot)} v_i(\cdot)$, with the controls $v_i(\cdot) \in L^2(0,1;V_i)$, $i = 1, \ldots, k$. The shape space of a multi-shape is a space of the form $M = M_1 \times \cdots \times M_k$. Let $q_0 = (q_{1,0}, \ldots, q_{k,0}) \in M$. Similarly to the previous section we consider the problem of minimizing the functional

$$\sum_{i=1}^{k} \int_0^1 \|v_i(t)\|_{V_i}^2 dt + g(q_1(1), \ldots, q_k(1)),$$

over all time-dependent vector fields $v_i(\cdot) \in L^2(0,1;V_i)$, $i = 1, \ldots, k$, and with $q_i(1) = \varphi_i(1) \cdot q_{i,0}$ where φ_i is the flow generated by v_i (note that, here, the problem is written without constraint).

As in Section 2.2, the kernel formulation of this optimal control problem consists of minimizing the functional

$$\frac{1}{2} \int_0^1 \sum_{i=1}^{k} \langle u_i(t), K_{q_i(t),i} u_i(t) \rangle_{X_i^*, X_i} dt + g(q(1)). \tag{13}$$

over all measurable functions $u(\cdot) = (u_1(\cdot), \ldots, u_k(\cdot)) \in L^2(0,1; X_1^* \times \cdots \times X_k^*)$, where the curve $q(\cdot) = (q_1(\cdot), \ldots, q_k(\cdot)) : [0,1] \to M$ is the solution of $q(0) = q_0$ and

$$\dot{q}(t) = K_{q(t)} u(t) = \left(K_{1,q_1(t)} u_1(t), \ldots, K_{k,q_k(t)} u_k(t) \right), \tag{14}$$

for almost every $t \in [0,1]$, with $K_{i,q_i} = \xi_{i,q_i} K_{V_i} \xi_{i,q_i}^*$ for $i = 1, \ldots, k$.

Obviously, without any further consideration, studying this space essentially amounts to studying each M_i separately, the only interaction possibly intervening from the final cost function g. More interesting problems arise however when the shapes can interact with each other, and are subject to consistency constraints. For example, assume that one studies a cat shape, decomposed into two parts for the body and the tail. Since these parts have very different properties, it makes sense to consider them *a priori* as two distinct shapes S_1 and S_2, with shape spaces $M_1 = \mathcal{C}^0(S_1, \mathbb{R}^3)$ and $M_2 = \mathcal{C}^0(S_2, \mathbb{R}^3)$, each of them being associated with RKHS's V_1 and V_2 respectively. Then, in order to take account for the tail being attached to the cat's body, the contact point of the body and the tail of the cat must belong to both shapes and be equal. In other words, if $q_1 \in M_1$ represents the body and $q_2 \in M_2$ the tail, then there must hold $q_1(s_1) = q_2(s_2)$ for some $s_1 \in S_1$ and $s_2 \in S_2$. This is a particular case of state constraints, i.e., constraints depending only on the state q of the trajectory.

Considering a more complicated example, assume that two (or more) shapes are embedded in a given background. Consider two states q_1 and q_2 in respective spaces $M_1 = \mathcal{C}^0(S_1, \mathbb{R}^d)$ and $M_2 = \mathcal{C}^0(S_2, \mathbb{R}^d)$ of $\mathbb{R}^d$. Assume that they represent the boundaries of two disjoint open subsets U_1 and U_2 of $\mathbb{R}^d$. We define a third space $M_3 = M_1 \times M_2$, whose elements are of the form $q_3 = (q_3^1, q_3^2)$. This shape space represents the boundary of the complement of $U_1 \cup U_2$ (this complement being the background). Each of these three shape spaces is acted upon by the diffeomorphisms of $\mathbb{R}^d$. Consider for every M_i an RKHS V_i of vector fields. The total shape space is then $M = M_1 \times M_2 \times M_3$, an element of which is given by $q = (q_1, q_2, q_3) = (q_1, q_2, q_3^1, q_3^2)$. Note that $\partial(U_1 \cup U_2) = \partial(U_1 \cup U_2)^c$. However, since (q_1, q_2) represents the left-hand side of this equality, and $q_3 = (q_3^1, q_3^1)$ the right-hand side, it only makes sense to impose the constraints $q_1 = q_3^1$ and $q_2 = q_3^2$. This model can be used for instance to study two different shapes or more that are required not to overlap during the deformation. This is exactly what we need to study sub-structures of the brain as mentioned in the introduction (see Figure 1).

In this example, one can even go further: the background does not need to completely folles the movements of the shapes. We can for example let the boundaries slide on each another. This imposes constraints on the speed of the shapes (and not just on the shapes themselves), of the form $C_q K_q u = 0$. See Section 5 for additional details.

Multi-shapes are of great interest in computational anatomy and provide an important motivation to study shape deformation under constraints.

2.4 Finite dimensional approximation of optimal controls

The purpose of this section is to show that at least one solution of Problem 1 can be approximated by a sequence of solutions of a family of nested optimal control problems on finite-dimensional shape spaces with finite-dimensional constraints. We assume throughout that $\ell \geqslant 1$.

Let $(Y^n)_{n \in \mathbb{N}}$ be a sequence of Banach spaces and $(C^n)_{n \in \mathbb{N}}$ be a sequence of continuous mappings $C^n : M \times V \to Y^n$ that are linear and continuous with respect to the second variable. Let $(g^n)_{n \in \mathbb{N}}$ be a sequence of continuous functions on M, bounded from below with a constant independent of n. For every integer n, we consider the problem of minimizing the functional

$$J_1^n(v) = \frac{1}{2} \int_0^1 \|v(t)\|_V^2 dt + g^n(q(1)),$$

over all $v(\cdot) \in L^2(0, 1; V)$ such that $C^n_{q(t)} v(t) = 0$ for almost every $t \in [0, 1]$, where $q(\cdot) : [0, 1] \to M$ is the curve defined by $q(0) = q_0$ and $\dot{q}(t) = \xi_{q(t)} v(t)$ for almost every $t \in [0, 1]$. It follows from Theorem 1 that there exists an optimal solution $v^n(\cdot) \in L^2(0, 1; V)$. We denote by $q^n(\cdot)$ the corresponding curve.

Proposition 5. *Assume that V is an RKHS of vector fields of class $\mathcal{C}_0^{\ell+1}$ on $\mathbb{R}^d$ and that the sequence $(\mathrm{Null}(C_q^n))_{n \in \mathbb{N}}$ is decreasing (in the sense of the inclusion) and satisfies*

$$\bigcap_{n \in \mathbb{N}} \mathrm{Null}(C_q^n) = \mathrm{Null}(C_q),$$

for every $q \in M$. Assume that g^n converges to g uniformly on every compact subset of M. Finally, assume that q_0 has compact support. Then the sequence $(v^n(\cdot))_{n \in \mathbb{N}}$ is bounded in $L^2(0, 1; V)$, and every cluster point of this sequence for the weak topology of $L^2(0, 1; V)$ is an optimal solution of Problem 1. More precisely, for every cluster point $\bar{v}(\cdot)$ of $(v^n(\cdot))_{n \in \mathbb{N}}$, there exists a subsequence such that $(v^{n_j}(\cdot))_{j \in \mathbb{N}}$ converges weakly to $\bar{v}(\cdot) \in L^2(0, 1; V)$, the sequence $(q^{n_j}(\cdot))_{j \in \mathbb{N}}$ of corresponding curves converges uniformly to $\bar{q}(\cdot)$, and $J_1^{n_j}(v^{n_j})$ converges to $\min J_1 = J_1(\bar{v})$ as j tends to $+\infty$, and $\bar{v}(\cdot)$ is a solution of Problem 1.

Proof. The sequence $(v^n(\cdot))_{n\in\mathbb{N}}$ is bounded in $L^2(0,1;V)$ as a consequence of the fact that the functions g^n are uniformly bounded below. Let $\bar{v}(\cdot)$ be a cluster point of this sequence for the weak topology of $L^2(0,1;V)$. Assume that $(v^{n_j}(\cdot))_{j\in\mathbb{N}}$ converges weakly to $\bar{v}(\cdot) \in L^2(0,1;V)$. Denoting by $\bar{q}(\cdot)$ the curve corresponding to $\bar{v}(\cdot)$, the sequence $(q^{n_j}(\cdot))_{j\in\mathbb{N}}$ converges uniformly to $\bar{q}(\cdot)$ (see Lemma 2). Using the property of decreasing inclusion, we have $C^N_{q^{n_j}(\cdot)}v^{n_j}(\cdot) = 0$ for every integer N and every integer $j \geqslant N$. Using the same arguments as in the proof of Theorem 1 (see Section 2.5), it follows that $C_{\bar{q}(\cdot)}\bar{v}(\cdot) = 0$. Finally, since $\int_0^1 \|\bar{v}(t)\|_V^2\,dt \leqslant \liminf \int_0^1 \|v^n(t)\|_V^2\,dt$, and since g^n converges uniformly to g on every compact subset of M, it follows that $J_1(\bar{v}) \leqslant \liminf J_1^{n_j}(v^{n_j})$.

Since every $v \in \mathrm{Null}(C_q)$ belongs as well to $\mathrm{Null}(C_q^{n_j})$, it follows that $J_1^{n_j}(v^{n_j}) \leqslant J_1^{n_j}(v)$, for every time-dependent vector field $v(\cdot) \in L^2(0,1;V)$ such that $C_{q(\cdot)}v(\cdot) = 0$, where $q(\cdot) : [0,1] \to M$ is the curve corresponding to $v(\cdot)$. Since g^n converges uniformly to g, one has $J_1^{n_j}(v) \to J_1(v)$ as $n \to +\infty$. It follows that $\limsup J_1^{n_j}(v^{n_j}) \leqslant \min J_1$.

We have proved that $J_1(\bar{v}) \leqslant \liminf J_1^{n_j}(v^{n_j}) \leqslant \limsup J_1^{n_j}(v^{n_j}) \leqslant \min J_1$, and therefore $J_1(\bar{v}) = \min J_1$, that is, $\bar{v}(\cdot)$ is an optimal solution of Problem 1, and $J_1^{n_j}(v^{n_j})$ converges to $\min J_1 = J_1(\bar{v})$ as j tends to $+\infty$. $\qquad\square$

Application: approximation with finite dimensional shape spaces. Let S^1 be the unit circle of $\mathbb{R}^2$, let $\ell \geqslant 1$ be an integer, $X = \mathcal{C}^\ell(S^1, \mathbb{R}^d)$ and let $M = \mathrm{Emb}^\ell(S^1, \mathbb{R}^d)$ be the space of parametrized simple closed curves of class $\mathcal{C}^\ell$ on $\mathbb{R}^d$. We identify X with the space of all mappings $f \in \mathcal{C}^\ell([0,1], \mathbb{R}^d)$ such that $f(0) = f(1)$, $f'(0) = f'(1)$, ..., $f^{(\ell)}(0) = f^{(\ell)}(1)$. The action of the group of diffeomorphisms of $\mathbb{R}^d$ on M, defined by composition, is smooth of order ℓ (see Section 2.2). Let $q_1 \in M$ and $c > 0$ fixed. We define g by

$$g(q) = c \int_{S^1} |q(t) - q_1(t)|^2\,dt,$$

for every $q \in M$. Consider pointwise kinetic constraints $C : M \times X \to \mathcal{C}^0(S^1, \mathbb{R}^m)$, defined by $(C_q\dot{q})(s) = F_{q(s)}\dot{q}(s)$ for every $s \in S^1$, with $F \in \mathcal{C}^0(\mathbb{R}^d, \mathcal{M}_{m,d}(\mathbb{R}))$, where $\mathcal{M}_{m,d}(\mathbb{R})$ is the set of real matrices of size $m \times d$.

Note that the multishapes constraints described in Section 2.3 are of this form.

Our objective is to approximate this optimal control problem with a sequence of optimal control problems specified on the finite-dimensional shape spaces $\mathrm{Lmk}_d(2^n)$.

For $n \in \mathbb{N}$, let ξ^n be the infinitesimal action of the group of diffeomorphisms on $\mathrm{Lmk}_d(2^n)$, the elements of which are denoted by $q^n = (x_1^n, \ldots, x_{2^n}^n)$. Define on the associated control problem the kinetic constraints $\tilde{C}^n_{q^n}\xi^n_{q_n}v = (F_{x_1^n}v(x_1^n), \ldots, F_{x_{2^n}^n}v(x_{2^n}^n))$, and the data attachment function

$$\tilde{g}^n(q^n) = \frac{c}{2^n} \sum_{r=1}^{2^n} |x_r^n - q_1(2^{-n}r)|.$$

Let $v^n(\cdot)$ be an optimal control of Problem 1 for the above optimal control problem specified on $\mathrm{Lmk}_d(2^n)$.

