CrossMark

Metamorphosis of images in reproducing kernel Hilbert spaces

Casey L. Richardson[1] · Laurent Younes[2]

Abstract Metamorphosis is a method for diffeomorphic matching of shapes, with many potential applications for anatomical shape comparison in medical imagery, a problem which is central to the field of computational anatomy. An important tool for the practical application of metamorphosis is a numerical method based on shooting from the initial momentum, as this would enable the use of statistical methods based on this momentum, as well as the estimation of templates from hyper-templates using morphing. In this paper we introduce a shooting method, in the particular case of morphing images that lie in a reproducing kernel Hilbert space (RKHS). We derive the relevant shooting equations from a Lagrangian frame of reference, present the details of the numerical approach, and illustrate the method through morphing of some simple images.

Keywords Groups of diffeomorphisms · Shape analysis · Deformable templates · Metamorphosis · Adjoint methods

✉ Laurent Younes
laurent.younes@jhu.edu

Casey L. Richardson
Casey.Richardson@jhuapl.edu

[1] Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, MD 20723-6099, USA

[2] Center for Imaging Science and Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218-2686, USA

Mathematics Subject Classifications (2010) 58E50

1 Introduction

Metamorphosis is a pattern matching framework that combines diffeomorphic mapping with variations in shape or image space; it has potential for interesting applications in shape analysis and computational anatomy [38, 43, 44]. One of its advantages is to allow for transgression of the diffeomorphic constraint, inducing changes in topology between the template and the target image, resulting in a transformation bringing the template to the target through the minimization of a geodesic cost associated to a Riemannian metric on the product space of shapes and deformations. For images, this is accomplished by allowing both deformations of the template as well as smooth changes in the template's intensity values. Through this combination of changes, the template is morphed into the target (see [25, 37, 45] for a precise description, and Section 2 for more details). As a result, metamorphosis is a "near-diffeomorphic" method that allows one to compare non-diffeomorphic objects within a well-specified Riemannian geometric framework. It therefore takes a special place among other "pure" diffeomorphic methods, like the large-deformation diffeomorphic metric mapping [8, 9, 45], diffeomorphic demons, [39], elastic matching approaches [16, 17], or optimal transportation methods [4, 10, 22, 40] — which properly speaking are not matching images, but densities, since optimal transport does not change the total mass of the matched functions.

In this paper, we generalize previously known results for image metamorphosis, and introduce a new shooting method for computing minimizers of the image metamorphosis matching functional, in the case where the images have some degree of smoothness (they are elements of a certain reproducing kernel Hilbert space). Our work builds upon [25], which introduced a general formulation of metamorphosis using the Euler-Poincaré framework, and then derived the continuous-time evolution equations for metamorphosis (EPMorph) in several concrete situations, such as image matching, density matching, and measure matching. This paper also suggested extensions of its analysis and numerics for further work, e.g. the numerics for morphing of discrete measures which was analyzed by the authors of this paper in [36]. In Section 11.2 of [25], Holm et al. apply metamorphosis to the case of images that are members of a reproducing kernel Hilbert space (RKHS), and then they propose the development of numerical methods for the EPMorph equations in this context. In this paper, we develop this idea into a shooting method for morphing RKHS images, by deriving the appropriate forward and adjoint equations, and then we present some numerical experiments that illustrate the use of such a method for simple examples of shape matching. We also complete the theoretical analysis of these methods, in a framework that covers a large range of applications.

The first part of the paper provides a formal presentation of the approach, leaving the detailed discussion of the hypotheses and rigorous proofs to the second part, constituted by Section 7. The basic notation and assumptions are presented in Section 2 together with the metamorphosis variational problem and associated optimality equations. Section 3 describes a family of singular solutions that satisfy

the optimality equations, providing a key component of the proposed numerical procedure. These singular solutions are then reinterpreted in Section 4 as the solutions that arise from a relaxation of the original problem replacing the infinite-dimensional boundary conditions in image space with a finite number of constraints. The numerical solution of the relaxed problem is then described in Section 5, with complements given in the Appendix. Section 6 then provides experimental results.

2 Mathematical setup

Reproducing kernel Hilbert spaces will be key elements in our construction. If X is a Banach or Hilbert space, we will denote by $(\mu \mid h)$ the pairing between a linear form $\mu \in X^*$ and a vector $h \in X$; the inner product in a Hilbert space X will be denoted by $\langle h, k \rangle_X$, $h, k \in X$. In the Hilbert case, we will denote by $\boldsymbol{K}_X$: the isometry map between X^* and X, such that $(\mu \mid h) = \langle \boldsymbol{K}_X \mu, h \rangle_X$, and by A_X its inverse, $A_X = \boldsymbol{K}_X^{-1}$. If X and Y are Banach and $A : X \to Y$ a bounded operator, we let $A^* : Y^* \to X^*$ be the conjugate, defined by $(A^* \mu \mid h) = (\mu \mid Ah)$. If $X = Y$ are Hilbert, we let A^T be the transpose, defined by $\left\langle A^T h, \tilde{h} \right\rangle_H = \left\langle h, A \tilde{h} \right\rangle$, or $A^T = \boldsymbol{K}_X A^* A_X$. We will also denote by A^T the transpose matrix of a finite-dimensional operator. Finally, if X, Y are two Banach spaces $\mathcal{L}(X, Y)$ denotes the set of bounded linear operators from X to Y, and the operator norm is denoted $\| \cdot \|_{\mathcal{L}(X,Y)}$. If $Y = X$, we will use $\mathcal{L}(X)$ instead of $\mathcal{L}(X, Y)$.

A Hilbert space X continuously embedded in $L^2(\mathbb{R}^d, \mathbb{R}^k)$ is a reproducing kernel Hilbert space (RKHS) if, for all $x \in \mathbb{R}^d$, the Dirac measure $\delta_x : X \to \mathbb{R}^k$, defined by $\delta_x(h) = h(x)$ is a bounded linear map. If X is an RKHS, and given $a \in \mathbb{R}^k$, we will denote by $a \cdot \delta_x$ the continuous linear form $(a \cdot \delta_x \mid h) = a \cdot h(x)$, where the latter denotes the usual dot product in $\mathbb{R}^k$. The kernel of X is then the matrix-valued function $(x, y) \mapsto K_X : (x, y)$ defined by

$$K_X(x, y)a = \boldsymbol{K}_X(a \cdot \delta_y)(x).$$

($K_X(x, y)$ is a k by k matrix, and k will be either d or 1 in the following discussion.)

Metamorphosis is a diffeomorphic registration framework: it is formulated using a certain subgroup of diffeomorphisms of $\mathbb{R}^d$ acting, as a left group action, on images (see [13, 25, 33, 36, 37] for more general classes of metamorphoses). This group, denoted Diff_V, is the set of all diffeomorphisms of $\mathbb{R}^d$ that can be attained as flows of time-dependent vector fields $v \in L^2([0, 1]; V)$, where V is a reproducing kernel Hilbert space continuously embedded in $\mathcal{B}^p := C_0^p(\mathbb{R}^d; \mathbb{R}^d)$ for some $p \geq 1$ (the space of C^p vector fields that decay to zero at infinity). More precisely, $\psi \in G$ if and only if $\psi = \varphi(1)$, where φ is the solution of

$$\dot{\varphi}(t) = v(t) \circ \varphi(t),$$
$$\varphi(0) = \text{id}$$

for some v satisfying $\int_0^1 \|v(t)\|_V^2 dt < \infty$. The group Diff_V is then embedded in the space Diff^p of diffeomorphisms ψ such that $\psi - \text{id}$ and $\psi^{-1} - \text{id}$ both belong to $\mathcal{B}^p$, which forms an open subset of the affine space $\text{id} + \mathcal{B}^p$.

In most this paper, the image space is a scalar RKHS, denoted H (we will weaken this assumption in some of the results of Section 7). To simplify the discussion, we will assume that H is equivalent to a Sobolev space $\mathcal{H}^r(\mathbb{R}^d)$ (the space of functions with square integrable partial derivatives up to order r) for some $r > d/2 + 1$, so that elements of H are differentiable. Assuming that $p \geq r$, we will consider the the action of C^p diffeomorphisms H given by $\varphi \cdot q = q \circ \varphi^{-1}$.

In order to connect two images $q^{(0)}$ and $q^{(1)}$ in H with a continuous path $q(t)$, image metamorphosis solves the optimal control problem

$$\frac{1}{2}\int_0^1 \|v(t)\|_V^2 dt + \frac{1}{2\sigma^2}\int_0^1 \|\zeta(t)\|_H^2 dt \longrightarrow \min$$
$$\text{subject to } \dot{q}(t) = \nabla q(t) \cdot v(t) + \zeta(t), \ q(0) = q^{(0)} \text{ and } q(1) = q^{(1)}. \ (1)$$

We will prove in Section 7 that, under some additional conditions, solutions of this problem exist and satisfy a Pontryagin maximum principle (PMP) that we derive formally here. Introduce the control-dependent Hamiltonian

$$H(p, q, v, \zeta) = (p \mid \nabla q \cdot v + \zeta) - \frac{1}{2}\|v\|_V^2 - \frac{1}{2\sigma^2}\|\zeta\|_H^2.$$

The PMP [3, 41] states that optimal solutions of Eq. 1 satisfy

$$\begin{cases} \dot{q}(t) = \partial_p H \\ \dot{p}(t) = -\partial_q H \\ (v, \zeta) = \text{argmax } H(p, q, \cdot, \cdot) \end{cases}$$

yielding

$$\begin{cases} \dot{q}(t) = \nabla q(t) \cdot v(t) + \zeta(t) \\ \dot{p}(t) + \nabla \cdot (p(t)v(t)) = 0 \\ \zeta(t) = \sigma^2 K_H p(t) \\ v(t) = -K_V(\nabla q(t) \cdot p(t)) \end{cases} \qquad (2)$$

We will use the following reformulation of problem (1). The evolution equation for q is an advection and is equivalent to

$$\dot{m}(t, \cdot) = \zeta(t, \varphi(t, \cdot))$$

with $m(t, \cdot) = q(t, \varphi(t, \cdot)) \in H$. Considering (φ, m) as a new state, we can define the problem

$$\frac{1}{2}\int_0^1 \|v(t)\|_V^2 dt + \frac{1}{2\sigma^2}\int_0^1 \|\zeta(t)\|_H^2 dt \longrightarrow \min$$
$$\text{subject to } \dot{\varphi}(t) = v(t) \circ \varphi(t), \ \dot{m}(t) = \zeta(t) \circ \varphi(t), \ m(0) = q^{(0)} \text{ and } m(1) = q^{(1)} \circ \varphi(1).$$
$$(3)$$

One of the interests of introducing (3) is that the formulation does not require m to be differentiable (in space) anymore (one can however use a generalized form of the evolution equation in Problem (1) to make this problem equivalent to Problem (3) — see [37]). Moreover, applying (still formally) the PMP to Problem (3) yields another set of optimality conditions that will be convenient later. Introduce a co-state $\rho = (\rho_\varphi, \rho_m) \in (\mathcal{B}^p)^* \times H^*$ and the Hamiltonian

$$H(\rho_\varphi, \rho_m, \varphi, m, v, \zeta) = \left(\rho_\varphi \mid v \circ \varphi\right) + \left(\rho_m \mid \zeta \circ \varphi\right) - \frac{1}{2}\|v\|_V^2 - \frac{1}{2\sigma^2}\|\zeta\|_H^2.$$

For $\varphi \in \text{Diff}_V$, introduce the operators $T_\varphi : v \to v \circ \varphi$ and $\tilde{T}_\varphi : \zeta \to \zeta \circ \varphi$, respectively from V to $\mathcal{B}^p$ and from H to itself. The PMP then gives the equations

$$\begin{cases} \dot{\varphi} = v \circ \varphi \\ \dot{m} = \zeta \circ \varphi \\ \dot{\rho}_\varphi = -\partial_\varphi\left(\rho_\varphi \mid v \circ \varphi\right) - \partial_\varphi(\rho_m \mid \zeta \circ \varphi) \\ \dot{\rho}_m = 0 \\ v = K_V T_\varphi^* \rho_\varphi \\ \zeta = \sigma^2 K_H \tilde{T}_\varphi^* \rho_m \end{cases} \quad (4)$$

These conditions imply, in particular, that ρ_m is constant. The boundary condition $m(1) \circ \varphi(1) = q^{(1)}$ implies a boundary condition for ρ, namely that $\left(\rho_\phi(1) \mid w\right) + (\rho_m \mid z) = 0$ whenever

$$z = \nabla m(1) \cdot D\varphi(1)^{-1} w,$$

since $\nabla q^{(1)} \circ \varphi(1) = D\varphi(1)^{-T} \nabla q^{(1)}$. This yields

$$\left(\rho_\varphi(1) \mid w\right) + \left(\rho_m \mid \nabla m(1) \cdot D\varphi(1)^{-1} w\right) = 0.$$

for all $w \in \mathcal{B}^p$, or, replacing w by $D\varphi(1)w$,

$$\left(\rho_\varphi(1) \mid D\varphi(1)w\right) + (\rho_m \mid \nabla m(1) \cdot w) = 0 \quad (5)$$

holding for all $w \in \mathcal{B}^p$.

Note that system (4) implies that

$$\partial_t \left(\left(\rho_\varphi(t) \mid D\varphi(t)w\right) + (\rho_m \mid \nabla m(t) \cdot w)\right) =$$
$$-\left(\rho_\varphi(t) \mid Dv(t) \circ \varphi(t) D\varphi(t)w\right) - (\rho_m \mid \nabla\zeta(t) \circ \varphi(t) \cdot D\varphi(t)w)$$
$$+\left(\rho_\varphi(t) \mid Dv(t) \circ \varphi(t) D\varphi(t)w\right) + (\rho_m \mid \nabla\zeta(t) \circ \varphi(t) \cdot D\varphi(t)w) = 0,$$

for which we have used $\partial_t D\varphi(t) = Dv(t) \circ \varphi(t)D\varphi(t)$ and $\partial_t \nabla m(t) = D\varphi(t)^T \nabla\zeta(t) \circ \varphi(t)$. This implies that the linear form

$$\mu(t) : w \mapsto \left(\rho_\varphi(t) \mid D\varphi(t)w\right) + (\rho_m(t) \mid \nabla m(t) \cdot w)$$

is invariant along (4), and the boundary condition (5) propagates over all times, i.e., $\mu(t) = 0$ over $[0, 1]$.

