

## JACOBI FIELDS IN GROUPS OF DIFFEOMORPHISMS AND APPLICATIONS

BY

LAURENT YOUNES

*Department of Applied Mathematics and Statistics and Center for Imaging Science, Johns Hopkins University, 3400 N. Charles St., Baltimore MD 21218*

**Abstract.** This paper presents a series of applications of the Jacobi evolution equations along geodesics in groups of diffeomorphisms. We describe, in particular, how they can be used to perform implementable gradient descent algorithms for image matching, in several situations, and illustrate this with 2D and 3D experiments. We also discuss parallel translation in the group, and its projection on shape manifolds, and focus in particular on an implementation of these equations using iterated Jacobi fields.

**1. Introduction.** This paper derives new algorithms for the analysis of image deformations generated by geodesics in groups of diffeomorphisms. The general framework is *large deformation diffeomorphisms*, developed in particular in [10, 27, 6, 4]. A primary application of this framework is the research of correspondences between two images. Matching in this context is parametrized by a diffeomorphism, say  $\varphi$ , and the optimal  $\varphi$  is given by the minimizer of an energy of the form [10, 27, 6] (letting  $\text{id}$  be the identity function)

$$E(\varphi) = D(\text{id}, \varphi)^2 + \frac{1}{\sigma^2} \|I_0 \circ \varphi^{-1} - I_1\|_2^2. \quad (1)$$

In this energy,  $I_0$  and  $I_1$  are the images being compared.  $I_0 \circ \varphi^{-1}$  is the image deformed by the diffeomorphism  $\varphi$ , which is therefore required to be close to  $I_1$ . The closeness is measured by the squared  $L^2$ -norm between functions, which we have denoted  $\|\cdot\|_2^2$ , and weighted by the parameter  $1/\sigma^2$ . All images and diffeomorphisms are assumed to be defined on an open subset  $\Omega$  of  $\mathbb{R}^d$ , and space integrals are taken over  $\Omega$ . The first term,  $D(\text{id}, \varphi)^2$ , is a squared distance between the identity transformation and  $\varphi$ . This distance is obtained by placing a Riemannian structure on the group of diffeomorphisms [26, 19, 17, 18, 24]. This will be summarized later in this section.

---

Received March 20, 2006.

2000 *Mathematics Subject Classification.* Primary 58E50; Secondary 53C22.

*Key words and phrases.* Groups of diffeomorphisms, Jacobi fields, image registration, shape analysis, deformable templates.

This work is partially supported by NSF DMS-0456253.

*E-mail address:* laurent.younes@jhu.edu

©2007 Brown University  
Reverts to public domain 28 years from publication

Such an approach has the big advantage of always providing diffeomorphic (non-ambiguous) correspondences, while conserving a notable degree of smoothness, even for very large deformations. The counterpart is an increase of the computational complexity compared to other “static” methods used in this context [5, 22, 23]. The framework is also somewhat more demanding mathematically. We now take a little time introducing it.

As introduced by [9], we build diffeomorphisms as flows associated to nonhomogeneous ordinary differential equations (ODE’s) of the form  $y_t = v(t, y)$ . (Here and in the rest of the paper derivatives and partial derivatives are indicated by subscripts:  $y_t = dy/dt$ .) The function  $v$  is a time-dependent vector field which becomes an auxiliary variable for the matching problem. It is assumed to vanish on the boundary of  $\Omega$  (or at infinity if  $\Omega$  is not bounded). It generates a time-dependent diffeomorphism (the associated flow), defined by

$$\varphi_t^v(t, x) = v(t, x), \quad \varphi^v(0, x) = x$$

(again,  $\varphi_t^v$  is the partial derivative of  $\varphi^v$  with respect to time). Diffeomorphisms generated in this way have been shown to form a group that we shall denote  $G_V$ . This leads us to formulate the *inexact matching problem* as the search for the minimum of the energy

$$\tilde{E}(v) = \int_0^1 \|v(t)\|_V^2 + \frac{1}{\sigma^2} \|I_0 \circ \varphi^v(1)^{-1} - I_1\|_2^2, \quad (2)$$

where  $v(t)$  and  $\varphi(t)$  are the functions  $x \mapsto v(t, x)$  and  $x \mapsto \varphi(t, x)$  and  $\|\cdot\|_V$  refers to a norm over vector fields that will be discussed later. This problem, formulated in [10, 27], has been numerically solved in [6], with an alternative procedure provided in [4].

This problem is equivalent to (1) if the minimum there is searched over  $\varphi$  in  $G_V$ . This is because

$$D(\varphi, \tilde{\varphi}) := \min \left\{ \sqrt{\int_0^1 \|v(t)\|_V^2 dt} : \tilde{\varphi} = \varphi^v(1) \circ \varphi \right\}$$

is a distance between diffeomorphisms, and  $\min_v \tilde{E}(v) = \min_\varphi E(\varphi)$  with

$$E(\varphi) = \min_{v: \varphi^v(1) = \varphi} \tilde{E}(v) = D(\text{id}, \varphi)^2 + \frac{1}{\sigma^2} \|I_0 \circ \varphi^{-1} - I_1\|_2^2.$$

The distance  $D$  is in fact a Riemannian metric on  $G_V$ . Metrics that are written in this form are such that the optimal  $v$  satisfies an evolution equation, which has been described in abstract form by Arnold [3, 2]. This equation also derives from an application of the Euler-Poincaré principle, as described in [12, 15], and has been called EPDiff. It coincides with the Euler equation for compressible fluids in the case when  $\|v(t)\|_V = \|v(t)\|_2$ , the  $L^2$ -norm. Another type of norm on  $V$  (called the  $H_\alpha^1$ -norm) relates to models of waves in shallow water and provides the Camassa-Holm equation [7]. A discussion of EPDiff in the particular case of template matching is provided in [18], and a parallel with the solitons emerging from the Camassa-Holm equation is discussed in [13]. For us, the interest in this equation is that it makes possible specifying the optimal  $v$  by the initial condition  $v(0)$ ; therefore, minimizing  $\tilde{E}$  can be rewritten in a problem involving  $v(0)$  only. However, because the dependency of  $\varphi^v$  in  $v(0)$  is quite complex, the algorithms provided in [6, 4] do not use this property and instead directly minimize  $\tilde{E}$  in terms of

the time-dependent vector field  $v(t)$ . One of the goals of this paper is to provide an alternative procedure addressing the problem uniquely in terms of  $v(0)$ .