Proposition 6. *Every cluster point of the sequence $(v^n(\cdot))_{n\in\mathbb{N}}$ for the weak topology on $L^2(0,1;V)$ is a solution of Problem 1 specified on $M = \mathrm{Emb}^\ell(S^1, \mathbb{R}^d)$ with constraints and minimization functional respectively given by C and g defined above.*

Proof. Define the submersions $P^n : M \to \mathrm{Lmk}_d(2^n)$ by

$$P^n(q) = (q(2^{-n}), \ldots, q(2^{-n}r), \ldots, q(1)).$$

Let $g^n = \tilde{g}^n \circ P^n$ and $C_q^n = \tilde{C}_{P^n(q)}dP_q$. In other words, g^n (resp. C^n) are the lifts of $\tilde{g}^n$ (resp. $\tilde{C}^n$) from $\mathrm{Lmk}_d(2^n)$ to M through P^n. Using Remark 9 on lifted shape spaces, we infer that the optimal

18

control $v_n(\cdot)$ of Problem 1 specified on the finite-dimensional space $\mathrm{Lmk}_d(2^n)$ with constraints $\tilde{C}^n$ and data attachment $\tilde{g}^n$ is also optimal for Problem 1 specified on the infinite dimensional set M with constraints C^n and data attachment g^n. Now, if $v \in \mathrm{Null}(C^n)$ for every integer n, then $F_{q(s)}v(q(s)) = 0$ for every $s = 2^{-n}r$ with $n \in \mathbb{N}$ and $r \in \{1,\ldots,2^n\}$. The set of such s is dense in $[0,1]$, and $s \mapsto F_{q(s)}v(q(s))$ is continuous. Therefore $F_{q(s)}v(q(s)) = 0$ for every $s \in [0,1]$, that is, $C_{q(\cdot)}\xi_{q(\cdot)}v(\cdot) = 0$. Since the converse is immediate, we get $\mathrm{Null}(C_{q(\cdot)}) = \bigcap_{n\in\mathbb{N}}\mathrm{Null}(C^n_{q(\cdot)})$. Finally, since $q(\cdot)$ is a closed curve of class at least $\mathcal{C}^\ell$ with $\ell \geqslant 1$, it is easy to check that

$$g^n(q) = \frac{c}{2^n}\sum_{r=1}^{2^n}|q(2^{-n}r) - q_1(2^{-n}r)|$$

converges to g, uniformly on every compact subset of M. Therefore, Proposition 5 can be applied to the sequence $(v^n(\cdot))_{n\in\mathbb{N}}$, which completes the proof. □

Remark 10. The same argument works as well if we replace S^1 with any compact Riemannian manifold $\mathcal{S}$, and applies to the vertices of increasingly finer triangulations of $\mathcal{S}$.

2.5 Proof of Theorem 1

Let $(v_n(\cdot))_{n\in\mathbb{N}}$ be a sequence of $L^2(0,1;V)$ such that $J_1(v_n)$ converges to its infimum. Let $(\varphi_n)_{n\in\mathbb{N}}$ be the corresponding sequence of deformations and let $(q_n(\cdot))_{n\in\mathbb{N}}$ be the sequence of corresponding curves (one has $q_n(t) = \varphi_n(t) \cdot q_0$ thanks to Remark 3). Since g is bounded below, it follows that the sequence $(v_n(\cdot))_{n\in\mathbb{N}}$ is bounded in $L^2(0,1;V)$. The following lemma is well known (see [42]), but we provide a proof for the sake of completeness.

Lemma 2. *There exist $\bar{v}(\cdot) \in L^2(0,1;V)$, corresponding to the deformation $\bar{\varphi}$, and a sequence $(n_j)_{j\in\mathbb{N}}$ of integers such that $(v_{n_j}(\cdot))_{j\in\mathbb{N}}$ converges weakly to $\bar{v}(\cdot)$ and such that, for every compact subset U of $\mathbb{R}^d$,*

$$\sup_{t\in[0,1]}\|(\varphi_{n_j}(t,\cdot) - \bar{\varphi}(t,\cdot))_{|U}\|_\ell \xrightarrow[j\to+\infty]{} 0.$$

Proof of Lemma 2. Since the sequence $(v_n(\cdot))_{n\in\mathbb{N}}$ is bounded in the Hilbert space $L^2(0,1;V)$, there exists a subsequence $(v_{n_j}(\cdot))_{j\in\mathbb{N}}$ converging weakly to some $\bar{v} \in L^2(0,1;V)$. Besides, using (2) for $i = \ell + 1$ and the Ascoli theorem, we infer that for every compact subset U of $\mathbb{R}^d$, the sequence $(\varphi_{n_j})_{j\in\mathbb{N}}$ is contained in a compact subset of the space $\mathcal{C}^0([0,1],\mathcal{C}^\ell_0(U,\mathbb{R}^d))$. Considering a compact exhaustion of $\mathbb{R}^d$ and using a diagonal extraction argument, we can therefore extract a subsequence $(\varphi_{n_{j_k}})_{k\in\mathbb{N}}$ with limit $\bar{\varphi}$ such that, for any compact subset U of $\mathbb{R}^d$,

$$\sup_{t\in[0,1]}\|(\varphi_{n_{j_k}}(t,\cdot) - \bar{\varphi}(t,\cdot))_{|U}\|_\ell \xrightarrow[k\to+\infty]{} 0. \tag{15}$$

To complete the proof, it remains to prove that $\bar{\varphi}$ is the deformation induced by $\bar{v}(\cdot)$. On the first hand, we have $\lim_{k\to+\infty}\varphi_{n_{j_k}}(t,x) = \bar{\varphi}(t,x)$, for every $x \in \mathbb{R}^d$ and every $t \in [0,1]$. On the second hand, one has, for every $k \in \mathbb{N}$,

$$\left|\varphi_{n_{j_k}}(t,x) - x - \int_0^t \bar{v} \circ \bar{\varphi}(s,x)\,ds\right| = \left|\int_0^t \left(v_{n_{j_k}} \circ \varphi_{n_{j_k}}(s,x) - \bar{v} \circ \bar{\varphi}(s,x)\right)ds\right|$$

$$\leqslant \left|\int_0^t v_{n_{j_k}} \circ \varphi_{n_{j_k}}(x) - v_{n_{j_k}} \circ \bar{\varphi}(s,x)\,ds\right| + \left|\int_0^t v_{n_{j_k}} \circ \bar{\varphi}(s,x) - \bar{v} \circ \bar{\varphi}(s,x)\,ds\right|.$$

Set $C = \sup_{n \in \mathbb{N}} \int_0^1 \|v_n(t)\|_1 \, dt$, and define $\chi_{[0,t]} \delta_{\bar{\varphi}(\cdot, x)} \in V^*$ by $(\chi_{[0,t]} \delta_{\bar{\varphi}(\cdot, x)} | v)_{L^2(0,1;V)} = \int_0^t v \circ \bar{\varphi}(s, x) \, ds$. Then,

$$
\left| \varphi_{n_{j_k}}(t, x) - x - \int_0^t \bar{v} \circ \bar{\varphi}(s, x) \, ds \right| \leqslant C \sup_{s \in [0,t]} |\varphi_{n_{j_k}}(s, x) - \bar{\varphi}(s, x)|
$$
$$
+ \left| (\chi_{[0,t]} \delta_{\bar{\varphi}(\cdot, x)} | v_{n_{j_k}} - \bar{v})_{L^2(0,1;V)} \right|,
$$

which converges to 0 as j tends to $+\infty$ thanks to (15) and to the weak convergence $v_{n_{j_k}}(\cdot)$ to $\bar{v}(\cdot)$. We thus conclude that $\bar{\varphi}(t, x) = x + \int_0^t \bar{v} \circ \bar{\varphi}(s, x) \, ds$, which completes the proof. $\qquad\square$

Setting $\bar{q}(t) = \bar{\varphi}(t) \cdot q_0$ for every $t \in [0, 1]$, one has $\dot{\bar{q}}(t) = \xi_{\bar{q}(t)} \bar{v}(t)$ for almost every $t \in [0, 1]$, and it follows from the above lemma and from the fact that q_0 has compact support that

$$
\sup_{t \in [0,1]} \|\bar{q}(t) - q_{n_j}(t)\|_X \xrightarrow[j \to +\infty]{} 0.
$$

The operator $v(\cdot) \mapsto C_{\bar{q}(\cdot)} v(\cdot)$ is linear and continuous on $L^2(0, 1; V)$, so it is also weakly continuous [16]. We infer that the sequence $C_{q_{n_j}(\cdot)} v_{n_j}(\cdot)$ converges weakly to $C_{\bar{q}(\cdot)} \bar{v}(\cdot)$ in $L^2(0, 1; Y)$. Since $C_{q_{n_j}(\cdot)} v_{n_j}(\cdot) = 0$ for every $j \in \mathbb{N}$, it follows that $C_{\bar{q}(\cdot)} \bar{v}(\cdot) = 0$. In other words, the time-dependent vector field $\bar{v}(\cdot)$ satisfies the constraints.

It remains to prove that $\bar{v}(\cdot)$ is indeed optimal. From the weak convergence of the sequence $(v_{n_j}(\cdot))_{j \in \mathbb{N}}$ to $\bar{v}(\cdot)$ in $L^2(0, 1; V)$, we infer that

$$
\int_0^1 \|\bar{v}(t)\|_V^2 \, dt \leqslant \liminf_{j \to +\infty} \int_0^1 \|v_{n_j}(t)\|_V^2 \, dt.
$$

Besides, since g is lower continuous, $\liminf_{j \to +\infty} g(q_{n_j}(1)) \geqslant g(\bar{q}(1))$. Since $J_1(v_{n_j})$ converges to $\inf J_1$, it follows that $J_1(\bar{v}) = \inf J_1$.

3 Constrained geodesic equations in shape spaces

In this section, we derive first-order necessary conditions for optimality in Problem 1. We extend the well-known Pontryagin maximum principle (PMP) from optimal control theory to our infinite-dimensional framework, under the assumption that the constraints are surjective. This allows us to derive the constrained geodesic equations for shape spaces, and we show how they can be reduced to simple Hamiltonian dynamics on the cotangent space of the shape space.

3.1 First-order optimality conditions: PMP in shape spaces

We address the Pontyagin maximum principle in a slightly extended framework, considering a more general control system and a more general minimization functional than in Problem 1.

Let V be a Hilbert space, and let X and Y be Banach spaces. Let M be an open subset of X. Let $\xi : M \times V \to X$ and $C : M \times V \to Y$ be mappings of class $\mathcal{C}^1$. Let $L : M \times V \to \mathbb{R}$ and $g : M \to \mathbb{R}$ be functions of class $\mathcal{C}^1$. We assume that there exist continuous functions $\gamma_0 : \mathbb{R} \to \mathbb{R}$ and $\gamma_1 : X \to \mathbb{R}$ such that $\gamma_0(0) = 0$ and

$$
\|dL_{q', v'} - dL_{q, v}\|_{X^* \times V^*} \leqslant \gamma_0(\|q' - q\|_X) + \gamma_1(q - q')\|v' - v\|_V, \tag{16}
$$

for all $(q, q') \in M^2$ and all $(v, v') \in V^2$.

Let $q_0 \in M$. We consider the optimal control problem of minimizing the functional

$$J(q, v) = \int_0^1 L(q(t), v(t)) \, dt + g(q(1)) \tag{17}$$

over all $(q(\cdot), v(\cdot)) \in H_{q_0}^1(0, 1; M) \times L^2(0, 1; V)$ such that $\dot{q}(t) = \xi_{q(t)} v(t)$ and $C_{q(t)} v(t) = 0$ for almost every $t \in [0, 1]$. We define the *Hamiltonian* $H : M \times X^* \times V \times Y^* \to \mathbb{R}$ by

$$H(q, p, v, \lambda) = \langle p, \xi_q v \rangle_{X^*, X} - L(q, v) - \langle \lambda, C_q v \rangle_{Y^*, Y}. \tag{18}$$

It is a function of class $\mathcal{C}^1$. Using the canonical injection $X \hookrightarrow X^{**}$, we have $\partial_p H = \xi_q v$.

Remark 11. The estimate (16) on L is exactly what is required to ensure that the mapping $(q, v) \mapsto \int_0^1 L(q(t), v(t)) \, dt$ be well defined and Fréchet differentiable for every $(q, v) \in H_{q_0}^1(0, 1; M) \times L^2(0, 1; V)$. Indeed, the estimate $L(q(t), v(t)) \leqslant L(q(t), 0) + \gamma_1(q(t), q(t)) \|v(t)\|_V^2$ implies the integrability property. The differentiability is an immediate consequence of the following estimate, obtained by combining (16) with the mean value theorem: for every $t \in [0, 1]$, and for some $s_t \in [0, 1]$, one has

$$|L(q(t) + \delta q(t), v(t) + \delta v(t)) - L(q(t), v(t) - dL_{q(t), v(t)}(\delta q(t), \delta v(t))|$$
$$\leqslant \Big(\gamma_0(\|\delta q(t)\|_X) + \gamma_1(s_t \delta q(t)) \|\delta v(t)\| \Big) (\|\delta q(t)\|_X + \|\delta v(t)\|_V).$$

Theorem 2. *Assume that the linear operator $C_{q(t)} : V \to Y$ is surjective for every $q \in M$. Let $(q(\cdot), v(\cdot)) \in H_{q_0}^1(0, 1; X) \times L^2(0, 1; V)$ be an optimal solution of the above optimal control problem. Then there exist $p(\cdot) \in H^1(0, 1; X^*)$ and $\lambda(\cdot) \in L^2(0, 1; Y^*)$ such that $p(1) + dg_{q(1)} = 0$ and*

$$\begin{aligned}
\dot{q}(t) &= \partial_p H(q(t), p(t), v(t), \lambda(t)), \\
\dot{p}(t) &= -\partial_q H(q(t), p(t), v(t), \lambda(t)), \\
\partial_v H(q(t), p(t), v(t), \lambda(t)) &= 0,
\end{aligned} \tag{19}$$

for almost every $t \in [0, 1]$.

Remark 12. This theorem is the extension of the usual PMP to our specific infinite-dimensional setting. Any quadruple $q(\cdot), p(\cdot), v(\cdot), \lambda(\cdot)$ solution of the above equations is called an *extremal*. This is a "weak" maximum principle, in the sense that we derive the condition $\partial_v H$ along any extremal, instead of the stronger maximization condition

$$H(q(t), p(t), v(t), \lambda(t)) = \max_{w \in \text{Null}(C_{q(t)})} H(q(t), p(t), w, \lambda(t))$$

for almost every $t \in [0, 1]$. Note however that, in the case of shape spaces, $v \mapsto H(q, p, v)$ is strictly concave and hence both conditions are equivalent.