Finally, we let the reader check that one can pass from solutions of Eq. 2 to solutions of Eq. 4 with the change of variables $q(t) \circ \phi(t) = m(t)$ and

$$(p(t) \mid z) = (\rho_m \mid z \circ \phi(t)).$$

Note also that the boundary condition can be rewritten in terms of $q = m \circ \varphi^{-1}$ as

$$\left(\rho_\varphi(t) \mid w\right) = (\rho_m \mid \nabla q(t) \cdot w). \tag{6}$$

3 Singular solutions

It was recognized in [25] that system (2) admits a family of singular solutions. These solutions are obtained directly from Eq. 4 by taking ρ_φ and ρ_m in the form

$$\rho_\varphi(t) = \sum_{k=1}^{N} z_k(t) \cdot \delta_{x_k^{(0)}} \tag{7}$$

$$\rho_m = \sum_{k=1}^{N} \alpha_k \delta_{x_k^{(0)}} \tag{8}$$

Here, $x^{(0)} = \left\{x_k^{(0)}\right\}_{k=1}^{N}$ is a collection of points, or particles, in $\mathbb{R}^d$, $z(t) = \{z_k(t)\}_{k=1}^{N}$ is a collection of time-dependent vectors in $\mathbb{R}^d$, $\alpha = \{\alpha_k\}_{k=1}^{N}$ is a time-independent collection of scalars.

Introduce the trajectories $x_k(t) := \varphi\left(t, x_k^{(0)}\right)$. Using this notation, we have

$$\left(T^*_{\varphi(t)}\rho_\varphi(t) \mid w\right) = \left(\rho_\varphi(t) \mid w \circ \phi(t)\right) = \sum_{k=1}^{N} z_k(t) \cdot w(x_k(t))$$

so that

$$T^*_{\varphi(t)}\rho_\varphi(t) = \sum_{k=1}^{N} z_k(t) \cdot \delta_{x_k(t)}$$

and Eq. 4 implies that (using the reproducing kernel of V)

$$v(t, \cdot) = \sum_{\ell=1}^{N} K_V(\cdot, x_\ell(t)) z_\ell(t).$$

Similarly, one gets

$$\zeta(t, \cdot) = \sigma^2 \sum_{\ell=1}^{N} K_H(\cdot, x_\ell(t)) \alpha_\ell.$$

The third equation in Eq. 4 gives, for $w \in \mathcal{B}^p$,

$$\sum_{k=1}^{N} \dot{z}_k(t) \cdot w\left(x_k^{(0)}\right) = -\sum_{k=1}^{N} z_k(t) \cdot Dv(x_k(t))w\left(x_k^{(0)}\right) - \sum_{k=1}^{N} \alpha_k \nabla \zeta(x_k(t)) \cdot w\left(x_k^{(0)}\right)$$

from which we get

$$\dot{z}_k(t) = -Dv(x_k(t))^T z_k(t) - \alpha_k \nabla \zeta(x_k(t)).$$

Using the expansions of v and ζ and the fact that $\dot{x}_k = v(t, x_k)$, we obtain the fact that Eqs. 7 and 8 provide solutions of Eq. 4 as soon as x, m and z satisfy the coupled dynamical system

$$\begin{cases} \dot{x}_k(t) = \sum\limits_{\ell=1}^{N} K_V(x_k(t), x_\ell(t))z_\ell(t) \\ \dot{m}_k(t) = \sum\limits_{\ell=1}^{N} K_H(x_k(t), x_\ell(t))\alpha_\ell \\ \dot{z}_k(t) = -\sum\limits_{\ell=1}^{N} \nabla_1 K_V(x_k(t), x_\ell(t))z_\ell(t) \cdot z_k(t) - \frac{1}{\sigma^2}\sum\limits_{\ell=1}^{N} \nabla_1 K_H(x_k(t), x_\ell(t))\alpha_k\alpha_\ell \end{cases} \tag{9}$$

(with the notation $m_k(t) = m\left(t, x_k^{(0)}\right)$). The boundary condition applied to ρ_φ and ρ_m is

$$\sum_{k=1}^{N} z_k(t) \cdot w = -\sum_{k=1}^{N} \alpha_k \nabla m\left(t, x_k^{(0)}\right) \cdot D\varphi\left(t, x_k^{(0)}\right)^{-1} w$$

yielding

$$z_k(t) = -\alpha_k D\varphi\left(t, x_k^{(0)}\right)^{-T} \nabla m\left(x_k^{(0)}\right) = -\alpha_k \nabla q(t, x_k(t))$$

Note that, given the initial positions $\left\{x_k^{(0)}\right\}$, and initial image $q^{(0)}$, the above system is uniquely specified by the choice of the scalar field α, since $z_k(0) = -\alpha_k \nabla q^{(0)}\left(x_k^{(0)}\right)$. The solutions $\{x_k, z_k\}$ then determine the controls v and ζ for all t and $x \in \mathbb{R}^d$, which define in turn the evolving image q. This will allow us to design a shooting method for computing metamorphoses that will look for initial conditions that bring trajectories to a desired endpoint.

4 Discrete relaxed problem

Equation 9 are optimality equations for the following relaxation of Problem (3):

$$\frac{1}{2}\int_0^1 \|v(t)\|_V^2 dt + \frac{1}{2\sigma^2}\int_0^1 \|\zeta(t)\|_H^2 dt \longrightarrow \min$$

$$\text{subject to } \dot{x}_k(t) = v(t, x_k(t)), \ \dot{m}_k(t) = \zeta(t, x_k(t)), \ m_k(0) = q^{(0)}\left(x_k^{(0)}\right),$$

$$\text{and } m_k(1) = q^{(1)}(x_k(1)).$$

$$(10)$$

This is just (3) with boundary conditions only enforced at the initial and final points of the trajectories $x_k(t), k = 1, \ldots, N$. Because the constraints only depend on the evaluation of v and ζ along the discrete trajectories, the optimal ones should minimize their respective norms subject to the values taken at these points. Well-known results on RKHS's [7, 42] imply that these optimal solutions must assume the form

$$v(t, \cdot) = \sum_{k=1}^N K_V(\cdot, x_k(t))z_k(t)$$

$$\zeta(t, \cdot) = \sum_{k=1}^N K_H(\cdot, x_k(t))\alpha_k(t)$$

for some coefficients z and α, and that their norms are given by

$$\|v\|_V^2 = \sum_{k,\ell=1}^N z_k(t) \cdot K_V(x_k(t), x_\ell(t))z_\ell(t)$$

$$\|\zeta(t)\|_H^2 = \sum_{k,\ell=1}^N K_H(x_k(t), x_\ell(t))\alpha_k(t)\alpha_\ell(t).$$

Solutions of Problem (10) are therefore solutions of the reduced problem

$$\frac{1}{2} \sum_{k,\ell=1}^N \int_0^1 z_k(t) \cdot K_V(x_k(t), x_\ell(t))z_\ell(t)dt \qquad (11)$$

$$+\frac{1}{2\sigma^2} \sum_{k,\ell=1}^N \int_0^1 K_H(x_k(t), x_\ell(t))\alpha_k(t)\alpha_\ell(t)dt \longrightarrow \min$$

$$\text{subject to } \ \dot{x}_k(t) = \sum_{\ell=1}^N K_V(x_k(t), x_\ell(t))z_\ell(t),$$

$$\dot{m}_k(t) = \sum_{\ell=1}^N K_H(x_k(t), x_\ell(t))\alpha_\ell(t),$$

$$m_k(0) = q^{(0)}\left(x_k^{(0)}\right) \text{ and } m_k(1) = q^{(1)}(x_k(1)).$$

The PMP associated to this problem derives, as before, from a control-dependent Hamiltonian

$$H_{\alpha,z}(p_x, p_m, x, m) = \sum_{k,\ell=1}^{N} p_{x,k}(t) \cdot K_V(x_k(t), x_\ell(t)) z_\ell(t)$$

$$+ \sum_{k,\ell=1}^{N} K_H(x_k(t), x_\ell(t)) p_{m,k}(t) \alpha_\ell(t)$$

$$- \frac{1}{2} \sum_{k,\ell=1}^{N} z_k(t) \cdot K_V(x_k(t), x_\ell(t)) z_\ell(t) dt$$

$$- \frac{1}{2\sigma^2} \sum_{k,\ell=1}^{N} K_H(x_k(t), x_\ell(t)) \alpha_k(t) \alpha_\ell(t) \qquad (12)$$

It is then easy to check that the optimality conditions $\partial_z H = 0$ and $\partial_\alpha H = 0$ imply that $p_x = z$ and $p_m = \alpha$; from $\partial_m H = 0$, one finds that α is constant; finally, the equation $\dot{z} = -\partial_x H$ yields an equation identical to the evolution of z in Eq. 9.

The boundary condition for Problem (11) is

$$z_k(1) = -\alpha_k \nabla q^{(1)}(x_k(1)).$$

This identity propagates over time as follows: define $\tilde{m}(t) \in H$ by $\partial_t \tilde{m} = \zeta(t) \circ \varphi(t)$ with $\tilde{m}(1) = q^{(1)} \circ \varphi(1)$. Define $\tilde{q}(t)$ such that $\tilde{m}(t) = \tilde{q}(t) \circ \varphi(t)$. Then

$$z_k(t) = -\alpha_k \nabla \tilde{q}(t, x_k(t))$$

at all times. To prove this statement write

$$\partial_t \nabla \tilde{m}(t) = D\varphi(t)^T \nabla \zeta(t) \circ \varphi(t)$$

on the first hand, and, on the other hand,

$$\partial_t \nabla \tilde{m}(t) = \partial_t (D\varphi(t)^T \nabla \tilde{q}(t) \circ \varphi(t))$$
$$= D\varphi(t)^T Dv(t) \circ \varphi(t)^T \nabla \tilde{q}(t) \circ \varphi(t) + D\varphi(t)^T \partial_t (\nabla \tilde{q}(t) \circ \varphi(t)).$$

Identifying the expressions, we find

$$\partial_t (\nabla \tilde{q}(t, x_k(t))) = -Dv(t, x_k(t))^T \nabla \tilde{q}(t, x_k(t)) + \nabla \zeta(t, x_k(t)).$$

This implies

$$\partial_t (z_k(t) + \alpha_k \nabla \tilde{q}(t, x_k(t))) = -Dv(t, x_k(t))^T (z_k(t) + \alpha_k \nabla \tilde{q}(t, x_k(t)))$$

proving that $D\varphi(t, x_k(0))^T (z_k(t) + \alpha_k \nabla \tilde{q}(t, x_k(t)))$ is conserved along the motion. This quantity therefore vanishes at all times as soon as it vanishes at time $t = 1$.

Note that this boundary condition differs from the one we had in the unrelaxed problem, because $\tilde{m}$ and $\tilde{q}$ are not necessarily identical to m and q. We have, actually, $q(t, x_k(t)) = \tilde{q}(t, x_k(t))$ for all k and t, since they have the same derivative and coincide at $t = 1$, but this identity does not hold for the the full functions $q(t, \cdot)$ and $\tilde{q}(t, \cdot)$, since the constraints at $t = 1$ only involve the particles. Note also that, if one initializes system (9) with $z_k(0) = -\alpha_k \nabla q^{(0)}\left(x_k^{(0)}\right)$, one also gets $z_k(t) =$

$-\alpha_k \nabla q(t, x_k(t))$ at all times. This can be an interesting constraint to enforce, since it is consistent with the continuous problem, even though this does not provide a solution of the relaxed problem.

5 Solution of the discrete problem

The discrete relaxation that was described in the previous sections has several advantages. The first one is that it will allow us to use a particle-based method to solve the metamorphosis problem, in a Lagrangian setting (i.e., using a shooting algorithm that is parametrized by the original conditions of the evolution equations). The algorithm is described in the remainder of this section. The second advantage is to lend itself to multi-grid methods in a rather straightforward way, since the grid size (the number of particles) can be freely specified when designing the approximation. One can in this way construct a family of nested problems simply by using nested grids in the approximation. One last important advantage is that using a shooting algorithm immediately provides a representation of the target in a "template-centered coordinate system" via the initial value of the pair (α, z), or simply α if the condition $z = -\alpha \nabla q$ is enforced. The interest of this representation is that it can be analyzed using linear methods (see Section 6.2).

One of the drawback of particle methods is that they become computationally intensive when the number of particles is large. This is specially true in our case because the computation involves several sums of kernel functions centered over particles which for a non-regular grid. Some acceleration can be found using fast multipole methods [14, 15, 21], which however have the drawback of breaking the positive definiteness of the kernel [27]. Hardware acceleration is also possible. While our implementation uses multiple processors, a significant gain would results from graphical processor implementations, for which particle methods are well adapted [21, 28].

We now describe a shooting method for the solution of Problem (11), in which we solve for $(\alpha_1, \ldots, \alpha_N)$ and $\left(z_1^{(0)}, \ldots, z_N^{(0)}\right)$ such that the solution of Eq. 9 initialized at $x_k(0) = x_k^{(0)}$, $m_k(0) = q^{(0)}\left(x_k^{(0)}\right)$ and $z_k(0) = z_k^{(0)}$ satisfies $m_k(1) = q^{(1)}(x_k(1))$ for $k = 1, \ldots, N$. Considering $x_k(\cdot)$ and $m_k(\cdot)$ as functions of α and $z^{(0)}$, we minimize

$$E(\alpha, z^{(0)}) = \sum_{k=1}^{N} (m_k(1) - q^{(1)}(x_k(1)))^2. \tag{13}$$

Here, we assume that $q^{(1)}$ is defined and known everywhere (by interpolation, for example). Computing the differential of E gives

$$dE = 2 \sum_{k=1}^{N} (m_k(1) - q^{(1)}(x_k(1)))(dm_k(1) - \nabla q^{(1)}(x_k(1)) \cdot dx_k(1)) \tag{14}$$

where dm_k and dx_k are differentials dual to infinitesimal changes in the discrete variables m_k and x_k.