The main ingredient for this purpose is the computation of the differential of  $\varphi^v$  with respect to the initial condition  $v(0)$ . That such a computation is relevant is clear from the dependency of  $\tilde{E}$  on  $\varphi^v(1)$ . These derivatives correspond to a well-known geometric object for the underlying Riemannian structure, the Jacobi fields [8, 14]. After introducing these fields, and in particular, the evolution equations they satisfy, we will be able to provide a new matching algorithm for images. We will also describe how Jacobi fields can be used to implement parallel translation in  $G_V$ .

In the following,  $\Omega$  will be an open subset of  $\mathbb{R}^d$ , assumed to be bounded for simplicity;  $V$  will be a Hilbert space of square integrable vector fields on  $\Omega$ . Elements of  $V$  will be assumed to have enough partial derivatives, and to vanish on  $\partial\Omega$ . We will denote the dot product on  $V$  by  $(v, w) \mapsto \langle v | w \rangle_V$  and the value of a linear form  $m \in V^*$  at  $v \in V$  by  $(m | v)$ . The duality operators  $K : V^* \rightarrow V$  and  $L = K^{-1} : V \rightarrow V^*$  are defined by  $(m | v) = \langle Km | v \rangle_V$  and  $\langle v | w \rangle_V = (Lv | w)$ . They are symmetric in the sense that  $(m | Kp) = (p | Km)$  and  $(Lv | w) = (Lw | v)$ .

If  $A$  is a linear operator from  $V$  to  $V$ , its dual,  $A^* : V^* \rightarrow V^*$ , is defined by  $(A^*m | v) = (m | Av)$ .

**2. Jacobi fields for diffeomorphisms.** The group of diffeomorphisms  $G_V$  is equipped with a right invariant metric coinciding with  $\langle | \rangle_V$  at the identity, as described in [2, 27, 12, 18], for example. A geodesic in this group (starting from the identity in the direction  $v(0)$ ) is given by the equation  $\varphi_t = v(t, \varphi)$ , where  $v$  satisfies the conservation of momentum: for all  $w \in V$ ,

$$(Lv(t), w) = (Lv(0), (d\varphi(t))^{-1}w \circ \varphi(t)).$$

This equation uniquely specifies  $Lv(t)$  as a linear form on  $V$ , given the initial momentum and the evolving diffeomorphism  $\varphi(t)$ .

For a diffeomorphism,  $\varphi$ , the transformation  $w \mapsto (d\varphi \circ \varphi^{-1})w \circ \varphi^{-1}$  is called the Adjoint of  $\varphi$  and denoted  $Ad_\varphi w$ . The conservation of momentum can therefore be written  $(Lv(t), w) = (Lv(0), Ad_{\varphi(t)^{-1}}w)$  or  $Lv(t) = (Ad_{\varphi(t)^{-1}})^*(Lv(0))$ . When  $m(0) = Lv(0)$  is a function (in which case  $(m(0), w)$  is simply given by  $\int_\Omega \langle m(0) | w \rangle_{\mathbb{R}^d} dx$ ), a change of variables provides the expression of  $(Ad_\psi)^*m(0)$ , which is

$$(Ad_\psi)^*m(0) = (d\psi)^T m(0) \circ \psi | \det d\psi|,$$

where  $\psi$  is a diffeomorphism (note that  $Lv(0)$  being a function is a rather particular case; in general,  $Lv(0)$  belongs to the dual,  $V^*$ , which contains functions, but also measures and more general distributions). Under some assumptions on the operator  $L$ , the system of equations

$$\begin{cases} \varphi_t = v(t, \varphi) \\ Lv(t) = (Ad_{\varphi(t)^{-1}})^*(Lv(0)) \end{cases}$$

has a unique solution over all times with the initial condition  $\varphi(0) = \text{id}$ , providing a time-dependent diffeomorphism that we will denote  $\varphi^{v(0)}(t)$ . A proof of this statement is provided in [24, Theorem 7] (which needs to be applied, using notation of [24], in

the particular case  $\sigma^2 = 0$ ). This is the exponential map on  $G_V$ , for its right-invariant Riemannian structure.

A Jacobi field (starting at 0) is a time-dependent vector field along this geodesic given by the differential at  $\varepsilon = 0$  of  $\varepsilon \mapsto \varphi^{v(0)+\varepsilon w(0)}(t)$ , for some  $w(0)$  in  $V$ . Our first task is to derive the equations that it satisfies, in a form that will be useful for our further computations.

2.1. *Evolution equation for Jacobi fields.* We consider a geodesic given by the initial condition  $v(0) \in V$  and denote  $\varphi(t) = \varphi^{v(0)}(t)$ ,  $\psi(t) = \varphi(t)^{-1}$ . Given  $w(0) \in V$ ,  $\delta\varphi$  and  $\delta\psi$  are the derivatives, with respect to  $\varepsilon$  and at  $\varepsilon = 0$ , of  $\varphi^{v(0)+\varepsilon w(0)}$  and of its inverse. We also introduce the vector fields  $\alpha$  and  $\beta$  defined by

$$\delta\varphi = \alpha \circ \varphi, \quad \delta\psi = \beta \circ \psi.$$

The Lie group bracket on diffeomorphisms is defined by  $ad_v \alpha = [v, \alpha] = dv \alpha - d\alpha v$ . The bracket is related to  $Ad$  by  $(Ad_{\varphi(\varepsilon)} w)_\varepsilon = ad_v Ad_{\varphi(0)} w$  with  $v$  given by  $\varphi_\varepsilon(0) = v \circ \varphi(0)$ .

We have the following result.

**THEOREM 1.** The time-dependent vector fields  $\alpha$  and  $\beta$  satisfy the equations

$$\alpha_t = K(Ad_\varphi^* Lw(0)) - K(ad_\alpha^* Ad_\psi^*(Lv(0))) + ad_v \alpha, \quad (3)$$

$$\beta_t = -Ad_\psi K Ad_\psi^*(Lv(0) + ad_\beta^*(Lv(0))) \quad (4)$$

with  $\alpha(0) = \beta(0) = 0$ .