Remark 13. It is interesting to note that, if we set $V = H^\ell(\mathbb{R}^d, \mathbb{R}^d)$ with ℓ large enough, and if $L(q, v) = \frac{1}{2} \int_{\mathbb{R}^d} |v(x)|^2 dx$ and $C_q v = \text{div} \, v$, then the extremal equations given in Theorem 2 coincide with the incompressible Euler equation. In other words, we recover the well-known fact that every divergence-free time-dependent vector field minimizing its L^2 norm (in t and x) must satisfy the incompressible Euler equation (see [5]).

Remark 14. Note that the surjectivity assumption is a strong one in infinite dimension. It is usually not satisfied in the case of shape spaces when Y is infinite dimensional. For instance, consider the shape spaces $M = \mathcal{C}^0(S, \mathbb{R}^d)$, with S a smooth compact Riemannian manifold. Let V be an RKHS of vector fields of class $\mathcal{C}_0^1$, acting on M as described in Section 2.2. Let T be a submanifold of S of class $\mathcal{C}^1$. Set $Y = \mathcal{C}^0(T, \mathbb{R}^d)$, and consider the kinetic constraints $C : M \times V \to Y$ defined by $C_q v = v \circ q_{|T}$. If q is differentiable along T, then no nondifferentiable map $f \in Y$ is in $\text{Range}(C_q)$.

Remark 15. It is possible to replace C_q with the orthogonal projection on $\text{Null}(C_q)^\perp$, which is automatically surjective. However in this case the constraints become fiber-valued, and for the proof of our theorem to remain valid, one needs to assume that there exists a Hilbert space V_1 such that, for every $q_0 \in M$, there exists a neighborhood U of q_0 in M such that $\bigcup_{q \in U} \text{Null}(C_q) \sim U \times V_1$.

Remark 16. An important consequence of the surjectivity of C_q is that the norm on Y is equivalent to the Hilbert norm induced from V by C_q, for every $q \in M$. The operator $C_q K_V C_q^* : Y^* \to Y$ is the isometry associated with this norm. In particular Y must be reflexive, and hence $L^2(0,1;Y^*) = L^2(0,1;Y)^*$ (whereas we only have an inclusion for general Banach spaces).

Remark 17. Theorem 2 withstands several generalizations. For instance it remains valid whenever we consider nonlinear constraints $C(q,v) = 0$ and a general Lagrangian L of class $\mathcal{C}^1$, with the following minor modifications.

- We must restrict ourselves to essentially bounded controls $v \in L^\infty(0,1;V)$ (instead of square-integrable controls). However, the estimate (16) on L is no longer required and we only need to assume that L is of class $\mathcal{C}^1$.

- The assumption "C_q surjective" must be replaced with "$\partial_v C(q,v)$" surjective for every $(q,v) \in M \times V$.

- $q(\cdot) \in W^{1,\infty}(0,1;X)$, $p(\cdot) \in W^{1,\infty}(0,1;X^*)$, and $\lambda(\cdot) \in L^\infty(0,1;Y^*)$.

The proof is similar to that of Theorem 2, although the regularity of the Lagrange multipliers $p(\cdot)$ and $\lambda(\cdot)$ is slightly more difficult to obtain.

Before proving Theorem 2, it can be noted that many versions of the PMP can be found in the existing literature for infinite-dimensional optimal control problems – for instance, with dynamics consisting of partial differential equations, and however, most of the time, without constraint on the state. Versions of the PMP with state constraints can also be found in the literature (see the survey [23]), most of the time in finite dimension, and, for the very few of them existing in infinite dimension, under the additional assumption that the constraints are of finite codimension. To the best of our knowledge, no version does exist that would cover our specific framework, concerning shape spaces, group actions, with an infinite number of constraints on the acting diffeomorphisms. The proof that we provide hereafter contains some subtleties such as Lemma 4, and hence Theorem 2 is a nontrivial extension of the usual PMP.

Proof of Theorem 2. We define the mapping $\Gamma : H^1_{q_0}(0,1;M) \times L^2(0,1;V) \to L^2(0,1;X) \times L^2(0,1;Y)$ by $\Gamma(q,v) = (\Gamma_1(q,v), \Gamma_2(q,v))$ with $\Gamma_1(q,v)(t) = \dot{q}(t) - \xi_{q(t)} v(t)$ and $\Gamma_2(q,v)(t) = C_{q(t)} v(t)$ for almost every $t \in [0,1]$. The mapping Γ stands for the constraints imposed to the unknowns of the optimal control problem.

The functional $J : H^1_{q_0}(0,1;M) \times L^2(0,1;V) \to \mathbb{R}$ and the mapping Γ are of class $\mathcal{C}^1$, and their respective differentials at some point (q,v) are given by

$$dJ_{(q,v)}.(\delta q, \delta v) = \int_0^1 \left(\langle \partial_q L(q(t), v(t)), \delta q(t) \rangle_{X^*,X} + \langle \partial_v L(q(t), v(t)), \delta v(t) \rangle_{V^*,V} \right) dt$$
$$+ \langle dg_{q(1)}, \delta q(1) \rangle_{X^*,X},$$

for all $(\delta q, \delta v) \in H^1_0(0,1;X) \times L^2(0,1;V)$, and $d\Gamma_{(q,v)} = (d\Gamma_{1(q,v)}, d\Gamma_{2(q,v)})$ with

$$(d\Gamma_{1(q,v)}.(\delta q, \delta v))(t) = \dot{\delta q}(t) - \partial_q \xi_{q(t)} v(t).\delta q(t) - \xi_{q(t)} \delta v(t),$$
$$(d\Gamma_{2(q,v)}.(\delta q, \delta v))(t) = \partial_q C_{q(t)} v(t).\delta q(t) + C_{q(t)} \delta v(t),$$

for almost every $t \in [0,1]$.

Lemma 3. *For every* $(q, v) \in H^1_{q_0}(0, 1; M) \times L^2(0, 1; V)$, *the linear continuous mapping* $d\Gamma_{(q,v)}$: $H^1_0(0, 1; X) \times L^2(0, 1; V) \to L^2(0, 1; X) \times L^2(0, 1; Y)$ *is surjective. Moreover the mapping* $\partial_q \Gamma_{1(q,v)}$: $H^1_0(0, 1; M) \to L^2(0, 1; X)$ *is an isomorphism.*

Proof. Let $(q, v) \in H^1_{q_0}(0, 1; M) \times L^2(0, 1; V)$ and $(a, b) \in L^2(0, 1; X) \times L^2(0, 1; Y)$. Let us prove that there exists $(\delta q, \delta v) \in H^1_0(0, 1; X) \times L^2(0, 1; V)$ such that

$$a(t) = \dot{\delta q}(t) - \partial_q \xi_{q(t)} v(t).\delta q(t) - \xi_{q(t)} \delta v(t), \tag{20}$$

$$b(t) = \partial_q C_{q(t)} v(t).\delta q(t) + C_{q(t)} \delta v(t), \tag{21}$$

for almost every $t \in [0, 1]$.

For every $\tilde{q} \in M$, $C_{\tilde{q}} : V \to Y$ is a surjective linear continuous mapping. It follows that $C_{\tilde{q}|(\mathrm{Null}(C_{\tilde{q}}))^\perp} : (\mathrm{Null}(C_{\tilde{q}}))^\perp \to Y$ is an isomorphism (note that $V = \mathrm{Null}(C_{\tilde{q}}) \oplus (\mathrm{Null}(C_{\tilde{q}}))^\perp$ since V is Hilbert). We set $A_{\tilde{q}} = (C_{\tilde{q}|(\mathrm{Null}(C_{\tilde{q}}))^\perp})^{-1} = K_V C^*_{\tilde{q}} (C_{\tilde{q}} K_V C^*_{\tilde{q}})^{-1}$. Note that $\tilde{q} \mapsto A_{\tilde{q}}$ is of class $\mathcal{C}^1$ in a neighbourhood of $q([0, 1])$.

Assume for the moment that $\delta q(\cdot)$ is known. Then we choose $\delta v(\cdot)$ defined by $\delta v(t) = A_{q(t)} (b(t) - \partial_q C_{q(t)}.\delta q(t))$ for almost every $t \in [0, 1]$, so that (21) is satisfied. Plugging this expression into (20) yields

$$\dot{\delta q}(t) - \partial_q \xi_{q(t)} v(t).\delta q(t) - \xi_{q(t)} A_{q(t)} (b(t) - \partial_q C_{q(t)} v(t).\delta q(t)) = a(t),$$

for almost every $t \in [0, 1]$. This is a well-posed linear differential equation with square-integrable coefficients in the Banach space X, which has a unique solution $\delta q \in H^1_0(0, 1; X)$ such that $\delta q(0) = 0$. This proves the statement.

Proving that the mapping $\partial_q \Gamma_{1(q,v)} : H^1_0(0, 1; M) \to L^2(0, 1; X)$, defined by $(\partial_q \Gamma_{1(q,v)}.\delta q)(t) = \dot{\delta q}(t) - \partial_q \xi_{q(t)} v(t).\delta q(t)$ for almost every $t \in [0, 1]$, is an isomorphism follows the same argument, by Cauchy uniqueness. $\qquad\square$

Let $(q, v) \in H^1_{q_0}(0, 1; M) \times L^2(0, 1; V)$ be an optimal solution of the optimal control problem. In other words, (q, v) is a minimizer of the problem of minimizing the functional J over the set of constraints $\Gamma^{-1}(\{0\})$ (which is a $\mathcal{C}^\infty$ manifold as a consequence of Lemma 3 and of the implicit function theorem). Since $d\Gamma_{(q,v)}$ is surjective (note that this fact is essential since we are in infinite dimension), it follows from [27, Theorem 4.1] that there exists a nontrivial Lagrange multiplier $(p, \lambda) \in L^2(0, 1; X)^* \times L^2(0, 1; Y)^*$ such that $dJ_{(q,v)} + (d\Gamma_{(q,v)})^*(p, \lambda) = 0$. Moreover since Y is reflexive one has $L^2(0, 1; Y)^* = L^2([0, 1], Y^*)$, and hence we can identify λ with a square-integrable Y^*-valued measurable mapping, so that the Lagrange multipliers relation yields

$$
\begin{aligned}
0 &= \left\langle dJ_{(q,v)} + (d\Gamma_{(q,v)})^*(p, \lambda), (\delta q, \delta v) \right\rangle \\
&= \left\langle p, \dot{\delta q} \right\rangle_{L^2(0,1;X)^*, L^2(0,1;X)} - \left\langle p, \partial_q \xi_q v.\delta q \right\rangle_{L^2(0,1;X)^*, L^2(0,1;X)} - \left\langle p, \xi_q \delta v \right\rangle_{L^2(0,1;X)^*, L^2(0,1;X)} \\
&\quad + \int_0^1 \left\langle \partial_q L(q(t), v(t)) + (\partial_q C_{q(t)} v(t))^* \lambda(t), \delta q(t) \right\rangle_{X^*, X} dt \\
&\quad + \int_0^1 \left\langle \partial_v L(q(t), v(t)) + C^*_{q(t)} \lambda(t), \delta v(t) \right\rangle_{V^*, V} dt + \left\langle dg_{q(1)}, \delta q(1) \right\rangle_{X^*, X},
\end{aligned}
\tag{22}
$$

for all $(p, \lambda) \in L^2(0, 1; X)^* \times L^2([0, 1], Y^*)$ and all $(\delta q, \delta v) \in H^1_0(0, 1; M) \times L^2(0, 1; V)$. Note that the space $L^2(0, 1; X)^*$ can be different from $L^2(0, 1; X^*)$ (unless X^* satisfies the Radon-Nikodym property, but there is no reason to consider such a Banach space X), and hence *a priori* p cannot be obviously identified with a square-integrable X^*-valued measurable mapping. Anyway, in the next lemma we show that this identification is possible, due to a hidden regularity property in (22).

Lemma 4. *We can identify p with an element of $L^2([0,1], X^*)$, so that*

$$\langle p, r \rangle_{L^2(0,1;X^*), L^2(0,1;X)} = \int_0^1 \langle p(t), r(t) \rangle_{X^*, X} \, dt,$$

for every $r \in L^2(0,1;X)$.

Proof of Lemma 4. For every $r \in L^2(0,1;X)$, we define $\delta q \in H_0^1(0,1;X)$ by $\delta q(s) = \int_0^s r(t) \, dt$ for every $s \in [0,1]$ (Bochner integral in the Banach space X), so that $r = \dot{\delta q}$. Defining $\alpha \in L^2(0,1;X)^*$ by

$$
\begin{aligned}
\langle \alpha, f \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} = {} & \langle p, \partial_q \xi_q v . f \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} \\
& - \int_0^1 \langle \partial_q L(q(t), v(t)) + (\partial_q C_{q(t)} v(t))^* \lambda(t), f(t) \rangle_{X^*, X} \, dt,
\end{aligned}
\tag{23}
$$

for every $f \in L^2(0,1;X)$, and taking $\delta v = 0$ in (22), we get

$$\langle p, r \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} = \langle \alpha, \delta q \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} - \langle dg_{q(1)}, \delta q(1) \rangle_{X^*, X}. \tag{24}$$

Let us express δq in another way with respect to r. By definition, one has $\delta q(s) = \int_0^s r(t) \, dt$ for every $s \in [0,1]$, and this can be also written as $\delta q(s) = \int_0^1 \chi_{[t,1]}(s) r(t) \, dt$, with $\chi_{[t,1]}(s) = 1$ whenever $s \in [t,1]$ and 0 otherwise. In other words, one has $\delta q = \int_0^1 \chi_{[t,1]} r(t) \, dt$ (Bochner integral). For every $t \in [0,1]$, we define the operator $A_t : X \to L^2(0,1;X)$ by $A_t x = \chi_{[t,1]} x$. It is clearly linear and continuous. Then, we have $\delta q = \int_0^1 A_t(r(t)) \, dt$, and therefore, using (24),

$$\langle p, r \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} = \left\langle \alpha, \int_0^1 A_t(r(t)) \, dt \right\rangle_{L^2(0,1;X)^*, L^2(0,1;X)} - \left\langle dg_{q(1)}, \int_0^1 r(t) \, dt \right\rangle_{X^*, X}.$$

Now, interchanging the Bochner integrals and the linear forms, we infer that

$$\langle p, r \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} = \int_0^1 \langle \alpha, A_t(r(t)) \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} \, dt - \int_0^1 \langle dg_{q(1)}, r(t) \rangle_{X^*, X} \, dt,$$

and then, using the adjoint $A_t^* : L^2(0,1;X)^* \to X^*$, we get

$$\langle p, r \rangle_{L^2(0,1;X)^*, L^2(0,1;X)} = \int_0^1 \langle A_t^* \alpha - dg_{q(1)}, r(t) \rangle_{X^*, X} \, dt.$$

Since this identity holds true for every $r \in L^2(0,1;X)$, it follows that p can be identified with an element of $L^2(0,1;X^*)$, still denoted by p, with $p(t) = A_t^* \alpha - dg_{q(1)}$ for almost every $t \in [0,1]$. $\square$

Still using the notation introduced in the proof of Lemma 4, now that we know that $p \in L^2(0,1;X^*)$, we infer from (23) that α can as well be identified with an element of $L^2([0,1], X^*)$, with

$$\alpha(t) = (\partial_q \xi_{q(t)} v(t))^* p(t) - \partial_q L(q(t), v(t)) - (\partial_q C_{q(t)} v(t))^* \lambda(t),$$

for almost every $t \in [0,1]$. Note that $\alpha(t) = \partial_q H(q(t), p(t), v(t), \lambda(t))$, where the Hamiltonian H is defined by (18).