To compute dE, we apply the well-known adjoint method to compute derivatives of functions of solutions of dynamical systems. Writing $\theta(t) = (x, m, z)$, and defining F so that Eq. 9 is $\dot{\theta} = F(\theta, \alpha)$, we let $\boldsymbol{\theta}(t, \theta^{(0)}, \alpha)$ denote the solution of this equation with initial condition $\theta(0) = \theta^{(0)}$ and parameter α. Given variations $\delta\alpha$ and $\delta\theta^{(0)}$, then

$$\delta\theta(t) := \partial_{\theta^{(0)}}\boldsymbol{\theta}.\delta\theta^{(0)} + \partial_\alpha\boldsymbol{\theta}.\delta\alpha$$

satisfies the ODE

$$\partial_t\delta\theta = \partial_\theta F(\theta, \alpha).\delta\theta + \partial_\alpha F(\theta, \alpha).\delta\alpha$$

with initial condition $\delta\theta(0) = \delta\theta^{(0)}$. Introduce the solution $\xi = (\xi_x, \xi_m, \xi_z)$ of the adjoint ODE

$$\partial_t\xi = -\partial_\theta F(\theta, \alpha).\xi$$

so that

$$\partial_t(\xi \cdot \delta\theta) = \xi \cdot \partial_\alpha F(\theta, \alpha).\delta\alpha. \tag{15}$$

If one takes

$$\xi(1) = \{-2(m_k(1) - q^{(1)}(x_k(1))\nabla q^{(1)}(x_k(1)), 0, 2(m_k(1) - q^{(1)}(x_k(1)), \}_{k=1}^N \tag{16}$$

then, from Eqs. 14 and 15,

$$dE.\delta\theta = \xi(1) \cdot \delta\theta(1) = \xi(0) \cdot \delta\theta^{(0)} + \left(\int_0^1 \partial_\alpha F(\theta, \alpha).\xi dt\right)^T \delta\alpha.$$

In other terms, defining $\xi(t)$ and $\eta(t)$ as solutions of the system

$$\begin{cases} \partial_t\xi = -\partial_\theta F(\theta, \alpha)^T\xi \\ \partial_t\eta = -\partial_\alpha F(\theta, \alpha)^T\xi \end{cases} \tag{17}$$

with $\xi(1)$ as above and $\eta(1) = 0$, one finds

$$\partial_{\theta^{(0)}}E = \xi(0) \text{ and } \partial_\alpha E = \eta(0).$$

Detailed expressions for system (17) expressed in terms of x, α and z are provided in the Appendix.

This system is used for the adjoint method to transport the discrete covector dE backwards in time, in order to find a descent direction for the optimization. In our implementation, the initial conditions $m^{(0)}$ and $x^{(0)}$ are fixed, and the optimization only operates on $z^{(0)}$ and α, yielding Algorithm 1.

If, as discussed at the end of Section 4, the minimization is run with the constraint $z_k^{(0)} = -\alpha_k\nabla q^{(0)}\left(x_k^{(0)}\right)$, the gradients obtained at step 1.2 of Algorithm 1 only have to be combined into $\tilde{\eta}_k(0) = \eta_k(0) - \nabla q^{(0)}\left(x_k^{(0)}\right) \cdot \xi_{z,k}(0)$ in order to update α. Note also that the obtained derivatives, $\xi_z(0)$ and $\eta(0)$ (or $\tilde{\eta}(0)$) can be conditioned according to their natural inner product before performing step 2, using the linear transformation $\eta(0) \mapsto \mathcal{K}_H(x^{(0)})^{-1}\eta(0)$ and $\xi_z(0) \mapsto \mathcal{K}_V(x^{(0)})^{-1}\xi_z(0)$, where $\mathcal{K}_H(x^{(0)})$ is the matrix with entries $K_H\left(x_k^{(0)}, x_l^{(0)}\right)$ and $\mathcal{K}_V(x^{(0)})$ is formed similarly with d by d blocks $K_V\left(x_k^{(0)}, x_l^{(0)}\right)$.

Algorithm 1 Shooting algorithm

Require: template $q^{(0)}$, target $q^{(1)}$; specify kernels K_V, K_H; matching parameter σ
$\quad \alpha \leftarrow 0, z^{(0)} \leftarrow 0$
$\quad$ **while** (not stop CG) **do**
$\quad\quad$ 1. Compute $\partial_{z^{(0)}} E$, $\partial_\alpha E$:
$\quad\quad\quad$ 1.1 Compute $dE = \partial_{x_k} E \, dx_k + \partial_{m_k} E \, dm_k$ given by Eq. 14
$\quad\quad\quad$ 1.2 Compute $\xi_z(0)$, $\eta(0)$: solve the adjoint system backwards in time
$\quad$ starting from dE at $t = 1$.
$\quad\quad$ 2. Update conjugate direction and perform line search
$\quad\quad$ 3. Update $z^{(0)}$, α
$\quad$ **end while**

---

6 Numerical experiments

We now illustrate our method with some simple numerical experiments. We used
Python for our implementation, making extensive use of the open source packages
Numpy, Scipy, and the f2py tool to integrate Fortran and Python [29]. The results in
the examples below are visualized using Paraview [23].

For all numerical results, we use

$$K_V(x, y) = \left(1 + u + 3u^2/7 + 2u^3/21 + u^4/105\right) e^{-u}.\mathrm{Id}_{\mathbb{R}^d}$$

and

$$K_H(x, y) = \left(1 + \tilde{u} + \tilde{u}^2/3\right) e^{-\tilde{u}}$$

with $u = |x - y|/\tau_V$ and $\tilde{u} = |x - y|/\tau_H$, where τ_V and τ_H are
width parameters associated to the reproducing kernels. These kernels provide
RKHS's equivalent to Sobolev spaces $\mathcal{H}^k(\mathbb{R}^d, \mathbb{R}^d)$ and $\mathcal{H}^r(\mathbb{R}^d)$ with $k = (9 + d)/2$ and $r = (5 + d)/2$, yielding respective inclusions in $\mathcal{B}^4$ and
$C_0^2(\mathbb{R}^d)$. All experiments are discretized on a 2D grid with isotropic resolution
$\Delta x_1 = \Delta x_2 = 1$.

6.1 Two-image comparisons

The first examples match images from the training set in the MNIST character recog-
nition database: the letter "D" and the digit "8". We use a discrete square with 72^2
points and a time discretization $\Delta t = 0.1$ (10 timesteps). Images from the charac-
ter database are upsampled at the sampling rate for this grid. We used $\tau_V = 1.5$ and
$\tau_h = 0.5$. Figure 1 illustrates the matching of two versions of the letter D (bottom
row at left, to bottom row at right). The top row shows the optimal evolution of the
template $m(t)$, while the bottom row shows the evolution of the deformed template
$q(t) = m(t) \circ \varphi(t)^{-1}$. Figure 2 shows matching of versions of the digit eight (top
left to bottom right), along with the deformed gridlines to visualize the minimizing
deformation.

Fig. 1 Morphing of letter D from MNIST training set: *top row* shows evolution of the template; *bottom row* shows evolution of deformed template

In Fig. 3, we show the metamorphosis of two leaves from the LeafSnap database [30], after downsampling the images to a grid of 100^2 and converting to grayscale images. Here, $\tau_V = 3.0$ and $\tau_h = 0.5$.

6.2 Momentum representation

Figure 4 shows the minimizing momenta α when matching the image on the top row to each of the seven images of the final row (which shows the final morphed
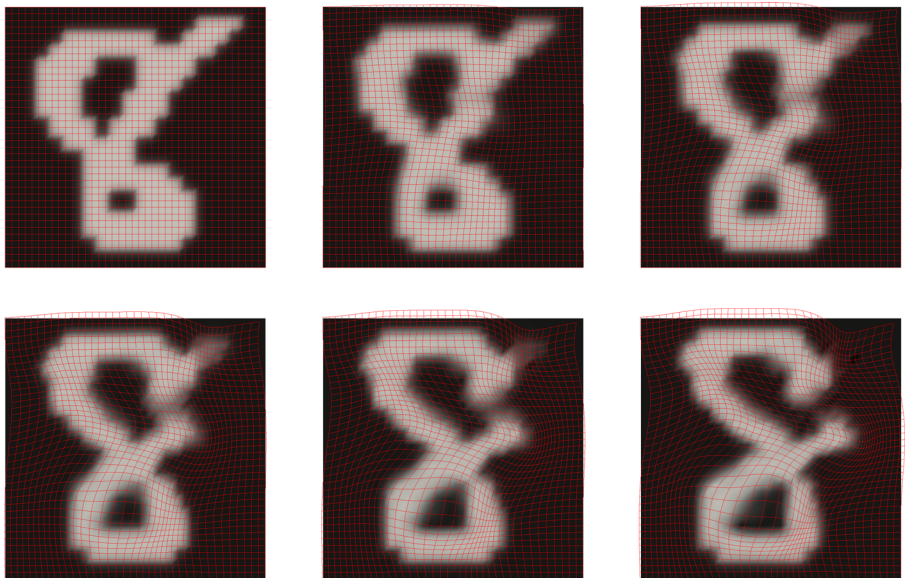

Fig. 2 Morphing of smoothed version of digit 8 from MNIST training set, where the coordinate grid is warped by the diffeomorphism and illustrated with grid lines
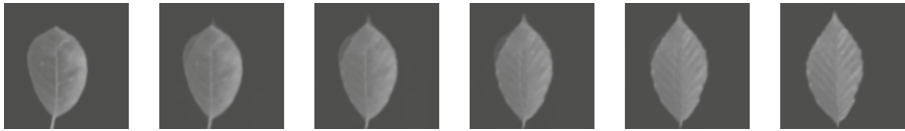
Fig. 3 Matching of two leaves from LeafSnap database using metamorphosis

image); the second row is an intensity map of the momenta. On the linear space of momenta, we can take linear combinations, as depicted in Fig. 5; this allows us to generate random images based on the ones obtained in Fig. 4, by solving (9) with initial momentum

$$\alpha(0) = \bar{\alpha}_0 + \frac{c}{\sqrt{n}} \sum_{k=1}^{7} \xi_k (\alpha_{0,k} - \bar{\alpha}_0) \tag{18}$$

where $\xi_1, \ldots, \xi_7$ are independent standard Gaussian random variables, $\alpha_{0,k}$ is the initial momentum obtained for the kth image in Fig. 4 and $\bar{\alpha}_0$ is their average. The covariance structure of the resulting random momentum $\alpha(0)$ coincides with the empirical covariance estimated from the seven examples.

6.3 Multiscale approach

Figure 6 provides an example in which two cardiac images are compared. The original images were 92×112, and we ran a multiscale procedure in which images were first smoothed and down-sampled by a factor four before a first run of the shooting algorithm was made, then up-sampled twice by a factor two, each time running the shooting algorithm initialized with the values of α obtained at the previous run.
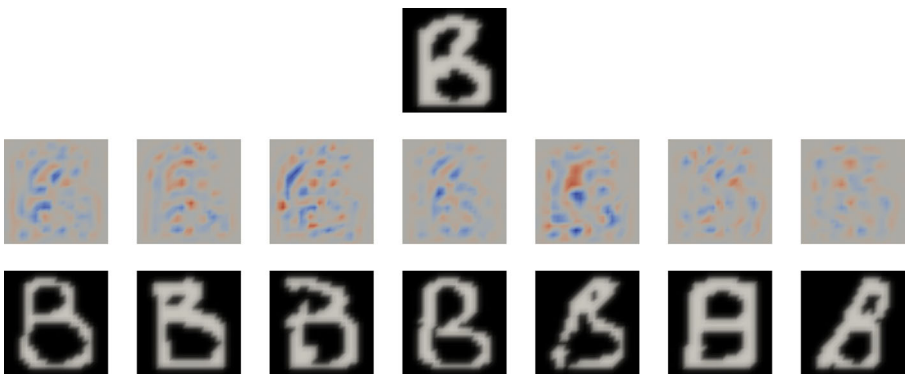

Fig. 4 Momentum field (α) for matching template to several targets for letter B in MNIST training set; figure in *top row* is the image chosen to be the template; *second row* shows α with color intensity indicating magnitude, negative values of α are colored *blue*, positive are *red*; *third row* shows final morphed image shooting from the α above

Fig. 5 Result of shooting with random momenta (described in Eq. 18) learned from letter matching. The figure provides ten independent samples

7 Rigorous results

7.1 Notation and preliminary results

We first recall our main assumptions. Images (m or q) belong to a Hilbert space H, with norm equivalent to the $\mathcal{H}^r(\mathbb{R}^d)$ norm for some integer $r \geq 0$, with notation for the $\mathcal{H}^r$ norm

$$\|u\|_{r,2}^2 = \sum_{|\alpha| \leq r} \|\partial_\alpha u\|_2^2$$

where α denotes a d-dimensional multi-index $(\alpha_1, \ldots, \alpha_d)$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$,

$$\partial_\alpha u = \frac{\partial^{|\alpha|} u}{\partial^{\alpha_1} x_1 \ldots \partial^{\alpha_d} x_d}$$

and $\| \ \|_2$ is the L^2 norm. We will use the usual notation $\mathcal{H}^r(\mathbb{R}^d)^* = \mathcal{H}^{-r}(\mathbb{R}^d)$. Most of the time, we will assume that $r > d/2 + k$ for some $k \geq 0$, which implies [2, 11] that H is continuously embedded in the space $C_0^k(\mathbb{R}^d)$ of k-times continuously differentiable functions that vanish at infinity, together with their first k derivatives, with norm

$$\|u\|_{k,\infty} = \sum_{|\alpha| \leq k} \|\partial_\alpha u\|_\infty.$$

We have denoted $\mathcal{B}^p$ the space $C_0^p(\mathbb{R}^d, \mathbb{R}^d)$, with norm $\| \ \|_{p,\infty}$, and we will denote $\| \ \|_{p,\infty,*}$ the associated norm on the dual space $(\mathcal{B}^p)^*$. We will assume that V is a Hilbert space which is continuously embedded in $\mathcal{B}^p$, with $p \geq \max(r, 1)$ at least, and $p \geq r + 1$ most of the time. If $v \in L^2([0, 1], \mathcal{B}^p)$ (which contains $L^2([0, 1], V)$), the associated flow, $\varphi^v(s, t, \cdot)$, solution of $\partial_t \varphi^v = v(t, \varphi^v)$ with $\varphi^v(s, s, x) = x$ takes values in $\text{Diff}^p(\mathbb{R}^d)$, the group of diffeomorphisms ψ such that $\psi - \text{id}$ and $\psi^{-1} - \text{id}$ both belong to $\mathcal{B}^p$. More precisely [3, 37, 45], there exists a continuous function c such that, for all $s, t \in [0, 1]$,

$$\|\varphi^v(s, t, \cdot) - \text{id}\|_{p,\infty} \leq c\left(\|v\|_{L^2([s,t],\mathcal{B}^p)}\right)\|v\|_{L^2([s,t],\mathcal{B}^p)}.$$
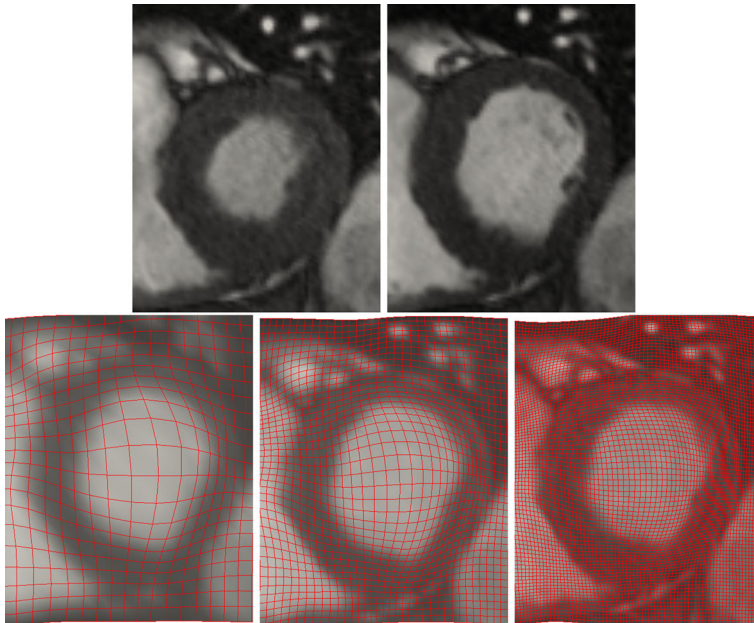
Fig. 6 Multiscale procedure applied to a cardiac image. From *top to bottom* and from *left to right*: First compared image (template), second compared image (target), final metamorphosis after down-sampling by four in every direction, final metamorphosis after down-sampling by two in every direction, final metamorphosis at original resolution