Also, the variation of the momentum  $Lv$  is given by

$$\delta Lv = Ad_\psi^* Lw(0) - ad_\alpha^* Ad_\psi^*(Lv(0)) = Ad_\psi^*(Lv(0) + ad_\beta^*(Lv(0))). \quad (5)$$

Like the big adjoint,  $ad_\beta^* m$  can be computed, at least when  $m$  is a  $C^1$  function, and is given by

$$ad_\beta^* m = (d\beta^T m + dm\beta + m \operatorname{div} \beta).$$

*Proof.* This result is proved (in a different form) in [16]. We make only a formal proof by identifying the equations satisfied by  $\alpha$  and  $\beta$ . We therefore work under the assumptions that the derivative in  $\varepsilon$  exists.

The inverse diffeomorphism,  $\psi$ , is governed by the equation  $\psi_t = -d\psi v$ . Taking the variation in  $\varepsilon$  yields

$$\beta_t \circ \psi + d\beta \circ \psi(-d\psi v) = -d(\beta \circ \psi)v - d\psi(\delta v),$$

which yields

$$\beta_t = -d\psi \circ \varphi(\delta v) \circ \varphi = -Ad_\psi(\delta v). \quad (6)$$

The expression of  $\delta v$  in terms of  $\beta$  can be computed from the conservation equation, which yields

$$\begin{aligned} (\delta(Lv), w) &= (\delta(Lv(0)), Ad_\psi w) + (Lv_0, \delta(Ad_\psi)w) \\ &= (Lv(0), Ad_\psi w) - (Lv(0), ad_\beta Ad_\psi w) \end{aligned}$$

so that

$$\delta Lv = Ad_\psi^* Lw(0) - Ad_\psi^* ad_\beta^*(Lv(0)),$$

which is the second part of (5). Plugging this expression, with  $\delta v = K\delta(Lv)$ , into (6) yields equation (4).

To prove (3), it suffices to notice that, since  $\delta\psi = -d\psi\alpha$ , we have  $\alpha = -Ad_\varphi\beta$  so that  $\alpha_t = -ad_v\alpha - Ad_\varphi\beta_t$ . This directly yields (3), while the first part of (5) comes from  $ad_\alpha = -Ad_\varphi ad_\beta Ad_\psi$ , which yields  $ad_\alpha^* Ad_\psi^* = -Ad_\psi^* ad_\beta^*$ .  $\square$

**3. Application: gradient descent in the group.** We now consider, as a first application, the issue of minimizing with gradient descent a function that depends on a diffeomorphism  $\varphi$  in  $G_V$ . We will consider functions of the form  $D(\text{id}, \varphi)^2 + E(\varphi^{-1})$ , where  $D$  is the geodesic distance on  $G_V$ . When  $\varphi = \varphi^{v(0)}(1)$  for some  $v(0) \in V$ , we have  $D(\text{id}, \varphi)^2 = \|v(0)\|_V^2 = (Lv(0) | v(0))$ . Since this representation can always be achieved (geodesics exist between any pair of points in  $G_V$ , [26]), there is no loss of generality in working directly with  $v(0)$  and minimizing

$$J(v(0)) = (Lv(0) | v(0)) + E(\varphi(1)^{-1}).$$

This is the problem we consider now.

3.1. *Derivative of J.* With our previous notation, the derivative of  $J(v(0) + \varepsilon w(0))$  (in  $\varepsilon$  at  $\varepsilon = 0$ ) is given by

$$2(Lv(0) | w(0)) + (dE(\varphi(1)^{-1}) | \beta(1) \circ \varphi(1)^{-1}),$$

where  $\beta(1)$  is given by (4). We now assume that  $dE(\varphi(1)^{-1})$  can be expressed as a function, which will be the case in the applications we are discussing here. Making a change of variable in the second term yields

$$2(Lv(0) | w(0)) + (|\det d\varphi(1)|dE(\varphi(1)^{-1}) \circ \varphi(1) | \beta(1)). \tag{7}$$

Note that the assumption that  $dE$  is a function will not apply to situations like landmark, curve or surface matching, for which it will typically be a measure. While computation is still possible in such cases, the change of variable will take another form.

We now want to express (7) in a “gradient form”,  $(\nabla J(v(0)) | w(0))$ . This will provide a gradient descent algorithm for  $J$ , in the form

$$v^{(n+1)}(0) = v^{(n)}(0) - \gamma K \nabla J(v^{(n)}(0)).$$

This requires the computation of the dual of the operation  $w(0) \mapsto \beta(1)$ , which is addressed in the next section.

3.2. *Dual of the Jacobi evolution.* Since (4) is linear in  $w(0)$  and  $\beta$ , we introduce a time-dependent operator,  $Q(t)$ , such that

$$\beta(t) = Q(t)w(0). \tag{8}$$

We already know that  $Q(0) = 0$ , since (4) is initialized at  $\beta = 0$ . Similarly, we let

$$\alpha(t) = R(t)w(0). \tag{9}$$

We will prove the following theorem.

THEOREM 2. We have

$$Q^*(t)w = - \int_0^t LAd_{\psi(t-s)}KAd_{\psi(t-s)}^*H(s,t)ds \quad (10)$$

with

$$H_s(s,t) = ad_{Ad_{\psi(t-s)}KAd_{\psi(t-s)}^*}^*H(Lv(0)) \quad (11)$$

and  $H(0,t) = w$ . We also have

$$R(t)^*w = -Q(t)^*Ad_{\varphi}^*w. \quad (12)$$

*Proof.* Introduce the operators  $V(t)w = -Ad_{\psi(t)}KAd_{\psi(t)}^*Lw$  and

$$U(t)\beta = -(Ad_{\psi(t)}KAd_{\psi(t)}^*ad_{\beta}^*)(Lv(0)).$$

With this notation,  $Q$  is the solution of the linear equation  $Q_t = V + UQ$  with  $Q(0) = 0$ . Let  $W(s,t)$  be the solution of the homogeneous equation,  $W_t = UW$ , with the initial condition  $W(s,s) = \text{Id}$ . The solution of the general equation is then given by

$$Q(t) = \int_0^t W(s,t)V(s)ds$$

so that

$$Q^*(t) = \int_0^t V^*(s)W^*(s,t)ds.$$

From  $W(s,t)W(t,s) = \text{Id}$ , we have

$$\left(\frac{d}{ds}W(s,t)\right)W(t,s) + W(s,t)\left(\frac{d}{ds}W(t,s)\right) = \left(\frac{d}{ds}W(s,t)\right)W(t,s) + W(s,t)U(s)W(t,s) = 0$$