Since $\alpha \in L^2(0,1;X^*)$, we have, for every $x \in X$,

$$
\begin{aligned}
\langle A_t^* \alpha, x \rangle_{X^*, X} = \langle \alpha, \chi_{[t,1]} x \rangle_{L^2(0,1,X^*), L^2(0,1;X)} = {} & \int_0^1 \langle \alpha(s), \chi_{[t,1]}(s) x \rangle_{X^*, X} \, ds \\
= {} & \int_t^1 \langle \alpha(s), x \rangle_{X^*, X} \, ds = \left\langle \int_t^1 \alpha(s) \, ds, x \right\rangle_{X^*, X},
\end{aligned}
$$

and therefore $A_t^* \alpha = \int_t^1 \alpha(s)\, ds$ for every $t \in [0,1]$. It follows that $p(t) = A_t^* \alpha - dg_{q(1)} = \int_t^1 \alpha(s)\, ds - dg_{q(1)}$, and hence, that $p \in H^1(0,1;X^*)$ and that $p(\cdot)$ satisfies the differential equation $\dot{p}(t) = -\alpha(t) = -\partial_q H(q(t), p(t), v(t), \lambda(t))$ for almost every $t \in [0,1]$ and $p(1) + dg_{q(1)} = 0$.

Finally, taking $\delta q = 0$ in (22) yields

$$\int_0^1 \langle \partial_v L(q(t), v(t)) + C_{q(t)}^* \lambda(t) - \xi_{q(t)}^* p(t), \delta v(t) \rangle_{V^*, V}\, dt = 0,$$

for every $\delta v \in L^2(0,1;X)$, and hence $\partial_v L(q(t), v(t)) + C_{q(t)}^* \lambda(t) - \xi_{q(t)}^* p(t) = 0$ for almost every $t \in [0,1]$, which exactly means that $\partial_v H(q(t), p(t), v(t), \lambda(t)) = 0$. The theorem is proved. $\qquad\square$

3.2 The geodesic equations in a shape space

We use the notations introduced in Section 2.2, and consider a shape space M of order $\ell \in \mathbb{N}$, V an RKHS of vector fields of class $\mathcal{C}_0^{\ell+1}$ on $\mathbb{R}^d$, and we set $L(q,v) = \frac{1}{2}\|v\|_V^2$. We assume that C and g are at least of class $\mathcal{C}^1$ and that $C_q : V \to Y$ is surjective for every $q \in M$.

In this context ξ is of class $\mathcal{C}^1$. Let us apply Theorem 2. We have $\partial_v H(q,p,v,\lambda) = \xi_q^* p - C_q^* \lambda - (v, \cdot)_V$, and the condition $\partial_v H(q,p,v,\lambda) = 0$ is equivalent to $v = K_V(\xi_q^* p - C_q^* \lambda)$. Then we have $\partial_p H = \xi_q v = \xi_q K_V(\xi_q^* p - C_q^* \lambda) = K_q p - \xi_q K_V C_q^* \lambda$, where $K_q = \xi_q K_V \xi_q^*$ is defined by (11). Besides, $C_q v = C_q K_V \xi_q^* p - C_q K_V C_q^* \lambda = 0$ if and only if $C_q K_V C_q^* \lambda = C_q K_V \xi_q^* p$. Since C_q is surjective, it follows that $C_q K_V C_q^*$ is invertible, and hence $\lambda = \lambda_{q,p} = (C_q K_V C_q^*)^{-1} C_q K_V \xi_q^* p$. The mapping $(q,p) \mapsto \lambda_{q,p}$ defined as such is of class $\mathcal{C}^1$ and is linear in p. In particular, $v = v_{q,p} = K_V(\xi_q^* p - C_q^* \lambda_{q,p})$ is a function of class $\mathcal{C}^1$ of q and p and is linear in p. We have obtained the following result.

Theorem 3 (Geodesic equations in shape spaces). *Let* $(q(\cdot), v(\cdot)) \in H_{q_0}^1(0,1;M) \times L^2(0,1;V)$ *be a solution of Problem 1. There exists* $p(\cdot) \in H^1(0,1;X^*)$ *such that*

$$v(t) = v_{q(t), p(t)} = K_V\left(\xi_{q(t)}^* p(t) - C_{q(t)}^* \lambda(t) \right),$$

for almost every $t \in [0,1]$*, and* $p(\cdot)$ *satisfies* $p(1) + dg_{q(1)} = 0$ *and the geodesic equations*

$$
\begin{aligned}
\dot{q}(t) &= K_{q(t)} p(t) - \xi_{q(t)} K_V C_{q(t)}^* \lambda(t), \\
\dot{p}(t) &= -\partial_q \langle p(t), \xi_{q(t)} v(t) \rangle_{X^*, X} + \partial_q \langle \lambda(t), C_{q(t)} v(t) \rangle_{Y^*, Y},
\end{aligned}
\tag{25}
$$

for almost every $t \in [0,1]$*, with*

$$\lambda(t) = \lambda_{q(t), p(t)} = (C_{q(t)} K_V C_{q(t)}^*)^{-1} C_{q(t)} K_V \xi_{q(t)}^* p(t).$$

Moreover the mapping $t \mapsto \frac{1}{2}\|v(t)\|^2$ *is constant, and one has*

$$J(v) = \frac{1}{2}\|v(0)\|_V^2 + g(q(1)).$$

Remark 18. It is interesting to note that, using an integration by parts, one can prove that curves $q(\cdot)$, for which such $p(\cdot)$ and $\lambda(\cdot)$ exist, are actually geodesics (i.e., critical points of the energy) with respect to the sub-Riemannian structure on M mentioned in Remark 4.

Remark 19. Defining the so-called reduced Hamiltonian $h : M \times X^* \to \mathbb{R}$ by

$$h(q, p) = H(q, p, v_{q,p}, \lambda_{q,p}),$$

we have a priori $\partial_q h = \partial_q H + \partial_v H(q, p, v_{q,p}, \lambda_{q,p}) \circ \partial_q(v_{q,p}) + \partial_\lambda H(q, p, v_{q,p}, \lambda_{q,p}) \circ \partial_q(\lambda_{q,p})$. But since $\lambda_{q,p}$ and $v_{q,p}$ are such that $\partial_v H(q, p, v_{q,p}, \lambda_{q,p}) = 0$ and $C_q v_{q,p} = \partial_\lambda H(q, p, v_{q,p}, \lambda_{q,p}) = 0$, it follows that $\partial_q h(q, p) = \partial_q H(q, p, v_{q,p}, \lambda_{q,p})$. Similarly, we have $\partial_p h(q, p) = \partial_p H(q, p, v_{q,p}, \lambda_{q,p})$. Therefore, in Theorem 3, the geodesics are the solutions of the Hamiltonian system

$$\dot{q}(t) = \partial_p h(q(t), p(t)), \quad \dot{p}(t) = -\partial_q h(q(t), p(t)).$$

Corollary 1. *Assume that the mappings C and ξ are of class $\mathcal{C}^2$. Then h is of class $\mathcal{C}^2$ as well, and for every $(q_0, p_0) \in M \times X^*$, there exists $\varepsilon > 0$ and there exists a unique solution $(q, p) : [0, \epsilon] \to M \times X^*$ of the geodesic equations (25) such that $(q(0), p(0)) = (q_0, p_0)$.*

Note that for most of shape spaces (at least, for all shape spaces given as examples in this paper), the mapping ξ is of class $\mathcal{C}^2$ whenever V is an RKHS of vector fields of class $\mathcal{C}^{\ell+2}$.

Example 5. A geodesic $(q(t), p(t)) = (x_1(t), \ldots, x_n(t), p_1(t), \ldots, p_n(t))$ on the landmark space $\mathrm{Lmk}_d(n)$ must satisfy the equations

$$\dot{x}_i(t) = \sum_{j=1}^{n} K(x_i(t), x_j(t)) p_j(t), \quad \dot{p}_j(t) = -\frac{1}{2} \sum_{j=1}^{n} \partial_{x_i} \left(p_i(t)^T K(x_i(t), x_j(t)) p_j(t) \right),$$

where K is the kernel of V.

Remark 20. Assume that the constraints are kinetic, i.e., are of the form $C_q \xi_q v = 0$ with $C_q : X \to Y$. Then $\dot{q}(t) = K_{q(t)}(p(t) - C_{q(t)}^* \lambda(t))$ and $\lambda_{q,p} = (C_q K_q C_q^*)^{-1} C_q K_q p$. Note that even if K_q is not invertible, the assumption that $C_q \xi_q$ is surjective means that $C_q K_q C_q^*$ is invertible. In particular, minimizing controls take the form $K_V \xi_q^* u$, with $u \in L^2(0, 1; X^*)$. This proves the contents of Proposition 4 (see Section 2.2).

Remark 21. Let M (resp., M'), open subset of a Banach space X (resp., X'), be a shape space of order ℓ (resp. of order $\ell' \leqslant \ell$). Assume that there is a dense and continuous inclusion $X \hookrightarrow X'$, such that $M \hookrightarrow M'$ is equivariant (in particular we have $X'^* \hookrightarrow X^*$). For every $q_0 \in M \subset M'$, there are more geodesics emanating from q_0 on M than on M' (indeed it suffices to consider initial momenta $p_0 \in X^* \setminus X'^*$). These curves are not solutions of the geodesic equations on M', and are an example of so-called *abnormal extremals* [1, 35]. Note that they are however not solutions of Problem 1 specified on X', since Theorem 2 implies that such solutions are projections of geodesics having an initial momentum in X'^*.

An example where this situation is encountered is the following. Let S be a compact Riemannian manifold. Consider $X = \mathcal{C}^{\ell+1}(S, \mathbb{R}^d)$ and $X' = \mathcal{C}^\ell(S, \mathbb{R}^d)$, with actions defined in Definition 2. If $p(0)$ is a $\ell + 1$-th order distribution, then the geodesic equations on X with initial momentum $p(0)$ yield an abnormal geodesic in X'.

Remark 22. Following Remark 21, if the data attachment function $g : X \to \mathbb{R}^+$ can be smoothly extended to a larger space X', then the initial momentum of any solution of Problem 1 is actually in $X'^* \subset X^*$.

For example, we set $g(q) = \int_S |q(s) - q_{target}(s)|^2 \, ds$, for some $q_{target} \in X$, with $X = \mathcal{C}^\ell(S, \mathbb{R}^d)$ and S a compact Riemannian manifold. Let $X' = L^2(S, \mathbb{R}^d)$. Then the momentum $p : [0, 1] \to X^*$ associated with a solution of Problem 1 (whose existence follows from Theorem 3) takes its values in $X'^* = L^2(S, \mathbb{R}^d)$.

26

The case of pure state constraints. Let us consider pure state constraints, i.e., constraints of the form $C(q) = 0$ with $C : M \to Y$. Recall that, if C is of class $\mathcal{C}^1$, then they can be transformed into the mixed constraints $dC_q.\dot{q} = dC_q.\xi_q v = 0$. In this particular case, the geodesic equations take a slightly different form.

Proposition 7. *Assume that $C : M \to Y$ is of class $\mathcal{C}^3$, and that $dC_q.\xi_q : V \to Y$ is surjective for every $q \in M$. Consider Problem 1 with the pure state constraints $C(q) = 0$. If $v(\cdot) \in L^2(0, 1; V)$ is a solution of Problem 1, associated with the curve $q(\cdot) : [0, 1] \to M$, then there exists $\tilde{p} : [0, 1] \to X^*$ such that*

$$\dot{q}(t) = K_{q(t)}\tilde{p}(t), \quad \dot{\tilde{p}}(t) = -\frac{1}{2}\partial_q \langle \tilde{p}(t), K_{q(t)}\tilde{p}(t)\rangle_{X^*,X} + \partial_q \langle \tilde{\lambda}_{q(t),\tilde{p}(t)}, C(q(t))\rangle_{Y^*,Y},$$

for almost every $t \in [0, 1]$, with

$$\tilde{\lambda}_{q,\tilde{p}} = (dC_q K_q dC_q^*)^{-1}\left(\frac{1}{2}dC_q.K_q\partial_q\langle\tilde{p}, K_q\tilde{p}\rangle_{X^*,X} - d^2C_q.(K_q\tilde{p}, K_q\tilde{p}) - dC_q.(\partial_q(K_q\tilde{p}).K_q\tilde{p})\right).$$

Moreover one has $dC_{q_0}.K_{q_0}\tilde{p}_0 = 0$ and $dg_{q(1)} + \tilde{p}(1) = dC_{q(1)}^*\nu$ for some $\nu \in Y^*$.