In the following, we will use the generic notation $c(\cdot)$ to represent some continuous function of its arguments (the actual function can change from an equation to another, even if we still denote it c). The notation cst will denote a generic constant.

The mapping $v \mapsto \varphi^v(s, t, \cdot)$ is differentiable from $L^2([0, 1], \mathcal{B}^p)$ to Diff^{p-1} with derivative

$$\partial_v \varphi^v(s, t, \cdot).h = \int_s^t D\varphi^v(u, t, \varphi^v(s, u, \cdot))h(u, \varphi^v(s, u, \cdot))du.$$

Moreover, one can show that, if $v, \tilde{v} \in L^2([s, t], \mathcal{B}^p)$, then

$$\|\varphi^v(s, t, \cdot) - \varphi^{\tilde{v}}(s, t, \cdot)\|_{p-1,\infty} \le c\left(\|v\|_{L^2([s,t],\mathcal{B}^p)}, \|\tilde{v}\|_{L^2([s,t],\mathcal{B}^p)}\right)\|v - \tilde{v}\|_{L^2([s,t],\mathcal{B}^p)}.$$

Note that $\|v\|_{L^2([s,t],\mathcal{B}^p)}$ is bounded, up to a multiplicative constant, by $\|v\|_{L^2([s,t],V)}$.

Finally, we note that weak convergence of a sequence v_n to a limit v in $L^2([0, 1], V)$ implies that φ^{v_n} converges to φ^v in the (p, ∞) norm over compact subsets of $\mathbb{R}^d$ [18, 45].

To simplify our expressions, we will simply denote $\varphi(t, x) = \varphi(0, t, x)$ when $s = 0$.

We let $\mathrm{Diff}_V \subset \mathrm{Diff}^p$ denote the group of diffeomorphisms that can be obtained from flows associated to some $v \in L^2([0, 1], V)$. For $\psi \in \mathrm{Diff}_V$, we introduced the translation operators $\boldsymbol{T}_\psi : V \to \mathcal{B}^p$ and $\tilde{\boldsymbol{T}}_\psi : H \to H$ defined by $\boldsymbol{T}_\psi v = v \circ \psi$

and $\tilde{\boldsymbol{T}}_\psi h = h \circ \psi$. The fact that $\tilde{\boldsymbol{T}}_\psi$ maps H onto itself (with $(\tilde{\boldsymbol{T}}_\psi)^{-1} = \tilde{\boldsymbol{T}}_{\psi^{-1}}$) is a consequence of H being equivalent to $\mathcal{H}^r(\mathbb{R}^d)$ and of $p \geq r$ (see justification below). The following lemma, which can be proved by induction, describes how $\tilde{\boldsymbol{T}}_\psi$ commutes with partial derivatives.

Lemma 1 *Let α be a multi-index. Assume that $z : \mathbb{R}^d \to \mathbb{R}$ has at least $|\alpha|$ continuous derivatives, and let $\psi \in \mathrm{Diff}^p$ with $p \geq |\alpha|$. One can write $\partial_\alpha(z \circ \psi^{-1}) \circ \psi$ in the form*

$$\partial_\alpha(z \circ \psi^{-1}) \circ \psi(y) = \sum_{\beta \leq \alpha} Q_\beta^\alpha(\psi)(y)\partial_\beta z(y)$$

where $Q_\beta^\alpha(\psi)(y)$ depends on derivatives of ψ at y, and can be written as a sum of terms

$$\sigma(D\psi)(\partial_{\gamma_1}\psi_{j_1})^{\ell_1} \cdots (\partial_{\gamma_k}\psi_{j_k})^{\ell_k}$$

with $|\gamma_q| > 1$ for $q = 1, \ldots, k$ and $|\beta| + \sum_{q=1}^k \ell_q(|\gamma_q| - 1) \leq |\alpha|$. In this expression, ψ_j denotes the jth coordinate of ψ and σ is a continuous function of $D\psi$, which can be expressed as the ratio of a polynomial in the coefficients of $D\psi$ divided by $|\det D\psi|$ to some power.

This result (or a similar version of it) can be found in many places in the literature: see [19, 20, 26] and their references. This lemma implies, in particular, that

$$\sum_{|\alpha| \leq r} |\partial_\alpha \tilde{\boldsymbol{T}}_{\psi^{-1}} z|^2 \leq c(\|\psi - \mathrm{id}\|_{p,\infty}) \sum_{|\alpha| \leq r} \tilde{\boldsymbol{T}}_{\psi^{-1}} |\partial_\alpha z|^2,$$

from which one obtains the continuity of $\tilde{\boldsymbol{T}}_{\psi^{-1}}$, with the operator norm $\|\tilde{\boldsymbol{T}}_{\psi^{-1}}\|_{\mathcal{L}(H,H)}$ a continuous function of $\|\psi - \mathrm{id}\|_{p,\infty}$.

We will use the following result. Assume that ψ_n is a sequence of diffeomorphisms of $\mathbb{R}^d$ that converges pointwise to a diffeomorphism ψ, and such that $\|\psi_n^{-1}\|_{1,\infty}$ is bounded. Then, for any $z \in L^2(\mathbb{R}^d)$, $z \circ \psi_n$ converges in L^2 to $z \circ \psi$. This can be proved by using the fact that for any $\varepsilon > 0$, one can find a compact subset of $\mathbb{R}^d$, A_ε, such that z is continuous on A_ε, $A_\varepsilon^c = \mathbb{R}^d \setminus A_\varepsilon$ has measure less than ε and $\|z\mathbf{1}_{A_\varepsilon^c}\|_2 \leq \varepsilon$. Assume without loss of generality that $\psi = \mathrm{id}$ and write

$$\|z \circ \psi_n - z\|_2^2 = \int_{\mathbb{R}^d} z^2(|\det D(\psi_n^{-1})| - 1)dx + \int_{A_\varepsilon} z(z \circ \psi_n - z)dx + \int_{A_\varepsilon^c} z(z \circ \psi_n - z)dx.$$

The last integral is less than $\|z \circ \psi_n - z\|_2 \|z\mathbf{1}_{A_\varepsilon^c}\|_2 \leq \mathrm{cst.}\|z\|_2\varepsilon$ and the rest can be made arbitrarily small by letting n go to infinity.

This result combined with Lemma 1 implies that, if $z \in \mathcal{H}^r(\mathbb{R}^d)$, then $\psi \mapsto \tilde{\boldsymbol{T}}_\psi z$ is continuous in ψ as a function from Diff^p ($p \geq r$) to $\mathcal{H}^r(\mathbb{R}^d)$ (or H). More generally, if $\psi_n \in \mathrm{Diff}^p$ and its p first derivatives converge to those of $\psi \in \mathrm{Diff}^p$ pointwise, with $\|\psi_n^{-1}\|_{p,\infty}$ bounded, then $\tilde{\boldsymbol{T}}_{\psi_n} z$ converges to $\tilde{\boldsymbol{T}}_\psi z$ in $\mathcal{H}^r(\mathbb{R}^d)$. Finally $\tilde{\boldsymbol{T}}_\psi$ is, in addition, differentiable in ψ in the following setting. If $z \in \mathcal{H}^{r+1}(\mathbb{R}^d)$,

then $\psi \mapsto \tilde{T}_\psi z$ is differentiable, as a function from Diff^p to H, with differential $u \mapsto \nabla z \circ \psi \cdot u$. Indeed, starting with a smooth z, one writes

$$\tilde{T}_{\psi+u}z - \tilde{T}_\psi z - (\nabla z \circ \psi) \cdot u = \int_0^1 (\nabla z \circ (\psi + \varepsilon u) - \nabla z \circ \psi) \cdot u \, d\varepsilon.$$

An application of Leibnitz formula yields

$$\|(\nabla z \circ (\psi + \varepsilon u) - \nabla z \circ \psi) \cdot u\|_{r,2} \le \mathrm{cst} \|\nabla z \circ (\psi + \varepsilon u) - \nabla z \circ \psi\|_{r,2} \|u\|_{r,\infty}$$

yielding

$$\|\tilde{T}_{\psi+u}z - \tilde{T}_\psi z - (\nabla z \circ \psi) \cdot u\|_{r,2} \le \mathrm{cst} \|u\|_{r,\infty} \int_0^1 \|\nabla z \circ (\psi + \varepsilon u) - \nabla z \circ \psi\|_{r,2} d\varepsilon,$$

which can be extended to arbitrary $z \in \mathcal{H}^{r+1}(\mathbb{R}^d)$ by density. The conclusion then follows from the continuity of $\psi \mapsto \nabla z \circ \psi$ as an $\mathcal{H}^r(\mathbb{R}^d, \mathbb{R}^d)$-valued mapping, since $\nabla z \in \mathcal{H}^r(\mathbb{R}^d, \mathbb{R}^d)$. From this, it also follows that $\psi \mapsto \tilde{T}_\psi^* \rho$ is differentiable in ψ, for the $\mathcal{H}^{-r}(\mathbb{R}^d)$ norm, as soon as $\rho \in \mathcal{H}^{1-r}(\mathbb{R}^d)$.

We will also be interested, for $\psi \in G_V$, in the operator $L_\psi = \tilde{T}_\psi^* A_H \tilde{T}_\psi$, where A_H is, as before, the duality isometry from H to H^*, with inverse K_H. L_ψ provides a bounded invertible mapping from H to H^*, and one has

$$\left(L_{\psi^{-1}} z \mid z \right) = \|z \circ \psi^{-1}\|_H^2.$$

Note that $\|L_\psi\|_{\mathcal{L}(H,H^*)} \le \|\tilde{T}_\psi\|_{\mathcal{L}(H)}^2$ and, using $L_\psi^{-1} = \tilde{T}_{\psi^{-1}} K_H \tilde{T}_{\psi^{-1}}^*$, $\|L_\psi^{-1}\|_{\mathcal{L}(H^*,H)} \le \|\tilde{T}_{\psi^{-1}}\|_{\mathcal{L}(H)}^2$. More generally, if $\psi \in \mathcal{B}^{r+k}$, then L_ψ maps $\mathcal{H}^{r+k}(\mathbb{R}^d)$ to $\mathcal{H}^{k-r}(\mathbb{R}^d)$ and $\|L_\psi\|_{\mathcal{L}(\mathcal{H}^{r+k}, \mathcal{H}^{k-r})} \le \mathrm{cst} \|\tilde{T}_\psi\|_{\mathcal{L}(\mathcal{H}^{r+k})} \|\tilde{T}_\psi\|_{\mathcal{L}(\mathcal{H}^{r-k})}$. Similarly, L_ψ^{-1} maps $\mathcal{H}^{k-r}(\mathbb{R}^d)$ to $\mathcal{H}^{r+k}(\mathbb{R}^d)$ with $\|L_\psi\|_{\mathcal{L}(\mathcal{H}^{k-r}, \mathcal{H}^{r+k})} \le \mathrm{cst} \|\tilde{T}_{\psi^{-1}}\|_{\mathcal{L}(\mathcal{H}^{r+k})} \|\tilde{T}_{\psi^{-1}}\|_{\mathcal{L}(\mathcal{H}^{r-k})}$.

From the differentiability of $\tilde{T}_\psi$ and $\tilde{T}_\psi^*$, one obtains the fact that $L_\psi z$ and $L_\psi^{-1} \rho$ are differentiable in ψ as soon as $z \in \mathcal{H}^{r+1}(\mathbb{R}^d)$ and $\rho \in \mathcal{H}^{1-r}(\mathbb{R}^d)$ (note that A_H maps $\mathcal{H}^{r+1}$ onto $\mathcal{H}^{1-r}$). One can go a little further by assuming that $p \ge r+1$ and that the norm on H results from a differential operator, i.e.,

$$\|z\|_H^2 = \left\| \sum_{|\alpha| \le r} b_\alpha \partial_\alpha z \right\|_2^2$$

for some coefficients b_α. One has, in this case,

$$\|z \circ \psi^{-1}\|_H^2 = \int_{\mathbb{R}^d} \left(\sum_{|\alpha| \le r} b_\alpha \circ \psi \, \partial_\alpha (z \circ \psi^{-1}) \circ \psi \right)^2 |\det D\psi| dy$$

and using Lemma 1 to expand the partial derivatives, one sees that the integrand can be written as a polynomial in the partial derivatives of z, with coefficients expressed as smooth functions of ψ and its first r derivatives. From this, one concludes that $L_{\psi^{-1}}$ is differentiable in ψ^{-1} for the $\mathcal{L}(H, H^*)$ operator norm, and so is the inverse map $L_{\psi^{-1}}^{-1}$.