so that  $(d/ds)W(s,t) = -W(s,t)U(s)$  and, passing to the dual:

$$\frac{d}{ds}W(s,t)^* = -U(s)^*W(s,t)^*.$$

To summarize,  $Q^*(t)$  is given by

$$Q^*(t) = \int_0^t V^*(t-s)W^*(t-s,t)ds$$

with  $(d/ds)W(t-s,t)^* = U(t-s)^*W(t-s,t)^*$ . So, fixing  $w$  and letting  $H(s,t) = W^*(t-s,t)w$ , we have

$$Q^*(t)w = \int_0^t V^*(t-s)H(s,t)ds$$

with  $H_s(s,t) = U^*(t-s)H(s,t)$ ,  $H(0,t) = w$ . We now proceed to the computation of  $U^*$  and  $V^*$ .

Since  $V = -Ad_{\psi}KAd_{\psi}^*L$ , we have  $V^* = -LAd_{\psi}KAd_{\psi}^*$ . Write

$$U\beta = -(Ad_{\psi}KAd_{\psi}^*L)(Kad_{\beta}^*)(Lv(0)).$$

We first compute  $B^*$  for  $B\beta = K(ad_{\beta})^*(Lv(0))$ . We have

$$\begin{aligned} (m \mid K(ad_{\beta})^*(Lv(0))) &= ((ad_{\beta})^*(Lv(0)) \mid Km) \\ &= -(Lv(0) \mid ad_{Km}\beta) \\ &= -((ad_{Km})^*Lv(0) \mid \beta) \end{aligned}$$

so that  $B^*m = -(ad_{Km})^*Lv(0)$ . Finally,

$$U^*m = -B^*(LAd_\psi K Ad_\psi^* m) = ad_{Ad_\psi K Ad_\psi^* m}^*(Lv(0)).$$

Putting everything together yields equation (10). Equation (12) is a consequence of the identity  $\alpha = -Ad_\varphi\beta$ .  $\square$

3.3. *Image matching.* We illustrate this with an algorithm for image matching. Here, we assume that  $I_0$  and  $I_1$  are smooth images. We minimize:

$$(Lv(0), v(0)) + (1/\sigma^2)\|I_0 \circ \varphi(1)^{-1} - I_1\|_2^2.$$

In this case,  $|\det d\varphi|\nabla E(\varphi^{-1}) \circ \varphi = (2/\sigma^2)|\det d\varphi|(I_0 - I_1 \circ \varphi)\nabla I_0$ , and the gradient is given by

$$J(v(0)) = 2Lv(0) + (2/\sigma^2)Q^*(1)(|\det d\varphi|(I_0 - I_1 \circ \varphi)\nabla I_0). \quad (13)$$

One can also add the constraint that  $Lv(0) = Z\nabla I_0$  for some  $Z$  (normality constraint, see [18]) and write a gradient descent algorithm directly in  $Z$ . In this case, mapping  $Z \mapsto Z + \varepsilon z$ , the first variation is

$$2(Lv(0) | K(z\nabla I_0)) + (2/\sigma^2)(Q^*(1)(|\det d\varphi|(I_0 - I_1 \circ \varphi)\nabla I_0) | K(z\nabla I_0)),$$

which can be written as

$$2(z | \langle \nabla I_0 | v(0) + (2/\sigma^2)Q^*(1)(|\det d\varphi|(I_0 - I_1 \circ \varphi)\nabla I_0) \rangle).$$

This gives the gradient descent algorithm on  $Z$ :

$$Z(n+1) = Z(n) - \gamma \langle \nabla I_0 | v(0) + (2/\sigma^2)Q^*(1)(|\det d\varphi|(I_0 - I_1 \circ \varphi)\nabla I_0) \rangle. \quad (14)$$

Recall that  $\varphi = \varphi(1)$  itself is given by the evolution equations

$$\varphi_t = v(t) \circ \varphi, \text{ with } Lv(t) = Ad_{\psi(t)}^*(Z\nabla I). \quad (15)$$

3.4. *Experiments.* Before providing experimental results using this algorithm, we provide a few details on our implementation. The computation of the gradient requires solving equation (14), together with (15) and (10). The gradient descent equation (14) is combined with a line search procedure (Golden Section [21]). The vector field  $|\det d\varphi|(I_0 - I_1 \circ \varphi)\nabla I_0$  is approximated by finite differences.

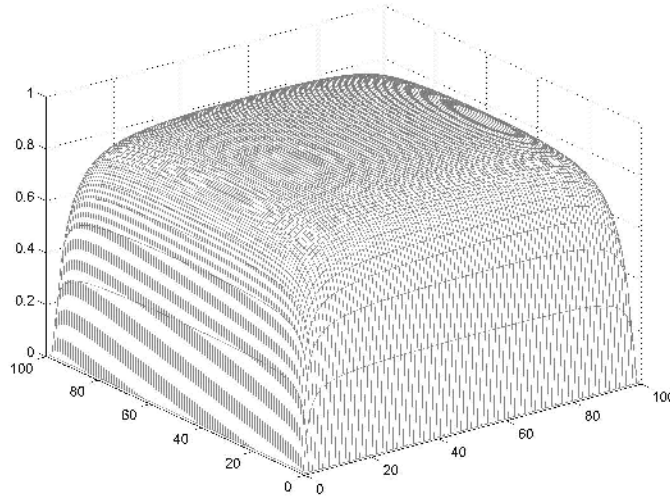
Solving (15) is done with a standard Euler scheme:

$$\varphi(t + \delta t) = \varphi(t) + \delta t v(t, \varphi(t))$$

with linear interpolation for the last term. We also solve simultaneously for  $\psi(t) = \varphi(t)^{-1}$  using  $\psi(t + \delta t) = \psi(t, \text{id} - \delta t v(t))$ . The time step  $\delta t$  is taken proportional to the norm of the initial momentum  $Lv(0) = Z\nabla I(0)$ , i.e., proportional to  $(Lv(0) | v(0))^{1/2}$ . The solution of (10) proceeds similarly, combining an Euler scheme with finite differences and linear interpolation.