Note that all functions involved are of class $\mathcal{C}^1$. Therefore, for a given initial condition $(q_0, \tilde{p}_0) \in T_{q_0}^*M$ with $dC_{q_0}.K_{q_0}p_0 = 0$, there exists a unique geodesic emanating from q_0 with initial momentum $\tilde{p}_0$.

Proof. The proof just consists in considering $\tilde{\lambda}(t) = \dot{\lambda}(t)$ and $\tilde{p} = p - dC_q^*.\lambda$, where p and λ are given by Theorem 3, and then in differentiating $C(q(t)) = 0$ twice with respect to time. $\square$

Remark 23. The mappings $\tilde{p}$ and $\tilde{\lambda}$ can also be obtained independently of Theorem 3 as Lagrange multipliers for the mapping of constraints $\Gamma(q, v) = (\dot{q} - \xi_q v, C(q))$, using the same method as in the proof of Theorem 2, under the assumptions that $C : X \to Y$ is of class $\mathcal{C}^2$ and $dC_q.\xi_q$ is surjective for every $q \in M$.

4   Algorithmic procedures

In this section we derive some algorithms in order to compute the solutions of the optimal control problem considered throughout. We first consider problems without constraint in Section 4.1, and then with constraints in Section 4.2.

4.1   Problems without constraints

Shape deformation analysis problems without constraint have already been studied in [9, 10, 11, 13, 17, 20, 21, 26, 29, 30, 34, 36] with slightly different methods, in different and specific contexts. With our previously developed general framework, we are now going to recover methods that are well known in numerical shape analysis, but with a more general point of view allowing us to generalize the existing approaches.

Gradient Descent. We adopt the notations, the framework and the assumptions used in Section 3.1, but without constraint. For $q_0 \in M$ fixed, we consider the optimal control problem of minimizing the functional J defined by (17) over all $(q(\cdot), v(\cdot)) \in H^1_{q_0}(0, 1; M) \times L^2(0, 1; V)$ such that $\dot{q}(t) = \xi_{q(t)}v(t)$ for almost every $t \in [0, 1]$. The Hamiltonian of the problem then does not involve the variable λ, and is the function $H : M \times X^* \times V \to \mathbb{R}$ defined by $H(q, p, v) = \langle p, \xi_q v\rangle_{X^*,X} - L(q, v)$.

We assume throughout that $g \in \mathcal{C}^1(X, \mathbb{R})$, that $\xi : M \times V \to X$ is of class $\mathcal{C}^2$, is a linear mapping in v, and that $L \in \mathcal{C}^2(X \times V, \mathbb{R})$ satisfies the estimate (16).

As in the proof of Theorem 2, we define the mapping $\Gamma : H_{q_0}^1(0, 1; M) \times L^2(0, 1; V) \to L^2(0, 1; X)$ by $\Gamma(q, v)(t) = \dot{q}(t) - \xi_{q(t)} v(t)$ for almost every $t \in [0, 1]$. The objective is to minimize the functional J over the set $E = \Gamma^{-1}(\{0\})$.

According to Lemma 3, the mapping $\partial_q \Gamma_{(q,v)}$ is an isomorphism for all $(q, v) \in E$. Therefore, the implicit function theorem implies that E is the graph of the mapping $v \mapsto q_v$ which to a control $v \in L^2(0, 1; V)$ associates the curve $q_v \in H_{q_0}^1(0, 1; X)$ solution of $\dot{q}_v(t) = \xi_{q_v(t)} v(t)$ for almost every $t \in [0, 1]$ and $q_v(0) = q_0$. Moreover this mapping is, like Γ, of class $\mathcal{C}^2$.

Then, as it was already explained in Remark 3 in the case where $L(q, v) = \frac{1}{2} \|v\|_V^2$, minimizing J over E is then equivalent to minimizing the functional $J_1(v) = J(q_v, v)$ over $L^2(0, 1; V)$.

Thanks to these preliminary remarks, the computation of the gradient of $J_{|E}$ then provides in turn a gradient descent algorithm.

Proposition 8. *The differential of J_1 is given by*

$$dJ_{1\,v}.\delta v = -\int_0^1 \partial_v H(q_v(t), p(t), v(t)).\delta v(t)\, dt,$$

for every $\delta v \in L^2(0, 1; V)$, where $p \in H^1([0, 1], X^*)$ is the solution of $\dot{p}(t) = -\partial_q H(q_v(t), p(t), v(t))$ for almost every $t \in [0, 1]$ and $p(1) + dg_{q_v(1)} = 0$. In particular we have

$$\nabla J_1(v)(t) = -K_V \partial_v H(q_v(t), p(t), v(t)),$$

for almost every $t \in [0, 1]$.

Remark 24. This result still holds true for Lagrangians L that do not satisfy (16), replacing $L^2(0, 1; V)$ with $L^\infty(0, 1; V)$. The gradient is computed with respect to the pre-Hilbert scalar product inherited from L^2.

Proof. Let $v \in L^2(0, 1; V)$ be arbitrary. For every $\delta v \in L^2(0, 1; V)$, we have $dJ_{1\,v}.\delta v = dJ_{(q_v,v)}.(\delta q, \delta v)$, with $\delta q = dq_v.\delta v$. Note that $(\delta q, \delta v) \in T_{(q_v,v)} E$ (the tangent space of the manifold E at (q_v, v)), since E is the graph of the mapping $v \to q_v$. Since $E = \Gamma^{-1}(\{0\})$, we have $\langle d\Gamma^*_{(q_v,v)} p, (\delta q, \delta v) \rangle = 0$ for every $p \in L^2(0, 1; X)^*$. Let us find some particular p such that $\langle dJ_{1(q_v,v)} + d\Gamma^*_{(q_v,v)} p, (\delta q, \delta v) \rangle$ only depends on δv.

Let $p \in H^1([0, 1], X^*)$ be the solution of $\dot{p}(t) = -\partial_q H(q_v(t), p(t), v(t))$ for almost every $t \in [0, 1]$ and $p(1) + dg_{q_v(1)} = 0$. Using the computations done in the proof of Theorem 2, we get

$$\langle dJ_{1(q_v,v)} + d\Gamma^*_{(q_v,v)} p, (\delta q, \delta v) \rangle = \int_0^1 \Big(\langle p(t), \dot{\delta q}(t) \rangle_{X^*, X} - \langle \partial_q H(q_v(t), p(t), v(t)), \delta q(t) \rangle_{X^*, X}$$
$$- \langle \partial_v H(q_v(t), p(t), v(t)), \delta v(t) \rangle_{V^*, V} \Big) dt + dg_{q_v(1)}.\delta q_v(1).$$

Integrating by parts and using the relations $\delta q(0) = 0$ and $p(1) + dg_{q_v(1)} = 0$, we obtain

$$\langle dJ_{1(q_v,v)} + d\Gamma^*_{(q_v,v)} p, (\delta q, \delta v) \rangle = -\int_0^1 \partial_v H(q_v(t), p(t), v(t)).\delta v(t)\, dt$$

Since $\langle d\Gamma^*_{(q_v,v)} p, (\delta q, \delta v) \rangle = 0$, the proposition follows. $\qquad\square$

In the case of shape spaces, which form our main interest here, we have $L(q, v) = \frac{1}{2} \|v\|_V^2$, and then $\partial_v H(q, p, v) = \xi_q^* p - (v, \cdot)_V$ and $K_V \partial_v H(q, p, v) = K_V \xi_q^* p - v$. It follows that

$$\nabla J_1(v) = v - K_V \xi_{q_v}^* p.$$

In particular, if $v = K_V \xi_q^* u$ for some $u \in L^2([0,1], X^*)$, then

$$v - h\nabla J_1(v) = K_v \xi_q^*(u - h(u - p)),$$

for every $h \in \mathbb{R}$. Therefore, applying a gradient descent algorithm does not change this form. It is then important to notice that this provides as well a gradient descent algorithm for solving Problem 2 (the kernel formulation of Problem 1) without constraint, and this in spite of the fact that $L^2([0,1], X^*)$ is not necessarily a Hilbert space. In this case, if $p \in H^1([0,1], X^*)$ is the solution of $\dot{p}(t) = -\partial_q H(q_v(t), p(t), v(t))$ for almost every $t \in [0,1]$ and $p(1) + dg_{q_v(1)} = 0$, then $u - p$ is the gradient of the functional J_2 defined by (12) with respect to the symmetric nonnegative bilinear form $B_q(u_1, u_2) = \int_0^1 \langle u_1(t), K_{q(t)} u_2(t) \rangle_{X^*, X} \, dt$.

This gives a first algorithm to compute unconstrained minimizers in a shape space. We next provide a second method using the space of geodesics.

Gradient descent on geodesics: minimization through shooting. Since the tools are quite technical, in order to simplify the exposition we assume that the shape space M is finite dimensional, i.e., that $X = \mathbb{R}^n$ for some $n \in \mathbb{N}$. The dual bracket $\langle p, w \rangle_{X^*, X}$ is then identified with the canonical Euclidean product $p^T w$, and $K_q = \xi_q K_V \xi_q^* : X^* \to X$ is identified with a $n \times n$ positive semi-definite symmetric matrix. Theorem 3 and Corollary 1 (see Section 3.2) imply that the minimizers of the functional J_1 defined by (10) coincide with those of the functional

$$\hat{J}_1(p_0) = \frac{1}{2} p_0^T K_{q_0} p_0 + g(q(1)), \tag{26}$$

where $v = K_V \xi_q^* p$ and $(q(\cdot), p(\cdot))$ is the geodesic solution of the Hamiltonian system $\dot{q}(t) = \partial_p h(q(t), p(t))$, $\dot{p}(t) = -\partial_q h(q(t), p(t))$, for almost every $t \in [0,1]$, with $(q(0), p(0)) = (q_0, p_0)$. Here, h is the reduced Hamiltonian (see Remark 19) and is given by $h(q, p) = \frac{1}{2} \langle p, \xi_q K_V \xi_q^* p \rangle_{X^*, X} = \frac{1}{2} \langle p, K_q p \rangle_{X^*, X}$. Therefore, computing a gradient of $\hat{J}_1$ for some appropriate bilinear symmetric non-negative product provides in turn another algorithm for minimizing the functional J_1. For example, if the inner product that we consider is the canonical one, then $\nabla \hat{J}_1(p_0) = K_{q_0} p_0 + \nabla(g \circ q(1))(p_0)$. The term $\nabla(g \circ q(1))(p_0)$ is computed thanks to the following well-known result.

Lemma 5. *Let $n \in \mathbb{N}$, let U be an open subset of $\mathbb{R}^n$, let $f : U \to \mathbb{R}^n$ be a complete smooth vector field on U, let G be the function of class $\mathcal{C}^1$ defined on U by $G(q_0) = g(q(1))$, where g is a function on U of class $\mathcal{C}^1$ and $q : [0,1] \to \mathbb{R}^n$ is the solution of $\dot{q}(t) = f(q(t))$ for almost every $t \in [0,1]$ and $q(0) = q_0$. Then $\nabla G(q_0) = Z(1)$ where $Z : [0,1] \to \mathbb{R}^n$ is the solution of $\dot{Z}(t) = df_{q(1-t)}^T Z(t)$ for almost every $t \in [0,1]$ and $Z(0) = \nabla g(q(1))$.*

In our case, we have $U = M \times \mathbb{R}^n$ and $f(q, p) = (\nabla_p h, -\nabla_q h) = (K_q p, -\frac{1}{2} \nabla_q (p^T K_q p))$. Note that we used the Euclidean gradient instead of the derivatives. This is still true thanks to the identification made between linear forms and vectors at the beginning of the section. We get $\nabla \hat{J}_1(p_0) = K_{q_0} p_0 + \alpha(1)$, where $Z(\cdot) = (z(\cdot), \alpha(\cdot))$ is the solution of $\dot{Z}(t) = df_{q(1-t), p(1-t)}^T Z(t)$ for almost every $t \in [0,1]$ and $Z(0) = (z(0), \alpha(0)) = (\nabla g(q(1)), 0)$.

In numerical implementations, terms of the form $df^T w$, with f a vector field and w a vector, require a long computational time since every partial derivative of f has to be computed. In our context however, using the fact that the vector field $f(q, p)$ is Hamiltonian, the computations can be simplified in a substantial way. Indeed, using the commutation of partial derivatives, we get

$$df_{(q,p)}^T Z = \begin{pmatrix} \nabla_q(f(q,p), Z) \\ \nabla_p(f(q,p), Z) \end{pmatrix} = \begin{pmatrix} \nabla_q(\nabla_p h^T z - \nabla_q h^T \alpha) \\ \nabla_p(\nabla_p h^T z - \nabla_q h^T \alpha) \end{pmatrix} = \begin{pmatrix} \partial_p(\nabla_q h).z - \partial_q(\nabla_q h).\alpha \\ \partial_p(\nabla_p h).z - \partial_q(\nabla_p h).\alpha \end{pmatrix}.$$

Replacing h with its expression, we get

$$df^T_{(q,p)}Z = \begin{pmatrix} \nabla_q(p^T K_q z) - \frac{1}{2}\partial_q(\nabla_q(p^T K_q p)).\alpha \\ K_q z - \partial_q(K_q p).\alpha \end{pmatrix}.$$

Therefore, instead of computing $d(\nabla_q(p^T K_q p))^T \alpha$, which requires the computation of all partial derivatives of $\nabla_q(p^T K_q)$, it is required to compute only one of them, namely the one with respect to α. Let us sum up the result in the following proposition.