Finally, let $\varphi : [0, 1] \to G_V$ be a continuous mapping (e.g., $\varphi = \varphi^v$ for some $v \in L^2([0, 1], \mathbb{R}^d)$). Define the operator

$$\boldsymbol{R}_\varphi = \int_0^1 \boldsymbol{L}_{\phi(t)^{-1}}^{-1} dt = \int_0^1 \tilde{\boldsymbol{T}}_{\varphi(t)} \boldsymbol{K}_H \tilde{\boldsymbol{T}}_{\varphi(t)}^* dt,$$

defined on H^*, with values in H. This operator is continuous in φ (for $\|\varphi\| = \sup_{t \in [0,1]} \|\varphi\|_{p,\infty}$), and is invertible. To prove the last statement, first notice that $\boldsymbol{R}_\varphi$ has closed range. Indeed, if $\boldsymbol{R}_\varphi \rho_n \to \xi$, then ρ_n is bounded because

$$\left(\rho_n \mid \boldsymbol{R}_\varphi \rho_n \right) = \int_0^1 \|\tilde{\boldsymbol{T}}_{\varphi(t)}^* \rho_n\|_{H^*}^2 dt \geq \left(\int_0^1 \|\tilde{\boldsymbol{T}}_{\varphi(t)^{-1}}\|_{\mathcal{L}(H)}^{-2} dt \right) \|\rho_n\|_{H^*}^2 \qquad (19)$$

so that

$$\|\rho_n\|_{H^*} \leq \|\boldsymbol{R}_\varphi \rho_n\|_H \left(\int_0^1 \|\tilde{\boldsymbol{T}}_{\varphi(t)^{-1}}\|_{\mathcal{L}(H)}^{-2} dt \right)^{-1}.$$

This implies that ρ_n has a weakly converging subsequence in H^*, say $\rho_n \rightharpoonup \rho$, which implies $\boldsymbol{R}_\varphi \rho_n \rightharpoonup \boldsymbol{R}_\varphi \rho$ so that $\xi = \boldsymbol{R}_\varphi \rho$. Thus, $\boldsymbol{R}_\varphi$ is one-to-one and has closed range, which implies that it is invertible.

From Eq. 19 and a similar upper bound for the inverse, we obtain the fact that $\|\boldsymbol{R}_\varphi\|_{\mathcal{L}(H^*, H)}$ and $\|\boldsymbol{R}_\varphi^{-1}\|_{\mathcal{L}(H, H^*)}$ are bounded by continuous functions of φ. From this, and the identity $\boldsymbol{R}_\varphi^{-1} - \boldsymbol{R}_{\varphi'}^{-1} = \boldsymbol{R}_\varphi^{-1} (\boldsymbol{R}_{\varphi'} - \boldsymbol{R}_\varphi) \boldsymbol{R}_{\varphi'}^{-1}$, it follow that $\boldsymbol{R}_\varphi^{-1}$ is also continuous in φ. The differentiability of $\boldsymbol{R}_\varphi$ in φ comes from the differentiability of $\boldsymbol{L}_\psi^{-1}$, so that $\varphi \mapsto \boldsymbol{R}_\varphi \rho$ is differentiable as soon as $\rho \in \mathcal{H}^{1-r}(\mathbb{R}^d)$. This statement holds also for $\rho \in \mathcal{H}^{-r}(\mathbb{R}^d)$ if $\| \ \|_H$ is associated to a differential operator. From these results and the continuity of the inverse map, one also concludes that $\boldsymbol{R}_\varphi^{-1} z$ is differentiable in φ if $z \in \mathcal{H}^{r+1}(\mathbb{R}^d)$ (or $\mathcal{H}^r(\mathbb{R}^d)$ if $\| \ \|_H$ is associated to a differential operator).

7.2 Existence of solutions of the boundary-value problem

We start with the existence of solutions for Problems (3) and (10).

Theorem 1 *Assume $r > d/2$ and $p \geq \max(1, r)$. Then Problems (3) and (10) have non-empty sets of solutions.*

Let $x^{(0,n)} = \{x_k^{(0,n)}\}_{k=1}^{N_n}$ be nested sets of points in $\mathbb{R}^d$ such that $\bigcup_n x^{(0,n)}$ is dense in $\mathbb{R}^d$. Let $(v^{(n)}, \zeta^{(n)}, \varphi^{(n)}, m^{(n)})$ be solutions of Problem (10) with $x^{(0)} = x^{(0,n)}$. Then, possibly after replacing them with subsequences, both $v^{(n)}$ and $\zeta^{(n)}$ weakly converge to limits v and ζ, while $\varphi^{(n)}$ and $m^{(n)}$ converge pointwise to the corresponding φ and m such that (v, ζ, φ, m) is a solution of Problem (3).

Proof Let $(v^{(n)}, \zeta^{(n)}, \varphi^{(n)}, m^{(n)})$ be a minimizing sequence for Problem (3). Then (using a subsequence if needed), the bounded sequences $v^{(n)}$ and $\zeta^{(n)}$ weakly converge to limits v and ζ in $L^2([0, 1], V)$ and $L^2([0, 1], H)$ respectively, with

$$\|v\|_{L^2([0,1],V)} \leq \liminf \|v^{(n)}\|_{L^2([0,1],V)} \text{ and } \|\zeta\|_{L^2([0,1],H)} \leq \liminf \|\zeta^{(n)}\|_{L^2([0,1],H)}.$$

This weak convergence for $v^{(n)}$ implies that $\varphi^{(n)}$ converges to φ uniformly on compact sets. For $x \in \mathbb{R}^d$, write

$$m^{(n)}(t, x) - m(t, x) = \int_0^t (\zeta^{(n)}(s, \varphi^{(n)}(s, x)) - \zeta^{(n)}(s, \varphi(s, x)))ds$$

$$+ \int_0^t (\zeta^{(n)}(s, \varphi(s, x)) - \zeta(s, \varphi(s, x)))ds.$$

Since the linear form

$$\zeta' \mapsto \int_0^t \zeta'(s, \varphi(s, x))ds$$

is continuous in $L^2([0, 1], H)$, the last term in the right-hand side converges to 0. Recall that K_H denote the reproducing kernel on H, defined by $K_H(\cdot, x) = \boldsymbol{K}_H \delta_x$. Rewrite the first term as

$$\int_0^t (\zeta^{(n)}(s, \varphi^{(n)}(s, x)) - \zeta^{(n)}(s, \varphi(s, x)))ds$$

$$= \int_0^t \left\langle K_H(\cdot, \varphi^{(n)}(s, x)) - K_H(\cdot, \varphi(s, x)), \zeta^{(n)}(s, \cdot) \right\rangle_H ds$$

$$\leq \left(\int_0^1 \| K_H(\cdot, \varphi^{(n)}(s, x)) - K_H(\cdot, \varphi(s, x)) \|_H^2 ds \right)^{1/2} \| \zeta^{(n)} \|_{L^2([0,1], H)}$$

$$= \left(\int_0^1 (K_H(\varphi^{(n)}(s, x), \varphi^{(n)}(s, x)) - 2K_H(\varphi^{(n)}(s, x), \varphi(s, x)) + K_H(\varphi(s, x), \varphi(s, x)))^2 ds \right)^{1/2}$$

$$\times \| \zeta^{(n)} \|_{L^2([0,1], H)}.$$

This last term goes to 0 because $r > d/2$ implies that K_H is continuous. As a consequence, we find that $m(1) = q^{(1)}(\varphi(1, x))$ is still satisfied at the limit, implying that (v, ζ, φ, m) is a solution of Problem (3). The proof for Problem (10) is exactly the same, since the only difference is that the constraint is enforced on a finite set instead of everywhere.

Now, let $(v^{(n)}, \zeta^{(n)}, \varphi^{(n)}, m^{(n)})$ be a sequence of solutions of Problem (10) with $x^{(0)} = x^{(0,n)}$. Since (10) is a relaxation of Problem (3), the optimal cost of the former is less than the optimal cost of the latter, implying that $v^{(n)}$ and $\zeta^{(n)}$ (or a subsequence) weakly converge to v and ζ with pointwise convergence of $\varphi^{(n)}$ and $m^{(n)}$ to φ and m as above. Since the sets $x^{(0,n)}$ are nested, the constraint $m(1) = q^{(1)}(\varphi(1, x))$ is satisfied for all x in their union, and therefore everywhere in $\mathbb{R}^d$ since the union is dense. Finally, since the cost of the limit is no larger than the lim inf of the costs of the sequence, which is itself no larger than the optimal cost of Problem (3), we find that (v, ζ, φ, m) is an optimal solution of Problem (3). $\square$

The existence of solutions for the continuous problem (3) is in fact true as soon as $r \geq 0$. Indeed, one can write

$$\int_0^1 \| \zeta(t) \|_H^2 dt = \int_0^1 \left(\boldsymbol{L}_{\varphi^v(t)^{-1}} \dot{m} \mid \dot{m} \right) dt$$

since $\dot{m}(t) = \zeta \circ \varphi^v(t)$, from which it results that the optimal m at fixed v is such that $L_{\varphi^v(t)^{-1}}\dot{m}$ remains constant over time. Letting $\sigma^2 \rho_m \in H^*$ denote this constant value (the normalization by σ^2 ensures that ρ_m coincides with the one introduced in Eq. 4), we get

$$m(t) - q^{(0)} = \left(\sigma^2 \int_0^t L^{-1}_{\varphi^v(t)^{-1}} dt \right) \rho_m$$

and using $m(1) = \tilde{T}_{\varphi^v(1)} q^{(1)}$, we get

$$\sigma^2 \rho_m = R^{-1}_{\varphi^v} \left(\tilde{T}_{\varphi^v(1)} q^{(1)} - q^{(0)} \right)$$

so that

$$\int_0^1 \|\zeta(t)\|_H^2 dt = \left(R^{-1}_{\varphi^v} \left(\tilde{T}_{\varphi^v(1)} q^{(1)} - q^{(0)} \right) \,\Big|\, \tilde{T}_{\varphi^v(1)} q^{(1)} - q^{(0)} \right).$$

The optimal v must therefore minimize

$$\frac{1}{2} \int_0^1 \|v(t)\|_V^2 dt + \frac{1}{2\sigma^2} \left(R^{-1}_{\varphi^v} \left(\tilde{T}_{\varphi^v(1)} q^{(1)} - q^{(0)} \right) \,\Big|\, \tilde{T}_{\varphi^v(1)} q^{(1)} - q^{(0)} \right) \qquad (20)$$

and an argument using minimizing sequences combined with the continuity of $\tilde{T}_\psi$ and R_φ leads to the existence of a minimizer (this generalizes the result proved in [37] in the L^2 case). Of course, the discretization in Problem (10) does not make sense for $r \leq d/2$, unless one replaces point evaluation by some other continuous linear forms on H, like evaluation against test functions. This would, however, have less practical interest, since test functions do not evolve in a computationally simple way under the action of diffeomorphisms.

7.3 Optimality conditions

We pass to the necessary conditions for optimal solutions of Problem (3), and now assume that $r > d/2 + 1$ so that H is embedded in $C_0^1(\mathbb{R}^d)$. Note that, since (10) can be reduced to Problem (11), which is finite dimensional, its optimality conditions follow from the standard Pontryagin maximum principle. For the infinite-dimensional case, we have:

Theorem 2 *Assume that both $q^{(1)}$ and $q^{(0)}$ belong to $\mathcal{H}^{r+1}(\mathbb{R}^d)$. Then, if (v, ζ, φ, m) is an optimal solution of Problem (3) there exist $\rho_\varphi \in (\mathcal{B}^p)^*$ and $\rho_m \in H^*$ such that Eq. 4 is satisfied, with*

$$\left(\rho_\varphi(t) \,\big|\, w \right) + \left(\rho_m \,\big|\, \nabla q^{(1)} \circ \varphi(1) \cdot w \right) = 0$$

for all $w \in \mathcal{B}^p$.

Proof Let (v, ζ, φ, m) be an optimal solution and let $\rho_m = R^{-1}_\varphi \left(\tilde{T}_{\varphi(1)} q^{(1)} - q^{(0)} \right)$. As remarked at the end of the previous section, the optimal ζ with fixed v is given by

$$\zeta(t) = \tilde{T}_{\varphi(t)^{-1}} L^{-1}_{\varphi^v(t)^{-1}} \rho_m = K_H \tilde{T}^*_{\varphi(t)} \rho_m, \qquad (21)$$

which is consistent with Eq. 4.

We now consider the optimal v when ζ is given by Eq. 21, which minimizes

$$\frac{1}{2}\|v\|^2_{L^2([0,1],V)} + \frac{1}{2\sigma^2}\left(\mathbf{R}_\varphi^{-1}\left(\tilde{\mathbf{T}}_{\varphi(1)}q^{(1)} - q^{(0)}\right) \,\Big|\, \tilde{\mathbf{T}}_{\varphi(1)}q^{(1)} - q^{(0)}\right)$$

$$\text{subject to } \dot{\varphi}(t) = v(t) \circ \varphi(t). \qquad (22)$$

If $z \in \mathcal{H}^{r+1}(\mathbb{R}^d)$ the mapping $\varphi \mapsto \left(\mathbf{R}_\varphi^{-1}z \,\big|\, z\right)$ is differentiable with differential

$$\partial_\varphi\left(\left(\mathbf{R}_\varphi^{-1}z \,\Big|\, z\right)\right).w = -2\int_0^1 \left(\partial_\phi(\tilde{\mathbf{T}}_{\varphi(t)}^*\eta).w \,\Big|\, \mathbf{K}_H\tilde{\mathbf{T}}_{\varphi(t)}^*\eta\right)dt$$

with $\eta = \mathbf{R}_\varphi^{-1}z$.