An important issue concerns dealing with the kernel and its inverse. We use a censored Gaussian kernel with null boundary conditions on the image domain  $\Omega$ . This kernel is defined, in the continuum, by

$$K(x, y) = \mu(x)e^{-\|x-y\|^2/(2\sigma^2)}\mu(y),$$

FIG. 1. Function  $\mu$  used for censored Gaussian kernels

where  $\mu$  is a smooth function that is positive on  $\Omega$ , but vanishes in  $\partial\Omega$  and has vanishing derivatives there also. With rectangular images, we selected  $\mu$  to be the function (defined on  $[0, 1]^d$ )

$$\mu_0(x) = \exp\left(-\sum_{i=1}^d (x_i(1-x_i))^{-1}\right)$$

rescaled to the image domain. A discrete version of this is provided in Figure 1. Inversion of the kernel requires us to solve the deconvolution problem  $Kf = g$ , which can lead to unstable procedures. This is stabilized in our algorithm by solving  $Kf + \varepsilon f = g$ , for a small  $\varepsilon > 0$ , using conjugate gradient. Note that no kernel inversion is needed to solve (15), and only one is required in (10).

Figure 2 provides examples of non-rigid alignment obtained with a small database of shape images. Figure 3 provides results of the same algorithm applied to 3D data (from the Biomedical Informatics Research Network – BIRN).

*3.5. Remark: Newton algorithm.* This section and the following describe other possible versions of this approach and are invitations to future work. We first discuss a possible implementation of the matching algorithm using Newton’s algorithm.

One can show that the optimal momentum must satisfy the equation, proved in [6],

$$Lv(0) = \frac{1}{\sigma^2} |\det d\varphi| (I_0 - I_1 \circ \varphi) \nabla I_0. \quad (16)$$

This can be deduced from our computations here using the facts that  $J(v(0)) = 0$  in equation (13) and the identity  $Q^*(1)(Lv(0)) = -Lv(0)$ , which we justify for completeness.



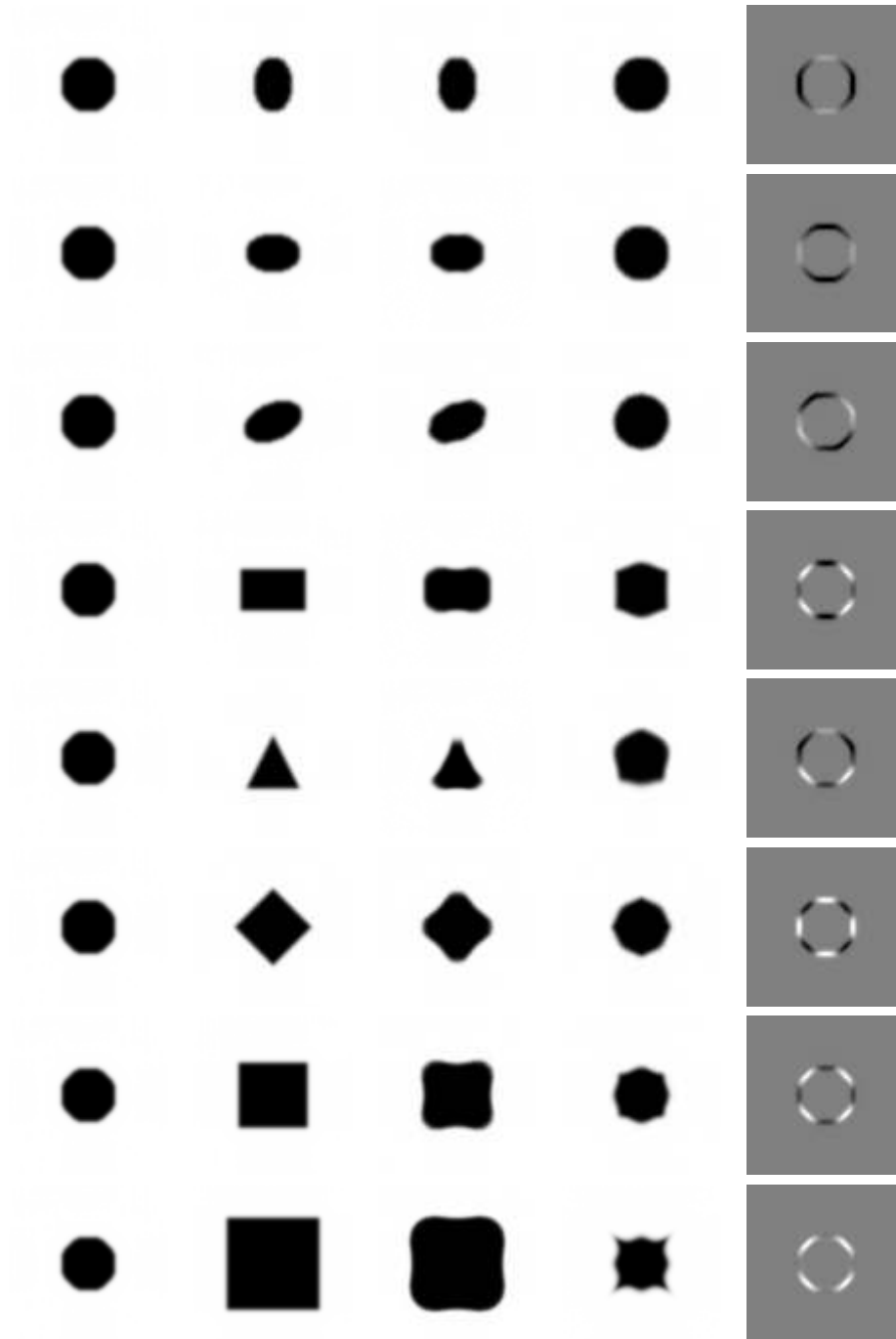


FIG. 2. First column: template image; second column: target image; third column: deformed template; fourth column: deformed target; fifth column: initial momentum ( $Lv(0)$ ). The initial momentum encodes the target shape along the boundary of the template.