Proposition 9. *We have* $\nabla \hat{J}_1(p_0) = K_{q_0} p_0 + \alpha(1)$, *where* $(z(\cdot), \alpha(\cdot))$ *is the solution of*

$$\dot{z}(t) = \nabla_q(p(1-t)^T K_{q(1-t)} z(t)) - \frac{1}{2}\partial_q(\nabla_q(p(1-t)^T K_{q(1-t)} p(1-t))).\alpha(t),$$

$$\dot{\alpha}(t) = K_{q(1-t)} z(t) - \partial_q(K_{q(1-t)} p(1-t)).\alpha(t)$$

with $(z(0), \alpha(0)) = (\nabla g(q(1)), 0)$, *and* $(q(t), p(t))$ *satisfies the geodesic equations* $\dot{q}(t) = K_{q(t)} p(t)$ *and* $\dot{p}(t) = -\frac{1}{2}\nabla_q(p(t)^T K_{q(t)} p(t))$ *for almost every* $t \in [0, 1]$, *with* $q(0) = q_0$ *and* $p(0) = p_0$.

A gradient descent algorithm can then be used in order to minimize $\hat{J}_1$ and thus J_1.

4.2 Problems with constraints

In this section, we derive several different methods devoted to solve numerically constrained optimal control problems on shape spaces. To avoid using overly technical notation in their whole generality, we restrict ourselves to the finite-dimensional case. The methods can however be easily adapted to infinite-dimensional shape spaces. We use the notation, the framework and the assumptions of Section 2.2.

Let $X = \mathbb{R}^n$ and let M be an open subset of X. For every $q \in M$, we identify K_q with a $n \times n$ symmetric positive semi-definite real matrix. Throughout the section, we focus on kinetic constraints and we assume that we are in the conditions of Proposition 4, so that these constraints take the form $C_q K_q u = 0$. Note that, according to Proposition 3, in this case Problems 1 and 2 are equivalent. Hence in this section we focus on Problem 2, and thanks to the identifications above the functional J_2 defined by (12) can be written as

$$J_2(u) = \frac{1}{2}\int_0^1 u(t)^T K_{q(t)} u(t)\, dt + g(q(1)).$$

Note (and recall) that pure state constraints, of the form $C(q) = 0$, are treated as well since, as already mentioned, they are equivalent to the kinetic constraints $dC_q.K_q u = 0$.

The augmented Lagrangian method. This method consists of minimizing iteratively unconstrained functionals in which the constraints have been penalized. Although pure state constraints are equivalent to kinetic constraints, in this approach they can also be treated directly. The method goes as follows. In the optimal control problem under consideration, we denote by $\lambda : [0, 1] \to \mathbb{R}^k$ the Lagrange multiplier associated with the kinetic constraints $C_q K_q u = 0$ (its existence is ensured by Theorem 2). We define the augmented cost function

$$J_A(u, \lambda_1, \lambda_2, \mu) = \int_0^1 L_A(q(t), u(t), \lambda(t), \mu)\, dt + g(q(1)),$$

where L_A, called augmented Lagrangian, is defined by

$$L_A(q, u, \lambda, \mu) = L(q, u) - \lambda^T C_q K_q u + \frac{1}{2\mu}|C_q K_q u|^2,$$

with, here, $L(q, u) = \frac{1}{2} u^T K_q u$. Let $q_0 \in M$ fixed. Choose an initial control u_0 (for example, $u_0 = 0$), an initial function λ_0 (for example, $\lambda_0 = 0$), and an initial constant $\mu_0 > 0$. At step ℓ, assume that we have obtained a control u_ℓ generating the curve q_ℓ, a function $\lambda_\ell : [0, 1] \to \mathbb{R}^k$, and a constant $\mu_\ell > 0$. The iteration $\ell \to \ell + 1$ is defined as follows. First, minimizing the unconstrained functional $u \mapsto J_A(u, \lambda_\ell, \mu_\ell)$ over $L^2(0, 1; \mathbb{R}^n)$ yields a new control $u_{\ell+1}$, generating the curve $q_{\ell+1}$ (see further in this section for an appropriate minimization method). Second, λ is updated according to

$$\lambda_{\ell+1} = \lambda_\ell - \frac{1}{\mu_\ell} C_{q_{\ell+1}} K_{q_{\ell+1}} u_{\ell+1}.$$

Finally, we choose $\mu_{\ell+1} \in (0, \mu_\ell]$ (many variants are possible in order to update this penalization parameter, as is well-known in numerical optimization).

Under some appropriate assumptions, as long as μ_ℓ is smaller than some constant $\beta > 0$, u_ℓ converges to a control u^* which is a constrained extremum of J_2. Note that it is not required to assume that μ_ℓ converge to 0. More precisely we infer from [25, Chapter 3] the following convergence result.

Proposition 10 (Convergence of the augmented Lagrangian method). *Assume that all involved mappings are least of class $\mathcal{C}^2$ and that $C_q K_q$ is surjective for every $q \in M$. Let u^* be an optimal solution of Problem 2 and let q^* be its associated curve. Let λ^* be the Lagrange multiplier (given by Theorem 2) associated with the constraints. We assume that there exist $c > 0$ and $\mu > 0$ such that*

$$(\partial_u^2 J_A)_{(u^*, \lambda^*, \mu)}.(\delta u, \delta u) \geqslant c \|\delta u\|_{L^2(0,1;\mathbb{R}^n)}^2, \tag{27}$$

for every $\delta u \in L^2(0, 1; \mathbb{R}^n)$. Then there exists a neighborhood of u^* in $L^2(0, 1; \mathbb{R}^n)$ such that, for every initial control u_0 in this neighborhood, the sequence $(u_\ell)_{\ell \in \mathbb{N}}$ built according to the above algorithm converges to u^*, and the sequence $(\lambda_\ell)_{\ell \in \mathbb{N}}$ converges to λ^*, as ℓ tends to $+\infty$.

Remark 25. Assumption (27) may be hard to check for shape spaces. As is well-known in optimal control theory, this coercivity assumption of the bilinear form $(\partial_u^2 J_A)_{(u^*, \lambda^*, \mu)}$ is actually equivalent to the nonexistence of *conjugate points* of the optimal curve q^* on $[0, 1]$ (see [14, 15] for this theory and algorithms of computation). In practice, computing conjugate points is a priori easy since it just consists of testing the vanishing of some determinants; however in our context the dimension n is expected to be large and then the computation may become difficult numerically.

Remark 26. Pure state constraints $C(q) = 0$ can either be treated in the above context by replacing $C_q K_q$ with $dC_q.K_q$, or can as well be treated directly by replacing $C_q K_q$ with $C(q)$ in the algorithm above.

Any of the methods described in Section 4.1 can be used in order to minimize the functional J_A with respect to u. For completeness let us compute the gradient in u of J_A at the point (u, λ, μ).

Lemma 6. *There holds*

$$\nabla_u J_A(u, \lambda, \mu) = u + C_q^T \lambda + \frac{1}{\mu} C_q^T C_q K_q u - p,$$

where $p(\cdot)$ is the solution of

$$\dot{p}(t) = -\partial_q (p(t)^T K_{q(t)} u(t)) + \partial_q L_A(q(t), u(t), \lambda(t), \mu)$$
$$= \partial_q \left(\left(\frac{u(t)}{2} - p(t) \right)^T K_{q(t)} u(t) \right) + \lambda(t)^T \partial_q C_{q(t)} u(t) + \frac{1}{2\mu} \partial_q \left(u(t)^T K_{q(t)} C_{q(t)}^T C_{q(t)} K_{q(t)} u(t) \right)$$

for almost every $t \in [0, 1]$ and $p(1) + dg_{q(1)} = 0$.

31

Remark 27. For pure state constraints $C(q) = 0$, there simply holds $\nabla_u J_A = u - p$ and the differential equation in p is

$$\dot{p}(t) = \partial_q \left(\left(\frac{u(t)}{2} - p(t) \right)^T K_{q(t)} u(t) \right) + \lambda(t)^T dC_{q(t)} + \frac{1}{\mu} C(q(t))^T dC_{q(t)}.$$

Proof of Lemma 6. We use Proposition 8 with $L(q, u, t) = L_A(q, u, \lambda(t), \mu)$, with λ and μ fixed (it is indeed easy to check that this proposition still holds true when the Lagrangian also depends smoothly on t). The differential of J_A with respect to u is then given by

$$d(J_A)_u.\delta u = \int_0^1 \left(\partial_u L_A(q(t), u(t), \lambda(t), \mu) - p(t)^T K_{q(t)} \right) \delta u(t)\, dt, \tag{28}$$

where $p(\cdot)$ is the solution of $\dot{p}(t) = -\partial_q(p(t)^T K_{q(t)} u(t)) + \partial_q L_A(q(t), u(t), \lambda(t), \mu)$ for almost every $t \in [0, 1]$ and $p(1) + dg_{q(1)} = 0$. To get the result, it then suffices to identify the differential $d(J_A)_u$ with the gradient $\nabla J_A(u)$ with respect to the inner product on $L^2(0, 1; R^n)$ given by $(u_1, u_2)_{L^2(0,1;R^n)} = \int_0^1 u_1(t)^T K_{q(t)} u_2(t)\, dt.$ □

The advantage of the augmented Lagrangian method is that, at every step, each gradient is "easy" to compute (at least as easy as in the unconstrained case). The problem is that, as in any penalization method, a lot of iterations are in general required in order to get a good approximation of the optimal solution, satisfying approximately the constraints with enough accuracy.

The next method we propose tackles the constraints without penalization.

Constrained minimization through shooting. We adapt the usual shooting method used in optimal control (see, e.g., [35]) to our context. For $(q, p) \in M \times X^* = M \times \mathbb{R}^n$, we define $\lambda_{q,p}$ as in Theorem 3 by

$$\lambda_{q,p} = (C_q K_q C_q)^{-1} C_q K_q p.$$

We also denote $\pi_q p = p - C_q^T \lambda_{q,p}$. In particular, $v_{q,p} = K_q \pi_q p$.

Remark 28. A quick computation shows that $\pi_q p$ is the orthogonal projection of p onto $\text{Null}(C_q)$ for the inner product induced by K_q.

According to Theorem 3, Corollary 1 and Remark 20 (see Section 3.2), the minimizers of J_2 have to be sought among the geodesics $(q(\cdot), p(\cdot))$ solutions of (25), and moreover $p(0)$ is a minimizer of the functional

$$\hat{J}_2(p_0) = \frac{1}{2}\|v_{q_0,p_0}\|_V^2 + g(q(1)) = \frac{1}{2}p_0^T \pi_{q_0} K_{q_0} \pi_{q_0} p_0 + g(q(1)).$$

The geodesic equations (25) now take the form

$$\dot{q}(t) = \partial_p h(q(t), p(t)) = K_{q(t)} \pi_{q(t)} p(t),$$

$$\dot{p}(t) = -\partial_q h(q(t), p(t)) = -\frac{1}{2}p(t)^T \pi_{q(t)}^T (\nabla_q K_{q(t)}) \pi_{q(t)} p(t) + \lambda_{q(t),p(t)}^T (\nabla_q C_{q(t)}) K_{q(t)} \pi_{q(t)} p(t),$$

where h is the reduced Hamiltonian (see Remark 19). It follows from Proposition 9 that $\nabla \hat{J}_2(p_0) = \pi_{q_0}^T K_{q_0} \pi_{q_0} p_0 + \alpha(1)$, where $Z(\cdot) = (z(\cdot), \alpha(\cdot))$ is the solution of

$$\dot{z}(t) = \partial_p(\nabla_q h(q(1-t), p(1-t))).z(t) - \partial_q(\nabla_q h(q(1-t), p(1-t))).\alpha(t)$$
$$\dot{\alpha}(t) = \partial_p(\nabla_p h(q(1-t), p(1-t))).z(t) - \partial_q(\nabla_p h(q(1-t), p(1-t))).\alpha(t)$$

with $(z(0), \alpha(0)) = (\nabla g(q(1)), 0)$. Replacing $\nabla_q h$ and $\nabla_p h$ by their expression, we get

$$\dot{z}(t) = p(1-t)^T \pi_{q(1-t)}^T (\nabla_q K_{q(1-t)}) \pi_{q(1-t)} z(t) - \lambda_{q(1-t),p(1-t)}^T (\nabla_q C_{q(1-t)}) K_{q(1-t)} \pi_{q(1-t)} z(t)$$

$$- \partial_q \Big(\frac{1}{2} p(1-t)^T \pi_{q(1-t)}^T (\nabla_q K_{q(1-t)}) \pi_{q(1-t)} p(1-t)$$

$$- \lambda_{q(1-t),p(1-t)}^T (\nabla_q C_{q(1-t)}) K_{q(1-t)} \pi_{q(1-t)} p(1-t) \Big) . \alpha(t)$$

$$\dot{\alpha}(t) = K_{q(1-t)} \pi_{q(1-t)} z(t) - \partial_q (K_{q(1-t)} \pi_{q(1-t)} p(1-t)) . \alpha(t)$$

In practice, the derivatives appearing in these equations can be efficiently approximated using finite differences.

This algorithm of constrained minimization through shooting has several advantages compared with the previous augmented Lagrangian method. The first is that, thanks to the geodesic reduction, the functional $\hat{J}_2$ is defined on a finite-dimensional space (at least whenever the shape space itself is finite dimensional) and hence $\nabla \hat{J}_2(p_0)$ is computed on a finite-dimensional space, whereas in the augmented Lagrangian method $\nabla_u J_A$ was computed on the infinite-dimensional space $L^2(0,1;\mathbb{R}^n)$.

A second advantage is that, since we are dealing with constrained geodesics, all resulting curves satisfy the constraints with a good numerical accuracy, whereas in the augmented Lagrangian method a large number of iterations was necessary for the constraints to be satisfied with an acceptable numerical accuracy.