Let $E(v,\varphi)$ denote the minimized term in Eq. 22. We assume that both $q^{(1)}$ and $q^{(0)}$ belong to $\mathcal{H}^{r+1}(\mathbb{R}^d)$, which implies that $q^{(1)} \circ \varphi(1) - q^{(0)} \in \mathcal{H}^{r+1}(\mathbb{R}^d)$ too. From the previous discussion and the expression of ρ_m, E is differentiable in φ, with

$$\partial_\varphi E.w = -\int_0^1 \left(\partial_\varphi(\tilde{\mathbf{T}}_{\varphi(t)}^*\rho_m).w(t) \,\Big|\, \zeta(t)\right)dt + \left(\rho_m \,\Big|\, \nabla q^{(1)} \circ \varphi(1) \cdot w(1)\right)$$

$$= -\int_0^1 \partial_\varphi\left(\rho_m \,\Big|\, \tilde{\mathbf{T}}_{\varphi(t)}\zeta(t)\right).w(t)dt + \left(\rho_m \,\Big|\, \nabla q^{(1)} \circ \varphi(1) \cdot w(1)\right)$$

Define $\mu(t) = \partial_\psi\left(\rho_m \,\Big|\, \tilde{T}_\psi\zeta(t)\right)_{|\psi=\varphi(t)}$. The derivative exists, since $\rho_m \in \mathcal{H}^{-r}(\mathbb{R}^d)$ and $\zeta \in \mathcal{H}^{r+1}(\mathbb{R}^d)$, and provides a a bounded linear form on $C^r(\mathbb{R}^d,\mathbb{R}^d)$. Define also the form $\nu : w \mapsto \left(\rho_m \,\big|\, (\nabla q^{(1)} \circ \varphi(1)) \cdot w\right)$, which is also bounded on $C^r(\mathbb{R}^d,\mathbb{R}^d)$. Define $\rho_\varphi(t)$ as the solution of the ODE

$$\dot{\rho}_\varphi = -\partial_\psi\left(\rho_\varphi \,\Big|\, T_{\varphi(t)}v(t)\right)_{|\psi=\varphi(t)} - \mu(t)$$

with $\rho_\varphi(1) = -\nu$ (this ODE is the third equation in Eq. 4). To see that this solution is well defined, first note that, for any given $\rho \in (\mathcal{B}^{p-1})^*$ and $w \in V$, the mapping $\psi \mapsto (\rho \,|\, w \circ \psi)$, defined on $\mathcal{B}^p$, is differentiable in ψ, with differential

$$\partial_\psi(\rho \,|\, w \circ \psi) \cdot \delta\psi = (\rho \,|\, Dw \circ \psi \cdot \delta\psi).$$

As a consequence, we have

$$\partial_\psi(\rho \,|\, w \circ \psi) \in C^{p-1}(\mathbb{R}^d,\mathbb{R}^d)^*,$$

with norm bounded by $\text{cst.}\|\rho\|_{p-1,\infty,*}\|w\|_{p,\infty}\|\psi\|_{p-1,\infty}$. The map $Q_{w,\psi} : \rho \mapsto \partial_\psi(\rho \,|\, w \circ \psi)$ therefore is a bounded linear map on $C^{p-1}(\mathbb{R}^d,\mathbb{R}^d)^*$, satisfying

$$\int_0^1 \|Q_{v(t),\varphi(t)}\|^2 dt < \infty$$

as soon as $\int_0^1 \|v(t)\|^2_{p,\infty}dt < \infty$, which is true for a minimizer of Eq. 22. Since both $\mu(t)$ and ν belong to $C_0^r(\mathbb{R}^d,\mathbb{R}^d)^* \subset C_0^{p-1}(\mathbb{R}^d,\mathbb{R}^d)^*$ (since $p \geq r+1$), the solution ρ_φ of $\dot{\rho}_\varphi = Q_{v,\varphi}\rho_\varphi + \mu$ initialized at $\rho_\varphi(1) = -\nu$ is uniquely defined over $[0,1]$.

If $\delta v \in L^2([0,1],V)$, the directional derivative $\delta\varphi := \partial_v\varphi^v.\delta v$ satisfies (since $\varphi = \varphi^v$)

$$\partial_t\delta\varphi(t) = \partial_\varphi(T_{\varphi(t)}v(t)) \cdot \delta\varphi(t) + \delta v(t) \circ \varphi(t)$$

with $\delta\varphi(0) = 0$. From the definition of ρ_φ, we have

$$\partial_t\big(\rho_\varphi \mid \delta\varphi\big) = -(\mu \mid \delta\varphi) + \big(\rho_\varphi \mid \delta v \circ \varphi\big)$$

so that

$$-\int_0^1 (\mu(t) \mid \delta\varphi(t))dt + (\nu(1) \mid \delta\varphi(1)) = -\int_0^1 \big(\rho_\varphi(t) \mid \delta v(t) \circ \varphi(t)\big)dt.$$

If v is an optimal solution of Problem (3), we must have

$$\int_0^1 (A_V v(t) \mid \delta v(t))dt - \int_0^1 \big(\rho_\varphi(t) \mid \delta v(t) \circ \varphi(t)\big)dt = 0$$

for all δv, which implies that

$$v(t) = K_V T_\varphi^* \rho_\varphi(t).$$

This is the fifth equation in Eq. 4, and completes the proof of Theorem 2. □

7.4 Existence of solutions of the initial-value problem

We now discuss the existence and uniqueness of solutions of Eq. 4 with initial conditions $\varphi(0) = \mathrm{id}$, $m(0) = m_0$, $\rho_\varphi(0) = \rho_{\varphi,0}$ and $\rho_m(0) = \rho_{m,0}$. We will assume that $\rho_{\varphi,0} \in (\mathcal{B}^{p-2})^*$ and $\rho_{m,0} \in \mathcal{H}^{1-r}(\mathbb{R}^d)$ with $p \geq r + 1$.

Since ρ_m is constant and m is obtained via quadrature given ζ and φ, we will focus on the subsystem

$$\begin{cases} \dot{\varphi}(t) = v(t) \circ \varphi(t) \\ \dot{\rho}_\varphi(t) = -\partial_{\varphi(t)}\big(\rho_\varphi(t) \mid v(t) \circ \varphi(t)\big) - \partial_{\varphi(t)}(\rho_m \mid \zeta(t) \circ \varphi(t)) \\ v(t) = K_V T_{\varphi(t)}^* \rho_\varphi(t) \\ \zeta(t) = \sigma^2 K_H \tilde{T}_{\varphi(t)}^* \rho_m \end{cases} \tag{23}$$

If $\rho \in (\mathcal{B}^{p-2})^*$ and $w \in V \subset \mathcal{B}^p$, the mapping $\psi \mapsto (\rho \mid w \circ \psi) = \big(\rho \mid T_\psi w\big)$ is differentiable in $\psi \in \mathrm{Diff}^p$ with $\partial_\psi(\rho \mid w \circ \psi).h = (\rho \mid Dw \circ \psi.h)$. One deduces from this that $\rho \mapsto \partial_\psi(\rho \mid w \circ \psi)$ is a bounded endomorphism of $(\mathcal{B}^{p-2})^*$ with operator norm bounded by $c(\|\psi - \mathrm{id}\|_{p-2,\infty})\|w\|_{p-1,\infty}$. If $\rho_m \in \mathcal{H}^{-1}(\mathbb{R}^d)$ and $\zeta \in \mathcal{H}^r(\mathbb{R}^d)$, we have $\partial_\psi(\rho_m \mid \zeta \circ \psi).h = (\rho_m \mid \nabla\zeta \circ \psi \cdot h)$.

From the expressions of v and ζ, one easily checks that

$$\partial_\psi\big(\rho_\varphi \mid v \circ \psi\big) = \frac{1}{2}\partial_\varphi\Big(\rho_\varphi \mid T_\psi K_V T_\psi^* \rho_\varphi\Big) \text{ and } \partial_\psi(\rho_m \mid \zeta \circ \psi)$$

$$= \frac{1}{2}\partial_\psi\Big(\rho_m \mid \tilde{T}_\psi K_H \tilde{T}_\psi^* \rho_m\Big).$$

One also has $\|v\|_V^2 = \big(\rho_\varphi \mid v \circ \varphi\big)$ and $\|\zeta\|_H^2 = \sigma^2(\rho_m \mid \zeta \circ \varphi)$, from which one deduces that, along any solution of Eq. 23, one has

$$\partial_t\Big(\|v\|_V^2 + \|\zeta\|_H^2/\sigma^2\Big) = 0$$

since this time derivative is equal to

$$\big(\partial_t\rho_\varphi \mid v \circ \varphi\big) + \partial_\varphi\big(\rho_\varphi \mid v \circ \varphi\big).\partial_t\varphi + \partial_\varphi(\rho_m \mid \zeta \circ \varphi).\partial_t\varphi$$

which vanishes since $\partial_t \varphi = v \circ \varphi$. This implies, in particular, that $v \in L^2([0, t], V)$ and $\zeta \in L^2([0, t], H)$ along any solution of Eq. 23 on the interval $[0, t]$.

Conversely, as soon as $v \in L^2([0, t], V)$ and $\zeta \in L^2([0, t], H)$, the equation

$$\dot{\rho}_\varphi = -\partial_\varphi(\rho_\varphi \mid v \circ \varphi) - \partial_\varphi(\rho_m \mid \zeta \circ \varphi) \tag{24}$$

is a well-defined linear equation on $(\mathcal{B}^{p-2})^*$, with a unique solution, since we assume $\rho_{\varphi,0} \in (\mathcal{B}^{p-2})^*$. Its solution can be made explicit by noting that

$$
\begin{aligned}
\partial_t\left(\rho_\varphi \mid D\varphi\, w\right) &= -\left(\rho_\varphi \mid Dv \circ \varphi\, D\varphi\, w\right) - (\rho_m \mid \nabla\zeta \circ \varphi \cdot D\varphi\, w) + \left(\rho_\varphi \mid Dv \circ \varphi\, D\varphi\, w\right) \\
&= -(\rho_m \mid \nabla(\zeta \circ \varphi) \cdot w),
\end{aligned}
$$

from which we conclude that

$$\left(\rho_\varphi(t) \mid D\varphi(t)w\right) = \left(\rho_{\varphi,0} \mid w\right) - \int_0^t (\rho_m \mid \nabla(\zeta(s) \circ \varphi(s)) \cdot w)\, ds.$$

Given this, we can summarize system (4) with a single consistency equation for v, namely, for all $w \in V$:

$$(A_V v(t) \mid w) = \left(\rho_{\varphi,0} \mid \mathbf{Ad}_{\varphi^v(t)^{-1}} w\right) - \sigma^2 \int_0^t \left(\rho_m \mid \nabla\left(L_{\varphi^v(s)^{-1}}^{-1}\rho_m\right) \cdot \mathbf{Ad}_{\varphi^v(t)^{-1}} w\right) dt, \tag{25}$$

in which we have introduced (for $\psi \in \mathrm{Diff}^{p-1}$) the "adjoint" operator $\mathbf{Ad}_\psi : w \mapsto (D\psi\, w) \circ \psi^{-1}$, as an operator from V to $\mathcal{B}^{p-2}$ and used the fact that $\zeta \circ \varphi = \sigma^2 \tilde{T}_\varphi K_H \tilde{T}_\varphi^* \rho_m = \sigma^2 L_{\varphi^{-1}}^{-1}\rho_m$. Equation 25 with $\sigma^2 = 0$ is of course the well-known momentum conservation equation over diffeomorphisms with a right-invariant metric [5, 6, 24, 32].

Let β^v denote the time-dependent linear form applied to w in the right-hand side of Eq. 25, which therefore can be summarized as $v(t) = K_V \beta^v(t)$. Fix a constant M. We first check that, for small enough t, $\beta^v(t) \in V^*$ as soon as $v \in L^2([0, t], V)$ and $\|v\|_{L^2([0,t],V)} \leq M$ implies $\|\beta^v\|_{L^2([0,t],V^*)} = \|K_V \beta^v\|_{L^2([0,t],V)} \leq M$ also.

We have, for $\psi \in \mathrm{Diff}^p$, $\mathbf{Ad}_{\psi^{-1}} w = \left((D\psi)^{-1} - \mathrm{Id}_{\mathbb{R}^d}\right) w \circ \psi + w \circ \psi$, from which one gets

$$\left(\rho_{\varphi,0} \mid \mathbf{Ad}_{\varphi^v(t)^{-1}} w\right) \leq c(\|\varphi(t) - \mathrm{id}\|_{p-1,\infty})\|\rho_{\varphi,0}\|_{p-2,\infty,*}\|w\|_{p-2,\infty}$$

Since

$$\|\varphi(t) - \mathrm{id}\|_{p-1,\infty} \leq c\left(\|v\|_{L^2([0,t],V)}\right)\|v\|_{L^2([0,t],V)}$$

we find

$$\left(\rho_{\varphi,0} \mid \mathbf{Ad}_{\varphi^v(t)^{-1}} w\right) \leq c\left(\|v\|_{L^2([0,t],V)}\right)\|\rho_{\varphi,0}\|_{p-2,\infty,*}\|w\|_{p-2,\infty}. \tag{26}$$

Since $L_{\psi^{-1}}^{-1}$ maps $\mathcal{H}^{1-r}(\mathbb{R}^d)$ onto $\mathcal{H}^{r+1}(\mathbb{R}^d)$ for $\psi \in \mathrm{Diff}^p$, we have

$$\|L_{\psi^{-1}}^{-1}\rho_m\|_{r+1,2} \leq \mathrm{cst}\|\tilde{T}_\psi\|_{L^2(\mathcal{H}^{r+1})}^2\|\rho_m\|_{1-r,2} \leq c(\|\psi\|_{r+1,\infty})\|\rho_m\|_{1-r,2}.$$

Combined with the previous estimate, this yields, for $\psi, \tilde{\psi} \in \mathrm{Diff}^p$,

$$\left(\rho_m \mid \nabla(L_{\psi^{-1}}^{-1}\rho_m) \cdot \mathbf{Ad}_{\tilde{\psi}^{-1}} w\right) \leq c(\|\psi\|_{r+1,\infty}, \|\tilde{\psi}\|_{r-1,\infty})\|\rho_m\|_{1-r,2}^2\|\|w\|_{r-2,\infty}.$$