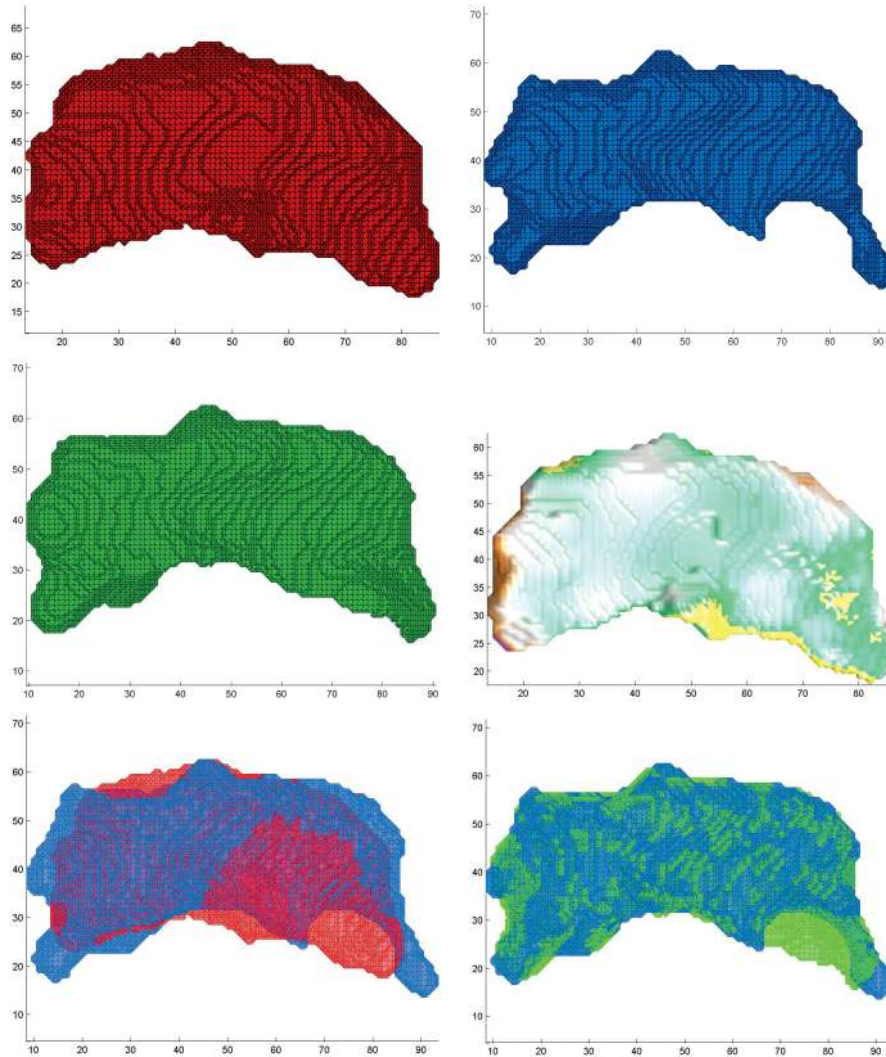


FIG. 3. Comparison of two 3D hippocampi shapes from the BIRN database. First row: Template, Target; second row: deformed Template and Momentum; third row: superimposed Template and Target, and deformed Template and Target. The color code for the Momentum is, from negative to positive: magenta, brown, white, green, yellow, white being 0.

This fact can be seen as a consequence of the identity, true on any Riemannian manifold [14]:

$$\langle v \mid w \rangle = \langle (d \exp(v))v \mid (d \exp(v))w \rangle.$$

With our notation, this gives  $(Lv(0) \mid w) = (LR(1)v(0) \mid R(1)w)$ . Using the fact that  $R(1)v(0) \circ \varphi(1) = (\partial/\partial \varepsilon) \varphi^{(1+\varepsilon)v(0)} = v(1) \circ \varphi(1)$ , we find the identity  $R(1)^*Lv(1) = Lv(0)$ .

Now, using (12), we have  $R(1)^*Lv(1) = -Q(1)^*Ad_{\varphi(1)}^*Lv(1) = -Q(1)^*Lv(0)$  by the conservation of momentum, which is the desired result.

Equation (16) can be seen as resulting from Pontryagin’s principle in optimal control. In finite dimensions, a standard approach is to find the control (here  $v(0)$ ) by directly solving this equation once a gradient descent has brought  $v(0)$  close enough to the solution (this is the shooting method). We now consider whether this is possible in our infinite-dimensional case, using, in particular, Newton’s method.

This method requires us to compute a small variation of (16). Taking  $v = v(0) + w(0)$ , a first-order expansion in  $w(0)$  is (with  $\alpha(1) = R(1)w(0)$  as in Theorem 1)

$$Lv(0) + Lw(0) = \frac{1}{\sigma^2} |\det d\varphi| (I_0 - I_1 \circ \varphi) \nabla I_0 + \frac{1}{\sigma^2} |\det d\varphi| ((I_0 - I_1 \circ \varphi) \operatorname{div}(\alpha(1)) \circ \varphi - \langle \nabla I_1 \circ \varphi | \alpha(1) \circ \varphi \rangle) \nabla I_0.$$

Newton’s method consists in solving this equation with respect to  $w(0)$ , then replacing  $v(0)$  by  $v(0) + w(0)$  and iterating. Obviously, the key problem is computing  $w(0)$  at each step. Letting

$$Aw(0) = w(0) - \frac{1}{\sigma^2} K(|\det d\varphi|(\operatorname{div}(\alpha(1)) \circ \varphi(I_0 - I_1 \circ \varphi) - \langle \nabla I_1 \circ \varphi | \alpha(1) \circ \varphi \rangle) \nabla I_0)$$

and  $b = -v(0) + \frac{1}{\sigma^2} K(|\det d\varphi|(I_0 - I_1 \circ \varphi) \nabla I_0)$ , we need to solve  $Aw(0) = b$ . Let  $A^T$  be define by  $\langle A^T w | w' \rangle_V = \langle w | Aw' \rangle_V$ , so that  $A^T = KA^*L$ . One possible approach is to solve  $A^T Aw(0) = A^T b$ , which can be done by conjugate gradient, as soon as  $A^T$  can be computed, which may be done as follows. First of all, we have

$$\begin{aligned} (m | K |\det d\varphi|(\operatorname{div}(\alpha(1)) \circ \varphi(I_0 - I_1 \circ \varphi) \nabla I_0) &= (|\det d\varphi|(\operatorname{div}(\alpha(1)) \circ \varphi(I_0 - I_1 \circ \varphi) \nabla I_0 | Km) \\ &= (\operatorname{div}(\alpha(1))(I_0 \circ \psi - I_1) \nabla I_0 \circ \varphi | Km \circ \psi) \\ &= -(\alpha(1) | \nabla \langle (I_0 \circ \psi - I_1) \nabla I_0 \circ \varphi | Km \circ \psi \rangle). \end{aligned}$$