This substantial gain is however counterbalanced by the computation of $\lambda_{q(t),p(t)}$, which requires solving of a linear equation at every time $t \in [0,1]$ along the curve (indeed, recall that $\lambda_{q,p} = (C_q K_q C_q^*)^{-1} C_q K_q p$). The difficulty here is not just that this step is time-consuming, but rather the fact that the linear system may be ill-conditioned, which indicates that this step may require a more careful treatment. One possible way to overcome this difficulty is to solve this system with methods inspired from quasi-Newton algorithms. This requires however a particular study that is beyond the scope of the present article (see [3] for results and algorithms).

5   Numerical examples

5.1   Matching with constant total volume

In this first example we consider a very simple constraint, namely, a constant total volume. Consider $S = S^{d-1}$, the unit sphere in $\mathbb{R}^d$, and let $M = \mathrm{Emb}^1(S, \mathbb{R}^d)$ be the shape space, acted upon with order 1 by $\mathrm{Diff}(\mathbb{R}^d)$. Consider as in [40] the RKHS V of smooth vector fields given by the Gaussian kernel K with positive scale σ defined by $K(x,y) = e^{-\frac{|x-y|^2}{\sigma^2}} I_d$. An embedding $q \in M$ of the sphere is the boundary of an open subset $U(q)$ with total volume given by $\mathrm{Vol}(U(q)) = \int_S q^* \omega$, where ω is a $(d-1)$-form such that $d\omega = dx_1 \wedge \cdots \wedge dx_d$. Let q_0 be an initial point and let q_1 be a target such that $\mathrm{Vol}(U(q_0)) = \mathrm{Vol}(U(q_1))$. We impose as a constraint to the deformation $q(\cdot)$ to be of constant total volume, that is, $\mathrm{Vol}(U(q(t))) = \mathrm{Vol}(U(q_0))$. The data attachment function is defined by $g(q) = d(q, q_1)^2$, with d a distance between submanifolds (see [41] for examples of such distances).

For the numerical implementation, we take $d = 2$ (thus $S = S^1$) and M is a space of curves, which is discretized as landmarks $q = (x_1, \ldots, x_n) \in \mathrm{Lmk}_2(n)$. The volume of a curve is approximately equal to the volume of the polygon $P(q)$ with vertices x_i, given by $\mathrm{Vol}(P(q)) = \frac{1}{2}(x_1 y_2 - y_2 x_1 + \cdots + x_n y_1 - y_n x_1)$.

If one does not take into account a constant volume constraint, a minimizing flow matching a circle on a translated circle usually tends to shrink it along the way (see Figure 2(a)). If the

volume is required to remain constant then the circle looks more like it were translated towards the target, though the diffeomorphism itself does not look like a translation (see Figure 2(b)).

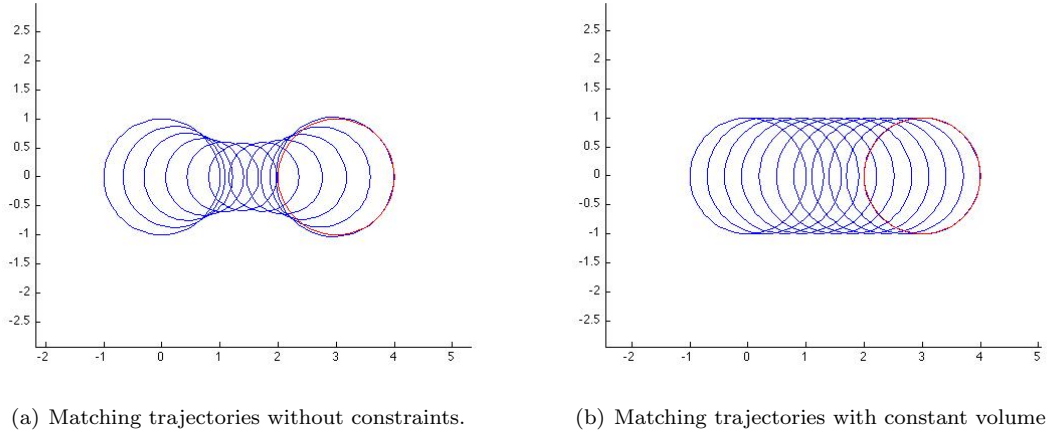(a) Matching trajectories without constraints.　　(b) Matching trajectories with constant volume.

Figure 2: Matching trajectories.

The implementation of the shooting method developed in Section 4.2 leads to the diffeomorphism represented on Figure 3.
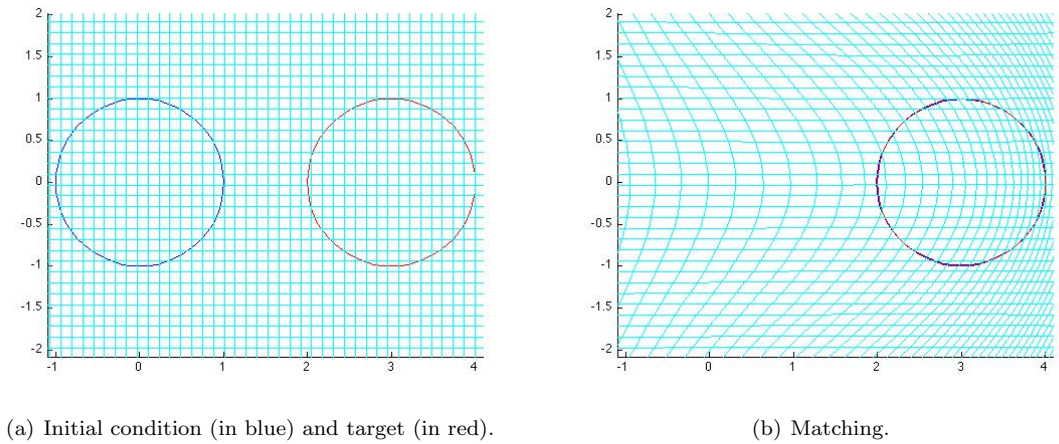

(a) Initial condition (in blue) and target (in red).　　(b) Matching.

Figure 3: Constant volume experiment.

5.2　Multishape matching

We consider the multishape problem described in Section 2.3. We define the shape spaces $M_1, \ldots, M_k$, by $M_j = \mathrm{Emb}(S_j, \mathbb{R}^d)$ for every $j \in \{1, \ldots, k-1\}$ and $M_k = M_1 \times \cdots \times M_{k-1}$ (background space).

For every $j \in \{1, \ldots, k\}$ we consider a reproducing kernel K_i and a reduced operator $K_{q,j}$ for every $q \in M_j$.

In the following numerical simulations, each q_j is a curve in $\mathbb{R}^2$, so that $S_1 = \cdots = S_k = S^1$, the unit circle. The function g appearing in the functional (13) is defined by

$$g(q_1, \ldots, q_k) = \sum_{j=1}^{k-1} \left(d(q_j, q^{(j)})^2 + d(q_k^j, q^{(j)})^2 \right),$$

where $q_k = (q_k^1, \ldots, q_k^{k-1})$ and $q^{(1)}, \ldots, q^{(k-1)}$ are given target curves. The distance d is a distance between curves (see [41] for examples of such distances). We consider two types of compatibility constraints between homologous curves q_j and q_k^j: either the identity (or stitched) constraint $q_j = q_k^j$, or the identity up to reparametrization (or sliding) constraint $q_k^j = q_j \circ f$ for some (time-dependent) diffeomorphism f of S^1. Note that, since the curves have the same initial condition, the latter constraint is equivalent to imposing that $\dot{q}_k^j - \dot{q}_j \circ f$ is tangent to q_k^j, which can also be written as

$$(v_k(t, q_k^j) - v_j(t, q_k^j)) \cdot \nu_k^j = 0,$$

where $v_j = K_j \xi_{q_j}^* u_j$ and ν_k^j is normal to q_k^j. In the numerical implementation, the curves are discretized into polygonal lines, and the discretization of the control system (14) and of the minimization functional (13) is done by reduction to landmark space, as described in section 2.4. The discretization of the constraint in the identity case is straightforward. For the identity up to reparametrization (or sliding) constraint, the discretization is slightly more complicated and can be done in two ways. A first way is to add a new state variable ν_k^j which evolves while remaining normal to q_k^j, according to

$$\dot{\nu}_k^j = -dv_k^j (q_k^j)^T \nu_k^j,$$

which can be written as a function of the control u_k and of the derivatives of K_k (this is an example of lifted state space, as discussed in Section 2.3). A second way, which is computationally simpler and that we use in our experiments, avoids introducing a new state variable and uses finite-difference approximations. For every $j = 1, \ldots, k-1$, and every line segment $\ell = [z_\ell^-, z_\ell^+]$ in q_k^j (represented as a polygonal line), we simply use the constraint $\nu_\ell \cdot (v_j(\ell) - v_k(\ell)) = 0$, where ν_ℓ is the unit vector perpendicular of $z_\ell^+ - z_\ell^-$ and $v_j(\ell) = \frac{1}{2}(v_j(z_\ell^-) + v_j(z_\ell^+))$. Note that the vertices z_ℓ^- and z_ℓ^+ are already part of the state variables that are obtained after discretizing the background boundaries q_k.

With these choices, the discretized functional and its associated gradient for the augmented Lagrangian method are obtained with a rather straightforward – albeit lengthy – computation. In Figure 4, we provide an example comparing the two constraints. In this example, we take $k = 2$ and use the same radial kernel $K_1 = K_2$ for the two shapes, letting $K_1(x, y) = \gamma(|x - y|/\sigma_1)$, with

$$\gamma(t) = (1 + t + 2t^2/5 + t^3/15)e^{-t}.$$

The background kernel is $K_3(x, y) = \gamma(|x - y|/\sigma_3)$, with $\sigma_1 = 1$ and $\sigma_3 = 0.1$. The desired transformation, as depicted in Figure 4, moves a curve with elliptical shape upwards, and a flower-shaped curve downwards, each curve being, in addition, subject to a small deformation. The compared curves have a diameter of order 1.

The solutions obtained using the stitched and sliding constraints are provided in Figures 5(a) and 5(b), in which we have also drawn a deformed grid representing the diffeomorphisms induced by the vector fields v_1, v_2 and v_3 in their relevant regions. The consequence of the difference between the kernel widths inside and outside the curves on the regularity of the deformation is
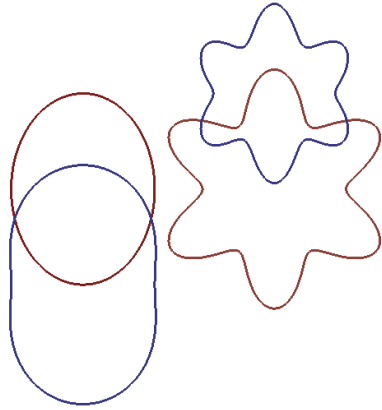
Figure 4: Multishape experiment: initial (blue) and target (red) sets of curves.

obvious in both experiments. One can also note differences in the deformation inside between the stitched and sliding cases, the second case being more regular thanks to the relaxed continuity constraints at the boundaries. Finally, we mention the fact that the numerical method that we illustrate here can be easily generalized to triangulated surfaces instead of polygonal lines.

This example, for which in curve contains a little more than 300 points, was run on a mac pro workstation with two 2.66 mhz quad-core intel processors, with python code combining fortran with openmp calls for the most computationally intensive parts. For stitched constraints, each gradient descent step took around 1.7s, while using on average 7 processors in parallel, with an approximate total computation to of 30 min. Sliding constraints are more demanding computationally. They require around 5s per gradient iteration, with a 3 hour total computation time.
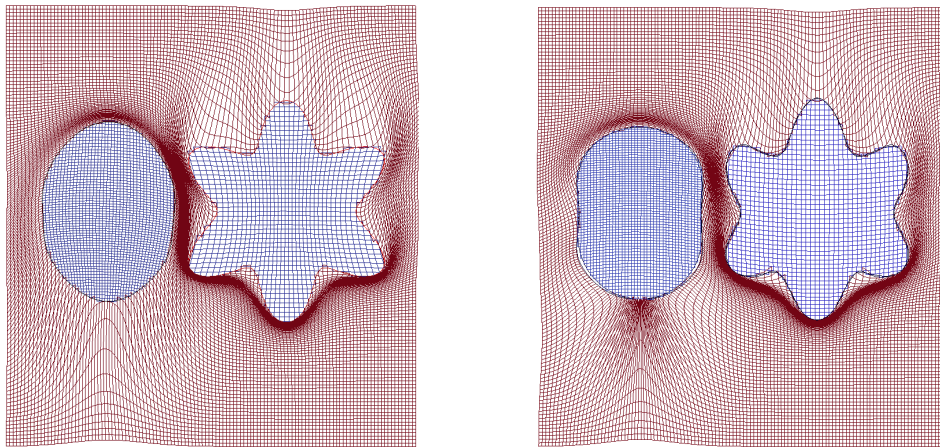
(a) Stitched constraints. (b) Sliding constraints.

Figure 5: Multishape experiment.

Simultaneous study of several shapes in images of the brain. We now return to the bran structures depicted in Figure 1.