Since $r \leq p - 1$, we can conclude that

$$\|\beta^v(t)\|_{p-2,\infty,*} \leq c(\|v\|_{L^2([0,t],V)}) \left(\|\rho_{\varphi,0}\|_{p-2,\infty,*} + t\|\rho_m\|_{1-r,2}^2\right), \qquad (27)$$

from which it follows that $\beta^v(t) \in (\mathcal{B}^{p-2})^* \subset V^*$ with

$$\|\beta^v\|_{L^2([0,t],V^*)}^2 \leq tc(\|v\|_{L^2([0,t],V)}) \left(\|\rho_{\varphi,0}\|_{p-2,\infty,*} + t\|\rho_m\|_{1-r,2}^2\right)^2.$$

If we assume that $\|v\|_{L^2([0,t],V)} \leq M$, we get $\|\beta^v\|_{L^2([0,t],V^*)} \leq M$ for $t \leq t_0$, where t_0 is chosen such that $t_0 c(M) \left(\|\rho_{\varphi,0}\|_{p-2,\infty,*} + t_0\|\rho_m\|_{1-r,2}^2\right)^2 \leq M$. It is important to notice that, beside universal constants and M, t_0 only depends on $\|\rho_{\varphi,0}\|_{p-2,\infty,*}^2$ and $\|\rho_m\|_{1-r,2}$. In the following, we take M large enough so that any solution of Eq. 4 must satisfy $\|v\|_{L^2([0,t_0],V)} \leq M$ for any $t_0 \leq 1$. This is possible since we have remarked that $\|v(t)\|_V^2 + \sigma^{-2}\|\zeta(t)\|_H^2$ remains constant along any solution of Eq. 4 so that, if $t_0 \leq 1$, one must have

$$\|v\|_{L^2([0,t_0],V)}^2 \leq \|v(0)\|_V^2 + \sigma^{-2}\|\zeta(0)\|_H^2 = \|\rho_{\varphi,0}\|_{V^*}^2 + \sigma^2\|\rho_m\|_{H^*}^2.$$

We now estimate the Lipschitz constant of $v \mapsto \beta^v$ on the ball of radius M of $L^2([0,t],V)$ for $t \leq t_0$. In the computations that follow, we will use repetitively the fact that $\|\varphi^v(t) - \mathrm{id}\|_{l,\infty} \leq c(M)$ for any $l \leq p$, as soon as $\|v\|_{L^2([0,t],V)} \leq M$ (recall that c is a notation for a generic continuous function). Recall also that $p \geq r + 1$. Writing, assuming $\max\left(\|v\|_{L^2([0,t_0],V)}, \|\tilde{v}\|_{L^2([0,t_0],V)}\right) \leq M$,

$$\mathbf{Ad}_{\varphi^v(t)^{-1}}w - \mathbf{Ad}_{\varphi^{\tilde{v}}(t)^{-1}}w = \left(D\varphi^v(t)^{-1} - D\varphi^{\tilde{v}}(t)^{-1}\right)w\circ\varphi^v(t) + D\varphi^v(t)^{-1}(w\circ\varphi^v(t) - w\circ\varphi^{\tilde{v}}(t))$$

and using Lemma 1 and Leibnitz formula, we get

$$\|\mathbf{Ad}_{\varphi^v(t)^{-1}}w - \mathbf{Ad}_{\varphi^{\tilde{v}}(t)^{-1}}w\|_{p-2,\infty} \leq c(M)\|\varphi^v(t) - \varphi^{\tilde{v}}(t)\|_{p-1,\infty}\|w\|_{p-1,\infty} \quad (28)$$

as soon as $\psi, \tilde{\psi} \in \mathrm{Diff}^{p-1}$ and $w \in \mathcal{B}^{p-1}$. This immediately implies

$$\left\|\left(\mathbf{Ad}^*_{\varphi^v(t)^{-1}} - \mathbf{Ad}^*_{\varphi(t)^{-1}}\right)\rho_{\varphi,0}\right\|_{p-1,\infty,*} \leq c(M)\|\rho_{\varphi,0}\|_{p-2,\infty,*}\|\varphi^v(t) - \varphi^{\tilde{v}}(t)\|_{p-1,\infty}.$$

Write

$$\left(\rho_m \,\middle|\, \nabla\left(L^{-1}_{\varphi^v(s)^{-1}}\rho_m\right) \cdot \mathbf{Ad}_{\varphi^v(t)^{-1}}w - \nabla\left(L^{-1}_{\tilde{\varphi}^v(s)^{-1}}\rho_m\right) \cdot \mathbf{Ad}_{\tilde{\varphi}^v(t)^{-1}}w\right)$$
$$= \left(\rho_m \,\middle|\, \left(\nabla\left(L^{-1}_{\varphi^v(s)^{-1}}\rho_m\right) - \nabla\left(L^{-1}_{\tilde{\varphi}^v(s)^{-1}}\rho_m\right)\right) \cdot \mathbf{Ad}_{\varphi^v(t)^{-1}}w\right)$$
$$+ \left(\rho_m \,\middle|\, \nabla\left(L^{-1}_{\tilde{\varphi}^v(s)^{-1}}\rho_m\right) \cdot \left(\mathbf{Ad}_{\varphi^v(t)^{-1}}w - \mathbf{Ad}_{\tilde{\varphi}^v(t)^{-1}}w\right)\right)$$
$$\leq \mathrm{cst}\,\|\rho_m\|_{1-r,2}\|L^{-1}_{\varphi^v(s)^{-1}}\rho_m - L^{-1}_{\tilde{\varphi}^v(s)^{-1}}\rho_m\|_{r,2}\|\mathbf{Ad}_{\varphi^v(t)^{-1}}w\|_{r-1,\infty}$$
$$+ \mathrm{cst}\,\|\rho_m\|_{1-r,2}\|L^{-1}_{\tilde{\varphi}^v(s)^{-1}}\rho_m\|_{r,2}\|\mathbf{Ad}_{\varphi^v(t)^{-1}}w - \mathbf{Ad}_{\tilde{\varphi}^v(t)^{-1}}w\|_{r-1,\infty} \quad (29)$$

We have (letting $\psi = \varphi^v(s)$ and $\tilde{\psi} = \varphi^{\tilde{v}}(s)$)

$$\|L_{\psi^{-1}}^{-1}\rho_m - L_{\tilde{\psi}^{-1}}^{-1}\rho_m\|_{r,2} = \|\tilde{T}_\psi K_H \tilde{T}_\psi^* \rho_m - \tilde{T}_{\tilde{\psi}} K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r,2}$$

$$\leq \|\tilde{T}_\psi\|_{r,2}\|K_H \tilde{T}_\psi^* \rho_m - K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r,2}$$

$$+\|\tilde{T}_\psi K_H \tilde{T}_{\tilde{\psi}}^* \rho_m - \tilde{T}_{\tilde{\psi}} K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r,2} \qquad (30)$$

Let us consider the last two terms separately. We have $\|\tilde{T}_\psi\|_{r,2} = c(\|\psi - \mathrm{id}\|_{r,\infty}) \leq c(M)$. Also,

$$\|K_H \tilde{T}_\psi^* \rho_m - K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r,2} \leq \mathrm{cst}\|\tilde{T}_\psi^* \rho_m - \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{-r,2}$$

$$= \mathrm{cst} \sup\left(\left(\rho_m \,\Big|\, \tilde{T}_\psi z - \tilde{T}_{\tilde{\psi}} z\right) : \|z\|_{r,2} \leq 1\right)$$

and we have

$$\left(\rho_m \,\Big|\, \tilde{T}_\psi z - \tilde{T}_{\tilde{\psi}} z\right) \leq \|\rho_m\|_{1-r,2}\|\tilde{T}_\psi z - \tilde{T}_{\tilde{\psi}} z\|_{r-1,2} \leq c(M)\|\rho_m\|_{1-r,2}\|\psi - \tilde{\psi}\|_{r-1,\infty}\|z\|_{r,2}$$

so that

$$\|\tilde{T}_\psi\|_{r,2}\|K_H \tilde{T}_\psi^* \rho_m - K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r,2} \leq c(M)\|\rho_m\|_{1-r,2}\|\psi - \tilde{\psi}\|_{r-1,\infty}. \qquad (31)$$

For the second term in Eq. 30, write

$$\|\tilde{T}_\psi K_H \tilde{T}_{\tilde{\psi}}^* \rho_m - \tilde{T}_{\tilde{\psi}} K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r,2} \leq c(M)\|\psi - \tilde{\psi}\|_{r,\infty}\|K_H \tilde{T}_{\tilde{\psi}}^* \rho_m\|_{r+1,2}$$

$$\leq c(M)\|\psi - \tilde{\psi}\|_{r,\infty}\|\tilde{T}_{\tilde{\psi}}^* \rho_m\|_{1-r,2}$$

$$\leq c(M)\|\psi - \tilde{\psi}\|_{r,\infty}\|\rho_m\|_{1-r,2}$$

From this and (30), (31), we get

$$\|L_{\psi^{-1}}^{-1}\rho_m - L_{\tilde{\psi}^{-1}}^{-1}\rho_m\|_{r,2} \leq c(M)\|\psi - \tilde{\psi}\|_{r,\infty}\|\rho_m\|_{1-r,2} \qquad (32)$$

Since

$$\|\mathbf{Ad}_{\varphi^v(t)^{-1}} w\|_{r-1,\infty} \leq c(\|\varphi^v(t) - \mathrm{id}\|_{r,\infty})\|w\|_{r-1,\infty}$$

we find that the first term in Eq. 29 is less than $c(M)\|\psi - \tilde{\psi}\|_{r,\infty}\|\rho_m\|_{1-r,2}^2\|w\|_{r-1,\infty}$.

For the second term in Eq. 29, we have $\|L_{\tilde{\varphi}^v(s)^{-1}}^{-1}\rho_m\|_{r,2} \leq c(M)\|\rho_m\|_{1-r,2}$ while, similarly to Eq. 28,

$$\|\mathbf{Ad}_{\varphi^v(t)^{-1}} w - \mathbf{Ad}_{\tilde{\varphi}^v(t)^{-1}} w\|_{r-1,\infty} \leq c(M)\|\varphi^v(t) - \varphi^{\tilde{v}}(t)\|_{r,\infty}\|w\|_{r,\infty}.$$

This finally gives the upper-bound

$$\left(\rho_m \,\Big|\, \nabla\left(L_{\varphi^v(s)^{-1}}^{-1}\rho_m\right) \cdot \mathbf{Ad}_{\varphi^v(t)^{-1}} w - \nabla\left(L_{\tilde{\varphi}^v(s)^{-1}}^{-1}\rho_m\right) \cdot \mathbf{Ad}_{\tilde{\varphi}^v(t)^{-1}} w\right)$$

$$\leq c(M)\|\rho_m\|_{1-r,2}^2\|w\|_{r,\infty} \sup_{s \leq t}\|\varphi^v(s) - \varphi^{\tilde{v}}(s)\|_{r,\infty}$$

so that (using $r \leq p - 1$)

$$\|\beta^v(t) - \beta^{\tilde{v}}(t)\|_{p-1,\infty,*} \leq c(M)\left(\|\rho_{\varphi,0}\|_{p-2,\infty,*} + t\|\rho_m\|_{1-r,2}^2\right)\sup_{s \leq t}\|\varphi^v(s) - \varphi^{\tilde{v}}(s)\|_{p-1,\infty}.$$

Using the fact that

$$\sup_{s \le t_0} \|\varphi^v(s) - \varphi^{\tilde{v}}(s)\|_{p-1,\infty} \le c(M)\|v - \tilde{v}\|_{L^2([0,t_0],V)}$$

we find that $v \mapsto K_V \beta^v$ is Lipschitz on the ball of radius M in $L^2([0, t_0], V)$, with Lipschitz constant less than $c(M)t_0 \left(\|\rho_{\varphi,0}\|_{p-2,\infty,*} + t_0\|\rho_m\|_{1-r,2}^2\right)$.

Reducing the value of t_0 if needed, one can make this upper-bound less than 1 to ensure that $v \mapsto K_V \beta^v$ has a unique fixed point in the ball of radius M in $L^2([0, t_0], V)$. This shows that system (23) has a unique solution (with the considered initial condition) over the interval $[0, t_0]$.

A valid choice for t_0 can therefore be made in terms of M, $\|\rho_{\varphi,0}\|_{p-2,\infty,*}$, and $\|\rho_m\|_{1-r,2}$ uniquely; since M can itself be chosen as a function of the last two norms, their values are sufficient to specify t_0. If we now define T_0 to be the largest time $T_0 \le 1$ such that a solution exists over all intervals $[0, t] \subset [0, T_0)$, we must have $T_0 = 1$ unless $\|\rho_\varphi(t)\|_{p-2,\infty,*}$ tends to ∞ when t tends to T_0 (recall that ρ_m is time-independent). Since $\rho_\varphi(t) = T^*_{\varphi(t)^{-1}} A_V v(t) = T_{\varphi(t)^{-1}} \beta^v(t)$, Eq. (27) shows that $\|\rho_\varphi(t)\|_{p-2,\infty,*}$ must remain bounded, showing that $T_0 = 1$ necessarily.

Since one can obviously replace the unit interval by any interval $[0, T]$, we have obtained the following result.

Theorem 3 *Assume that $p \ge 1 + d/2$ and $p \ge r + 1$. Then system (4) has a unique solution over any bounded interval as soon as $\rho_{\varphi,0} \in (\mathcal{B}^{p-2})^*$ and $\rho_m \in \mathcal{H}^{1-r}(\mathbb{R}^d)$.*

Note that, with metamorphosis, the boundary condition requires that $\left(\rho_{\varphi,0} \mid w\right) = (\rho_m \mid \nabla q_0 \cdot w)$. Assuming that $q_0 \in \mathcal{H}^1(\mathbb{R}^d)$ (which is restrictive only for $r = 0$), we see that $\rho_m \in \mathcal{H}^{1-r}(\mathbb{R}^d)$ implies that $\rho_{\varphi,0} \in (\mathcal{B}^{r-1})^* \subset (\mathcal{B}^{p-2})^*$ since $p \ge r + 1$, so that the regularity condition for $\rho_{\varphi,0}$ is automatically satisfied.

Remark In the previous result, we "lose" two derivatives in the initial condition for ρ_φ and one in ρ_m. This can be improved under more restrictive assumptions on the spaces V and H.