Similarly,

$$\begin{aligned} (m | K(|\det d\varphi| \langle \nabla I_1 \circ \varphi | \alpha(1) \circ \varphi \rangle \nabla I_0)) &= (\langle \nabla I_1 | \alpha(1) \rangle \nabla I_0 \circ \psi | Km \circ \psi) \\ &= (\langle Km \circ \psi | \nabla I_0 \circ \psi \rangle \nabla I_1 | \alpha(1)). \end{aligned}$$

The difference of the two expressions is

$$-(\alpha(1) | \langle \nabla I_0 \circ \psi | Km \circ \psi \rangle \nabla(I_0 \circ \psi) + (I_0 \circ \psi - I_1) \nabla \langle \nabla I_0 \circ \varphi | Km \circ \psi \rangle)$$

This implies that

$$A^*m = m - R(1)^*(\langle \nabla I_0 \circ \psi | Km \circ \psi \rangle \nabla(I_0 \circ \psi) + (I_0 \circ \psi - I_1) \nabla \langle \nabla I_0 \circ \varphi | Km \circ \psi \rangle)$$

and  $A^T = KA^*L$ .

3.6. *Remark: Symmetrizing the alignment process.* The matching function in (1) is asymmetric in  $I_0$  and  $I_1$ : if  $\varphi$  is a minimizer of  $E$  in (1), then  $\varphi^{-1}$  does not necessarily minimize  $E$  with  $I_0$  and  $I_1$  interchanged. This asymmetry is due to the data attachment term, since

$$\|I_0 \circ \varphi^{-1} - I_1\|_2 \neq \|I_1 \circ \varphi - I_0\|_2;$$

the first term,  $D(\text{id}, \varphi)$ , is symmetrical ( $D(\text{id}, \varphi) = D(\text{id}, \varphi^{-1})$ ) because of the right invariance of the metric. This asymmetry can be justified by the initial interpretation of this energy in terms of deformable template theory. In this context  $I_0$  is the template, and  $D(\text{id}, \varphi)$  is the deformation cost involved in the transformation  $I_0 \rightarrow I_0 \circ \varphi^{-1}$ . The data attachment term then can be interpreted as observation noise, explaining the difference between the ideal image (the deformed template) and the observed one,  $I_1$ .

In situations when the asymmetry template/target is not natural, it may be preferable to symmetrize the procedure. This can easily be done by slightly modifying the data attachment term to be

$$U(I_0, I_1, \varphi) = \int |\det d\varphi \circ \varphi^{-1}|^{-1/2} (I_0 \circ \varphi^{-1} - I_1)^2 dx = \int |\det d\psi|^{1/2} (I_0 \circ \psi - I_1)^2 dx$$

with  $\psi = \varphi^{-1}$ . Indeed, with this choice, we have

$$\begin{aligned} U(I_0, I_1, \varphi) &= \int |\det d\psi|^{1/2} (I_0 \circ \psi - I_1)^2 dx \\ &= \int |\det d\psi|^{1/2} (I_0 \circ \psi - I_1)^2 dx \\ &= \int |\det d\psi \circ \varphi|^{1/2} |\det d\varphi| (I_0 - I_1 \circ \varphi)^2 dx \\ &= U(I_1, I_0, \psi). \end{aligned}$$

With the previous notation  $\partial\psi = \beta \circ \psi$ , we can compute the variation of  $U$  to be

$$\begin{aligned} \partial U &= \int |\det d\psi|^{1/2} (I_0 \circ \psi - I_1) \langle \nabla I_0 \circ \psi \mid \beta \circ \psi \rangle dx \\ &\quad + \frac{1}{2} \int |\det d\psi|^{1/2} (I_0 \circ \psi - I_1)^2 (\text{div} \beta) \circ \psi dx \\ &= \int |\det d\varphi|^{1/2} (I_0 - I_1 \circ \varphi) \langle \nabla I_0 \mid \beta \rangle dx + \frac{1}{2} \int |\det d\varphi|^{1/2} (I_0 - I_1 \circ \varphi)^2 \text{div} \beta dx \\ &= \int |\det d\varphi|^{1/2} (I_0 - I_1 \circ \varphi) \langle \nabla I_0 \mid \beta \rangle dx \\ &\quad - \frac{1}{2} \int \langle \nabla (|\det d\varphi|^{1/2} (I_0 - I_1 \circ \varphi)^2) \mid \beta \rangle dx. \end{aligned}$$

Therefore, the new expression of the energy gradient (13) is

$$\begin{aligned} J(v(0)) &= 2Lv(0) + (2/\sigma^2)Q^*(1)(|\det d\varphi|^{1/2}(I_0 - I_1 \circ \varphi)\nabla I_0) \\ &\quad - (1/\sigma^2)Q^*(1)\nabla(|\det d\varphi|^{1/2}(I_0 - I_1 \circ \varphi)^2). \quad (17) \end{aligned}$$

Another way to enforce symmetry, which is generalizable to other types of data attachment terms, is to place the comparison at the midpoint of the geodesic  $t \mapsto \varphi(t)$ .

Indeed, consider the geodesic evolution in  $G_V$  starting at  $\text{id}$  with velocity  $v(0)$ , and let  $w(0) = -v(1)$ , where  $v(1)$  is the final velocity of the geodesic. Then, one can prove that

$$\varphi^{w(0)}(t) \circ \varphi^{v(0)}(1) = \varphi^{v(0)}(1-t)$$

and the final velocity of the geodesic starting from  $\text{id}$  with initial velocity  $w(0)$  is  $-v(0)$  (this is just traveling along the same geodesic of  $G_V$  in the reverse direction). Because the velocities have constant norm along geodesics, we have

$$\begin{aligned} & \|v(0)\|_V^2 + \lambda \|I_0 \circ \psi^{v(0)}(1/2) - I_1 \circ \varphi^{v(0)}(1) \circ \psi^{v(0)}(1/2)\|_2^2 \\ &= \|w(0)\|_V^2 + \lambda \|I_1 \circ \psi^{w(0)}(1/2) - I_0 \circ \varphi^{w(0)}(1) \circ \psi^{w(0)}(1/2)\|_2^2. \end{aligned}$$

This argument shows the symmetry of the matching method with the data attachment term replaced by  $U = \|I_0 \circ \psi(1/2) - I_1 \circ \varphi(1) \circ \psi(1/2)\|_2^2$ .