<center>(a) Initial shapes.   (b) Target shapes.</center>
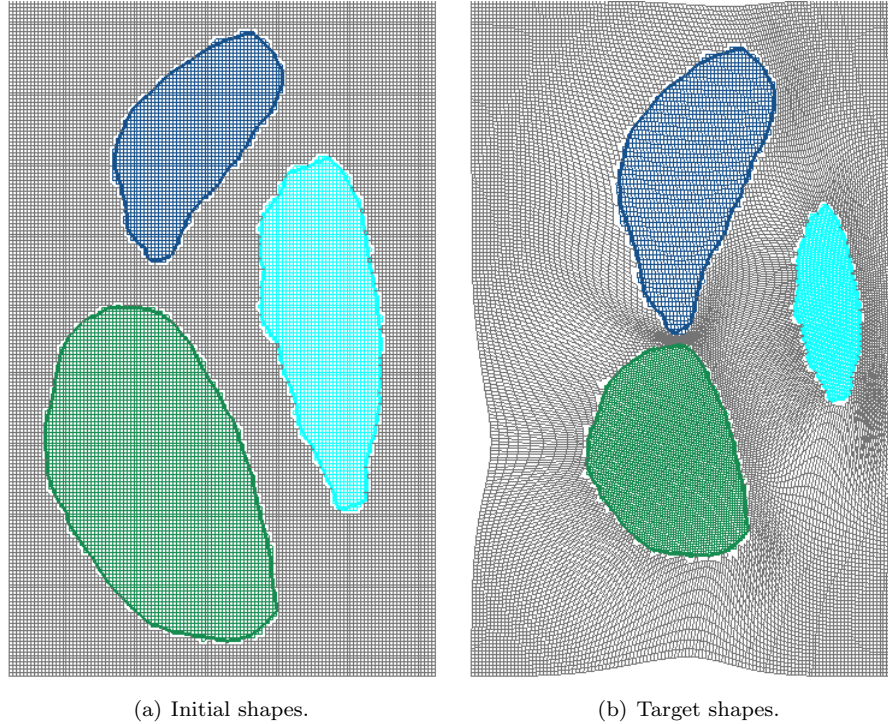
<center>Figure 6: Matching several shapes at once</center>

We apply the same method as above to match these three shapes onto the three corresponding target shapes of Figure 1(b) using stitching constraints. This requires four diffeomorphisms: one for each shape and one for the background. We use the reproducing kernels $K_i(x, y) = \gamma(|x-y|/\sigma_i)$, $i = 1, \ldots, 4$, with $\sigma_1 = \sigma_2 = \sigma_3 = 10$mm for the shapes and $\sigma_4 = 1$mm. An augmented Lagrangian method yields the deformed shapes of Figure 6(b).

We see that our approach solves all problems mentioned in the introduction: while the shapes cannot overlap, two of them end up quite close, and the deformation of each shape is evidently less severe than that of the background, which fits the physical facts much better than what would happen if we would use a single diffeomorphism for the whole multi-shape.

6 Conclusion and open problems

The purpose of this paper was to develop a very general framework for the analysis of shape deformations, along with practical methods to find an optimal deformation in that framework. The point of view of control theory gives powerful tools to attain this goal. This allows in particular for the treatment of constrained deformation, which had not been done before.

While our example focused on curve matching, and may seem somewhat academic compared to typical problems in computational anatomy, they provide a good illustration of the impact of

adding constraints in shape registration problems and their results are easy to interpret. Numerical methods and experimental results on brain structures for the more challenging problem of matching multiple surfaces are described in [4], to which we refer the interested reader.

Now that a concrete setting has been fixed, many new developments can be expected. First of all, the minimization algorithms in the case of constrained shapes are quite slow for a very high number of constraints. Moreover, we did not study any of the geometric aspects of shape deformation spaces. For example, we only briefly mentioned the infinite-dimensional sub-Riemannian structure that the RKHS induces on both groups of diffeomorphism and shape spaces. Sub-Riemannian geometry in infinite dimension and codimension is still a very open subject with very few results. The sub-Riemannian geometry in this paper is particularly difficult and interesting to study because the horizontal spaces may not be closed in the ambient space. A key difference with finite dimension is that some geodesics might exist that are neither normal nor abnormal.

More general control problems can be designed. One can, for example, use a second-order approach, with a control system taking the form

$$\dot{q} = \partial_p H(q, p), \quad \dot{p} = -\partial_q H(q, p) + f(q, u)$$

in which the original state is lifted to the cotangent space $(q \to (q, p))$, and the new control is u. These models have been introduced for shapes in [38], with $f(q, u) = u$, providing a way to interpolate smoothly between multiple shapes. We are currently exploring applications of this approach to model muscle-like motions, with external forces constrained to being collinear to the fibers.

Another extension to consider is that of inequality constraints, both linear and nonlinear. Such a generalization would have many applications. For example, one might wish to impose an upper bound on the Euclidean norm of the deformation vector field v along the shape. Let us take $M = \mathcal{C}^0(S^2, \mathbb{R}^d)$, the space of closed continuous surfaces in $\mathbb{R}^d$, and let us impose the constraints $\sup_{s \in S^2} |v(q(s))| \leqslant 1$. From what is known in the finite-dimensional case (see [23]), it is reasonable to expect that, for the corresponding Pontryagin maximum principle, Hamiltonian equations are similar to those of Theorem 2, with the particularity that, at almost every time t, the Lagrange multiplier $\lambda(t) \in \mathcal{C}^0(S^2, \mathbb{R})^* = \mathcal{M}(S^2)$ actually belongs to the space of *nonnegative* Radon measures on S^2 whose support is contained in $\{s \in S^2 \mid |v(t, q(t)(s))| = 1\}$. In other words, as a Radon measure, $\lambda(t)(s) \geqslant 0$ for every $s \in S^2$ and $\lambda(t)(s) > 0$ whenever $|v(t, q(t)(s))| = 1$. Proving that statement is certainly not that easy. Adapting the surjectivity assumption for the constraints seems problematic. A fortiori, extending the result to a more general setting such as that of Theorem 2 seems challenging. Note however that, in practice, one often works with finite-dimensional models and then these constraints are discretized into a finite number of inequality constraints. Such problems have been extensively studied in finite dimension (see, e.g., [23, 35]).

Another glaring issue comes from the assumption of the surjectivity of the constraints in Theorem 2. Indeed, in most practical cases, such as multishapes in Section 2.3, this assumption fails. In finite dimension, problems can occur when the rank of C_q changes with q. They are usually solved by taking higher-order derivatives of C_q on the sets on which it is not maximal. This does not seem easily possible with Banach spaces. Another problem comes from the incompatibility of topologies between the Hilbert space V and the Banach space Y in which the constraints are valued: C_q may not have a closed range, in which case we could find "missing" Lagrange multipliers. If $C_q(V)$ were constant, this could be solved simply by restricting Y to $C_q(V)$ equipped with the Hilbert topology induced by C_q and V, but since it is not constant, this might be impossible. It would be very interesting, both for control theory in general and for shape deformation analysis in particular, to find a way to address this problem.

References

[1] A. Agrachev, Y. Sachkov, *Control theory from the geometric viewpoint*, Encyclopaedia Math. Sciences **87**, Springer-Verlag, 2004.

[2] A. Agrachev, U. Boscain, G. Charlot, R. Ghezzi, M. Sigalotti. *Two-dimensional almost-Riemannian structures with tangency points.*, Ann. Inst. H. Poincaré Anal. Non Linéaire, **27** (2010), no 3, 793–807.

[3] S. Arguillère, *Approximating sequences of symmetric matrices with a generalized quasi-Newton algorithm*, Preprint Hal (2013).

[4] S. Arguillère, E. Trélat, A. Trouvé, L. Younès, *Multiple shape registration using constrained oprimal control*, Preprint (2014).

[5] V. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier **16** (1966), fasc. 1, 319–361.

[6] V. Arnold, *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics **60**, Springer-Verlag, New York, 1989.

[7] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.

[8] B. Avants, J.C. Gee, *Geodesic estimation for large deformation anatomical shape averaging and interpolation*, Neuroimage **23** (2004), S139–S150.

[9] R. Azencott, R. Glowinski, J. He, A. Jajoo, Y. Li, A. Martynenko, R.H.W. Hoppe, S. Benzekry, S.H. Little, W.A. Zoghbi, *Diffeomorphic matching and dynamic deformable surfaces in 3D medical imaging*, Comput. Methods Appl. Math. **10** (2010), no 3, 235–274.

[10] R. Azencott, R. Glowinski, A.M. Ramos, *A controllability approach to shape identification by elastic matching*, Appl. Math. Lett. **21** (2008), no 8, 861–865.

[11] M.F. Beg, M.I. Miller, A. Trouvé, L. Younes, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, Int. J. Comput. Vis. **61** (2005), no. 2, 139–157.

[12] *Sub-Riemannian geometry*, Progress in Mathematics **144**, A. Bellaïche and J.-J. Risler eds, Birkhäuser Verlag, Basel, 1996.

[13] S. Ben-Zekry, G. Lawrie, S. Little, W. Zogbhi, J. Freeman, A. Jajoo, S. Jain, J. He, A. MArtynenko, R. Azencott, *Comparative evaluation of mitral valve valve strain by deformation tracking in 3D-echocardiography*, Cardiovasc. Eng. Tech. **3** (2012), no. 4, 402-412.

[14] B. Bonnard, J.-B. Caillau, E. Trélat, *Second order optimality conditions in the smooth case and applications in optimal control*, ESAIM Control Optim. Calc. Var. **13** (2007), no. 2, 207–236.

[15] B. Bonnard, L. Faubourg, E. Trélat, *Mécanique céleste et contrôle des véhicules spatiaux (French) [Celestial mechanics and the control of space vehicles]*, Math. & Appl. (Berlin), 51, Springer-Verlag, Berlin, 2006.

[16] H. Brézis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[17] P. Dupuis, U. Grenander, M.I. Miller, *Variational problems on flows of diffeomorphisms for image matching*, Quart. Appl. Math. **56** (1998), no. 3, 587–600.

[18] S. Durrleman, S. Allassonnière, S. Joshi, *Sparse Adaptive Parameterization of Variability in Image Ensembles*, Int. J. of Computer Vision, Springer Verlag (Germany), **101** (2013), no. 1, 161–183.

[19] F. Gay-Balmaz, T. Ratiu, *Clebsch optimal control formulation in mechanics*, J. Geom. Mec. **3** (2011), no. 1, 47–79.

[20] J. Glaunès, A. Trouvé, L. Younes, *Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching*, in CVPR (2004), 712–718, Los Alamitos, IEEE Comput. Soc.

[21] J. Glaunès, A. Trouvé, L. Younes, *Modeling planar shape variation via Hamiltonian flows of curves*, in: Statistics and analysis of shapes, Model. Simul. Sci. Eng. Technol., 335–361, Birkhäuser Boston, 2006.

[22] U. Grenander, M.I. Miller, *Computational anatomy: an emerging discipline*, Current and future challenges in the applications of mathematics (Providence, RI, 1997), Quart. Appl. Math. **56** (1998), no. 4, 617–694.

[23] R.F. Hartl, S.P. Sethi, R.G. Vickson, *A survey of the maximum principles for optimal control problems with state constraints*, SIAM Rev. **37** (1995), no. 2, 181–218.

[24] D.D. Holm, J. Marsden, T.S. Ratiu, *Euler-Poincaré models of ideal fluids with nonlinear dispersion*, Phys. Rev. Lett. **80** (1998), no. 19, 4173–4176.

[25] K. Ito, K. Kunisch, *Lagrange multiplier approach to variational problems and applications*, Advances in Design and Control, Vol. 15, SIAM, Philadelphia, PA, 2008.

[26] S.C. Joshi, M.I. Miller, *Landmark matching via large deformation diffeomorphisms*, IEEE Transcript Image Processing **9** (2000), no. 8, 1357–1370.

[27] S. Kurcyusz, *On the existence and nonexistence of Lagrange multipliers in Banach spaces*, J. Optim. Theory Appl. **20** (1976), no. 1, 81–110.

[28] J.E. Marsden, T.S. Ratiu, *Introduction to mechanics and symmetry*, Texts in Applied Mathematics **17**, second edition, Springer-Verlag, New York, 1999.

[29] M.I. Miller, A. Trouvé, L. Younes, *On the metrics and Euler-Lagrange equations of computational anatomy*, Annu. Rev. Biomed. Eng. **4** (2002), 375–405.

[30] M.I. Miller, A. Trouvé, L. Younes, *Geodesic shooting for computational anatomy*, J. Math. Imaging Vision **24** (2006), no. 2, 209–228.

[31] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs **91**, American Mathematical Society, Providence, RI, 2002.

[32] H. Omori, *Infinite dimensional Lie transformation groups*, Lecture Notes in Mathematics, Vol. 427, Springer-Verlag, Berlin-New York, 1974.

[33] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mishchenko, *The mathematical theory of optimal processes*, A Pergamon Press Book, The Macmillan Co., New York, 1964.

[34] Y. Qin, J. W. He, R. Azencott, *A second order variational approach for diffeomorphic matching of 3D surfaces*, Numerical Analysis & Scientific Computing Preprint Series, no 27, 2014.

[35] E. Trélat, *Contrôle optimal (French) [Optimal control], Théorie & applications [Theory and applications]*, Math. Concrètes [Concrete Mathematics], Vuibert, Paris, 2005.

[36] A. Trouvé, *Action de groupe de dimension infinie et reconnaissance de formes*, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), no. 8, 1031–1034.

[37] A. Trouvé, *Diffeomorphism groups and pattern matching in image analysis*, International Journal of Computational Vision **37** (2005), no. 1, 17 pages.

[38] A. Trouvé, F.-X. Vialard, *Shape splines and stochastic shape evolutions: a second order point of view*, Quart. Appl. Math. **70** (2012), no. 2, pp. 219–251.

[39] A. Trouvé, L. Younes, *Local geometry of deformable templates*, SIAM J. Math. Anal. **37** (2005), no. 1, pp. 17–59.

[40] A. Trouvé, L. Younes, *Shape spaces*, in: Handbook of Mathematical Methods in Imaging, O. Scherzer ed., Springer New York, 2011, 1309–1362.

[41] M. Vaillant, J. Glaunès, *Surface Matching via Currents*, in: Information Processing in Medical Imaging, 1–5, Springer Berlin, Heidelberg, 2005, G. Christensen and M. Sonka (eds), Lecture Notes in Computer Science, Vol. 3565.

[42] L. Younes, *Shapes and diffeomorphisms*, Applied Mathematical Sciences **171**, Springer-Verlag, Berlin, 2010.

[43] L. Younes, *Constrained diffeomorphic shape evolution*, Found. Comput. Math. **12** (2012), no. 3, 295–325.