– Assume that the norm on H is specified by a differential operator. We have seen that $\psi \mapsto \left(\rho_m \mid L^{-1}_{\psi^{-1}} \rho_m\right)$ was a smooth function of $\psi \in \text{Diff}^p$ as soon as $\rho_m \in H$, with

$$\left(\rho_m \mid \nabla(L^{-1}_{\psi^{-1}} \rho_m) \cdot w\right) = (1/2)\partial_\psi \left(\rho_m \mid L^{-1}_{\psi^{-1}} \rho_m\right).w.$$

Using this property, one can carry on the estimates on the second term in β^v using only the assumption $\rho_m \in \mathcal{H}^{-r}$, and therefore extend the conclusion of the theorem to this case.

– If one makes the same hypothesis for V, namely that $V \sim \mathcal{H}^p(\mathbb{R}^d, \mathbb{R}^d)$ with $p \ge 1 + d/2$ and $p \ge r$ (note that this assumption only implies that V is embedded in $\mathcal{B}^1$), the associated group Diff_V is then included in the Hilbert manifold $\mathcal{D}^p$ of diffeomorphisms ψ such that $\psi - \text{id}$ and $\psi^{-1} - \text{id}$ both belong to $\mathcal{H}^p(\mathbb{R}^d, \mathbb{R}^d)$, on which the right invariant metric is a strong Riemannian metric

(i.e., the Riemannian topology coincides with the one induced by $\mathcal{H}^p(\mathbb{R}^d, \mathbb{R}^d)$). This is a consequence of Lemma 1 and of results on the stability of Sobolev spaces by products which imply that all terms $(\partial_{\gamma_1} \psi_{j_1})^{\ell_1} \cdots (\partial_{\gamma_k} \psi_{j_k})^{\ell_k} \partial_\beta z$ are square integrable as soon as $|\beta| + \sum_{q=1}^{k} \ell_q(|\gamma_q| - 1) \leq p$ [35]. (It has actually recently been showed that $\mathcal{D}^p$ coincides with Diff_V; see [12].) The right-invariant metric

$$\|(\xi, z)\|_{(\psi,q)}^2 = \|\xi \circ \psi^{-1}\|_V^2 + \sigma^{-1}\|z \circ \psi^{-1}\|_{r,2}^2$$

on the product space $\text{Diff}_V \times \mathcal{H}^r(\mathbb{R}^d)$ is then also a strong metric as soon as $r \leq p$, and since (4) is the geodesic equation on this manifold, its solutions are uniquely defined over arbitrary time intervals without loss of derivatives (see [1, 19, 20, 26, 31, 34], and the references therein, for more details).

8 Discussion

In this paper we developed new numerical tools, combined with an extension of known theoretical results, on image metamorphosis. We proposed, in particular, a particle-based optimization method for their estimation, based on the determination of initial conditions of the geodesic equation performed via a shooting method. The resulting algorithm allows for a numerically-stable sparse representation of the target image in a template-centered coordinate system, which was hard to achieve using previous methods. This improvement was made possible by the introduction of a Sobolev norm in image space, allowing for particle solutions that were not available when using an L^2 norm.

One of the limitations of the discretization scheme discussed in Section 3 is its asymmetry, since the evolving image is represented using a moving grid, x, which is specified at time $t = 0$, in the template coordinate frame (the continuous problem itself is symmetric, so that the asymmetry disappears in the discretization limit). Our scheme can, however, be modified to incorporate more symmetry by introducing a second set of particles, this time defined in the target coordinate frame. More precisely, one can add to Problem (10) another set of constraints, associated to a new grid y and image value n (in addition to x and m) in the form $\dot{y}_k = v(t, y_k)$, $\dot{n}_k = \zeta(t, y_k)$, $n_k(0) = q^{(0)}(y_k(0))$, $n_k(1) = q^{(1)}\left(y_k^{(1)}\right)$. The optimality equations are similar to those derived in Eq. 9 (the states are simply extended from x to (x, y) and from m to (m, n), with extended control variables z and α). The shooting algorithm must then be parametrized by the initial controls, as described in this paper, but also by the initial position of the y variables, with a new objective function

$$E = \sum_{k=1}^{N}(m_k(1) - q^{(1)}(x_k(1)))^2 + \sum_{k=1}^{N}\left(n_k(1) - q^{(1)}(y_k^{(1)})\right)^2 + \sum_{k=1}^{N}|y_k(1) - y_k^{(1)}|^2.$$

This symmetrized discretization scheme can be addressed along the same lines as the one studies in the present paper.

Appendix: Forward and adjoint systems

We here provide more details on the implementation of the adjoint method described in Section 5. We assume, in the following, that K_V is a scalar multiple of the identity matrix, and taking variations of Eq. 9 in the discrete variables x_k, z_k, m_k yields a forward system of equations that evolves these variations (note, to keep the equations below compact, we do not write the explicit evaluation of K_V and K_H at x_k, x_ℓ):

$$\partial_t(\delta x_k) = \frac{1}{\sigma^2} \sum_{\ell=1}^{N} ((\nabla_1 K_V) \cdot \delta x_k z_\ell(t) + (\nabla_2 K_V) \cdot \delta x_\ell z_\ell(t) + K_V \delta z_\ell(t)),$$

$$\partial_t(\delta z_k) = -\frac{1}{\sigma^2} \sum_{\ell=1}^{N} \left(z_\ell(t) \cdot z_k(t) \left(D_{11}^2 K_V\right)^T \delta x_k + z_\ell(t) \cdot z_k(t) \left(D_{12}^2 K_V\right)^T \delta x_\ell \right.$$
$$\left. + \nabla_1 K_V z_k(t) \cdot \delta z_\ell(t) + \nabla_1 K_V z_\ell(t) \cdot \delta z_k(t)) \right.$$
$$\left. - \sum_{\ell=1}^{N} \left(\alpha_k \alpha_\ell \left(D_{11}^2 K_H\right)^T \delta x_k - \alpha_k \alpha_\ell (D_{12}^2 K_H)^T \delta x_\ell - \nabla_1 K_H \alpha_k \delta \alpha_\ell - \nabla_1 K_H \alpha_\ell \delta \alpha_k \right) \right),$$

$$\partial_t(\delta m_k) = \sum_{\ell=1}^{N} (\alpha_\ell(\nabla_1 K_H) \cdot \delta x_k + \alpha_\ell(\nabla_2 K_H) \cdot \delta x_\ell + K_H \delta \alpha_\ell),$$

where $\delta x_k(t)$ denotes a variation in the value of the position of the node x_k at time t (and analogously for $\delta z_k, \delta m_k$), and K_V, K_H are treated as functions on $\mathbb{R}^d \times \mathbb{R}^d$, and so the subscripts for the gradient and Jacobian denote differentiation with respect to the first and second variables $x_k, x_\ell \in \mathbb{R}^d$.

Let $\xi_x, \xi_z, \xi_m, \eta_\alpha$ denote dual forms to the variations $\delta x, \delta z, \delta m$, and η_α the associated variation in α, as introduced in Eq. 17, which expands as (again without writing the evaluation of the kernel terms, and combining the summations for compactness of notation):

$$(\partial_t \xi_x)_k = \sum_{\ell=1}^{N} \left\{ -\frac{1}{\sigma^2} \left(\nabla_1 K_V z_\ell(t) \cdot \xi_{x,k}(t) + \nabla_1 K_V z_k(t) \cdot \xi_{x,\ell}(t) + z_\ell(t) \cdot z_k(t) D_{11}^2 K_V \xi_{z,k}(t) \right. \right.$$
$$\left. + z_k(t) \cdot z_\ell(t) D_{21}^2 K_V \xi_{z,\ell}(t) \right) + \alpha_k \alpha_\ell D_{11}^2 K_H \xi_{z,k}(t) + \alpha_\ell \alpha_k D_{21}^2 K_H \xi_{z,\ell}(t)$$
$$\left. - \alpha_\ell \nabla_1 K_H \xi_{m,k}(t) - \alpha_k \nabla_1 K_H \xi_{m,\ell}(t) \right\},$$

$$(\partial_t \xi_z)_k = \frac{1}{\sigma^2} \sum_{\ell=1}^{N} \left\{ -K_V \xi_{x,\ell}(t) + z_\ell(t)(\nabla_1 K_V) \cdot \xi_{z,k}(t) + z_\ell(t)(\nabla_2 K_V) \cdot \xi_{z,\ell}(t) \right\}$$

$$(\partial_t \eta_\alpha)_k = \sum_{\ell=1}^{N} \left\{ \alpha_\ell(\nabla_1 K_H) \cdot \xi_{z,k}(t) + \alpha_\ell(\nabla_2 K_H) \cdot \xi_{z,\ell}(t) - K_H \xi_{m,\ell}(t) \right\}.$$

Note that since no other variables depend on m_k in the forward system, the dual variable ξ_m is constant in time, and so we do not display its evolution in the list above.

References

1. Abraham, R.H., Smale, S.: Lectures of Smale on differential topology. Columbia University, Department of Mathematics (1963)
2. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Academic, New York (2003)
3. Agrachev, A.A., Sachkov, Y.: Control Theory from the Geometric Viewpoint. Springer, Berlin (2004)
4. Ambrosio, L.: Lecture Notes on Optimal Transport Problems. Springer, Berlin (2003)
5. Arnold, V.: Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. Annales de l'institut Fourier **16**(1), 319–361 (1966)
6. Arnold, V.: Les méthodes mathématiques de la mécanique classique. Editions Mir (1976)
7. Aronszajn, N.: Theory of reproducing kernels. Transactions of the American Mathematical Society, pp. 337–404 (1950)
8. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)
9. Faisal Beg, M., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int J. Comp. Vis. **61**(2), 139–157 (2005)
10. Brenier, Y.: Optimal transport, convection, magnetic relaxation and generalized boussinesq equations. Journal of Nonlinear Science **19**(5), 547–570 (2009)
11. Brézis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, Berlin (2011)
12. Bruveris, M., Vialard, F.-X.: On completeness of groups of diffeomorphisms (2014). arXiv:1403.2089
13. Camion, V., Younes, L.: Geodesic interpolating splines. Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, pp. 513–527 (2001)
14. Cheng, H., Greengard, L., Rokhlin, V.: A fast adaptive multipole algorithm in three dimensions. J. Comput. Phys. **155**(2), 468–498 (1999)
15. Darve, E.: The fast multipole method: numerical implementation. J. Comput. Phys. **160**(1), 195–240 (2000)
16. Droske, M., Rumpf, M.: A variational approach to nonrigid morphological image registration. SIAM J. Appl. Math. **64**(2), 668–687 (2004)
17. Droske, M., Rumpf, M.: Multiscale joint segmentation and registration of image morphology. IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2181–2194 (2007)
18. Dupuis, P., Grenander, U., Miller, M.: Variational problems on flows of diffeomorphisms for image matching. Quarterly of Applied Math (1998)
19. Ebin, D.: The manifold of Riemannian metrics. In: Proceedings Symposium AMS, vol. 15, pp. 11–40 (1970)
20. Ebin, D.G., Marsden, J.: Groups of diffeomorphisms and the motion of an incompressible fluid. Annals of Mathematics, pp. 102–163 (1970)
21. Gumerov, N.A., Duraiswami, R.: Fast multipole methods on graphics processors. J. Comput. Phys. **227**(18), 8290–8313 (2008)
22. Haker, S., Zhu, L., Tannenbaum, A., Angenent, S.: Optimal mass transport for registration and warping. Int. J. Comput. Vis. **60**(3), 225–240 (2004)
23. Henderson, A.: ParaView Guide, A Parallel Visualization Application. Kitware Inc. (2007)
24. Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler–Poincaré equations and semidirect products with applications to continuum theories. Adv. Math. **137**(1), 1–81 (1998)
25. Holm, D.D., Trouvé, A., Younes, L.: The Euler-Poincare theory of metamorphosis. Quart. Appl. Math. **67**, 661–685 (2009)
26. Inci, H., Kappeler, T., Topalov, P.: On the regularity of the composition of diffeomorphisms, vol. 226. American Mathematical Soc. (2013)
27. Jain, A., Younes, L.: A kernel class allowing for fast computations in shape spaces induced by diffeomorphisms. J. Comput. Appl. Math. **245**, 162–181 (2013)
28. Jain, S., Tward, D.J., Lee, D.S., Kolasny, A., Brown, T., Tilak Ratnanather, J., Miller, M.I., Younes, L.: Computational anatomy gateway: leveraging xsede computational resources for shape analysis. In: Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment, p. 54. ACM (2014)
29. Jones, E., Oliphant, T., Peterson, P., et al.: SciPy: open source scientific tools for Python (2001)

30. Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., John Kress, W., Lopez, I.C., Soares, J.V.B.: Leafsnap: a computer vision system for automatic plant species identification. In: Computer Vision–ECCV 2012, pp. 502–516. Springer, Berlin (2012)
31. Lang, S.: Introduction to Differentiable Manifolds. Interscience, New York (1962)
32. Marsden, J.E.: Lectures on Geometric Mechanics. Cambridge University Press, Cambridge (1992)
33. Miller, M.I., Younes, L.: Group action, diffeomorphism and matching: a general framework. Int. J. Comp. Vis. **41**, 61–84 (2001)
34. Misiołek, G., Preston, S.C.: Fredholm properties of Riemannian exponential maps on diffeomorphism groups. Invent. Math. **179**(1), 191–227 (2010)
35. Palais, R.S.: Foundations of Global Non-linear Analysis. Benjamin, New York (1968)
36. Richardson, C.L., Younes, L.: Computing metamorphoses between discrete measures. Journal of Geometric Mechanics **5**, 131–150 (2013)
37. Trouvé, A., Younes, L.: Local geometry of deformable templates. SIAM J. Math. Anal. **37**(1), 17–59 (2005)
38. Vaillant, M., Miller, M.I., Trouv'e, A., Younes, L.: Statistics on diffeomorphisms via tangent space representations. Neuroimage **23**(S1), S161—S169 (2004)
39. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage **45**(1), S61—S72 (2009)
40. Villani, C.: Optimal Transport: Old and New, vol. 338. Springer Science Business Media (2008)
41. Vincent, T.L., Grantham, W.J.: Nonlinear and Optimal Control Systems. Wiley, New York (1999)
42. Wahba, G.: Spline models for observational data. SIAM (1990)
43. Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J.C., Csernansky, J.G., Miller, M.I.: Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type. IEEE Trans. Med. Imaging **26**, 462–470 (2007)
44. Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G.: Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type. NeuroImage **20**, 667–682 (2003)
45. Younes, L.: Shapes and Diffeomorphisms, volume 171 of Applied Mathematical Sciences. Springer, Berlin (2010)