In this case, the variation  $\delta U$  is (after computation)

$$\delta U = 2 \int (I_0 - I_1 \circ \varphi(1)) |\det d\varphi(1/2)| (\langle \nabla I_0 \mid \beta(1) \rangle - \langle \nabla(I_1 \circ \varphi(1)) \mid \beta(1/2) - \beta(1) \rangle) dx,$$

so that the gradient is

$$\begin{aligned} & 2Q(1)^*((I_0 - I_1 \circ \varphi(1)) |\det d\varphi(1/2)| (\nabla I_0 + \nabla(I_1 \circ \varphi(1)))) \\ & - 2Q(1/2)^*((I_0 - I_1 \circ \varphi(1)) |\det d\varphi(1/2)| \nabla(I_1 \circ \varphi(1))). \end{aligned}$$

*3.7. Remark: Image Metamorphoses.* The previous results can be formally applied to image metamorphoses ([19, 25]). Image metamorphoses can be described via the action of a larger group, which is the semidirect product of the group  $G_V$  of diffeomorphisms and the space  $L^2$ . This is a special case. In general, metamorphoses (as defined in [25]) do not reduce to semi-direct products.

We consider the space  $G_V \times L^2$ , which consists of all elements  $(\varphi, h) \in G_V \times L^2$ , equipped with the product

$$(\varphi, h)(\tilde{\varphi}, \tilde{h}) = (\varphi \circ \tilde{\varphi}, \tilde{h} \circ \varphi^{-1} + h).$$

$G_V \times L^2$  is equipped with the right-invariant metric associated to the dot product on  $V \times L^2$ :

$$\langle (v, \delta) \mid (\tilde{v}, \tilde{\delta}) \rangle = \langle v \mid \tilde{v} \rangle_V + \frac{1}{\sigma^2} \langle \delta \mid \tilde{\delta} \rangle_{L^2}.$$

This product can also be written as

$$\langle (v, \delta) \mid (\tilde{v}, \tilde{\delta}) \rangle = \left( \hat{L}(v, \delta) \mid (\tilde{v}, \tilde{\delta}) \right),$$

where  $\hat{L}(v, \delta) = (Lv, \delta/\sigma^2)$ . In particular,  $\hat{K} = \hat{L}^{-1}$  is given by  $\hat{K}(m, \delta) = (Km, \sigma^2\delta)$ . The associated metric on  $G_V \times L^2$  induces a canonical metric on images via the projection  $(\varphi, h) \mapsto h$ . We associate curves in  $G_V \times L^2$  to time-dependent pairs  $(v(t), \delta(t)) \in V \times L^2$  via the equations (which are a generalization of  $\varphi_t = v \circ \varphi$ )

$$\varphi_t = v \circ \varphi, \quad h_t = \delta - \langle \nabla h \mid v \rangle. \tag{18}$$

The geodesic and Jacobi equations are formally the same on  $G_V \times L^2$  as they were for  $G_V$ . To make them more explicit, we only have to compute the adjoint maps and their conjugates for this larger structure. First of all, write

$$\begin{aligned} (\varphi, h)(\tilde{\varphi}, \tilde{h})(\varphi, h)^{-1} &= (\varphi \circ \tilde{\varphi}, \tilde{h} \circ \varphi^{-1} + h)(\varphi^{-1}, -h \circ \varphi) \\ &= (\varphi \circ \tilde{\varphi} \circ \varphi^{-1}, -h \circ \varphi \circ \tilde{\varphi}^{-1} \circ \varphi^{-1} + \tilde{h} \circ \varphi^{-1} + h). \end{aligned}$$

The Adjoint is obtained by taking  $(\tilde{\varphi}, \tilde{h}) = (\text{id} + \varepsilon v, \varepsilon \delta)$  in the previous formula and taking the derivative with respect to  $\varepsilon$ , yielding (at  $\varepsilon = 0$ ):

$$Ad_{(\varphi, h)}(v, \delta) = (Ad_\varphi v, \delta \circ \varphi^{-1} + \langle \nabla h \mid Ad_\varphi v \rangle),$$

where  $Ad_\varphi$  refers to the Adjoint in the group of diffeomorphisms as before. For the dual, we have, for  $m \in V^*$  and  $\mu \in L^2$ :

$$\begin{aligned} ((m, \mu) \mid Ad_{(\varphi, h)}(v, \delta)) &= (m \mid Ad_\varphi v) + (\mu \mid \delta \circ \varphi^{-1} + \langle \nabla h \mid Ad_\varphi v \rangle) \\ &= (Ad_\varphi^*(m + \mu \nabla h) \mid v) + (|\det d\varphi| \mu \circ \varphi \mid \delta) \end{aligned}$$

yielding

$$Ad_{(\varphi, h)}^*(m, \mu) = (Ad_\varphi^*(m + \mu \nabla h), |\det d\varphi| \mu \circ \varphi).$$

In particular, the conservation of momentum that characterizes geodesics (specified by (18)) in  $G \times L^2$  is  $Ad_{(\varphi, h)}^*(Lv, \delta/\sigma^2) = \text{const}$  or  $(Lv, \delta/\sigma^2) = Ad_{(\psi, -h \circ \varphi)}(Lv(0), \delta(0)/\sigma^2)$ . This gives

$$\begin{cases} Lv = Ad_\psi^*(Lv(0) - \frac{1}{\sigma^2} \delta(0) \nabla(h \circ \varphi)), \\ \delta = |\det d\psi| \delta(0) \circ \psi. \end{cases}$$

In the first equation, we have

$$Ad_\psi^*(\delta(0) \nabla(h \circ \varphi)) = |\det d\psi| \delta(0) \circ \psi d\psi^T d\varphi^T \circ \varphi \nabla h = \delta \nabla h.$$

Therefore we have

$$\begin{cases} Lv = Ad_\varphi^* Lv(0) + \delta \nabla h, \\ \delta = |\det d\psi| \delta(0) \circ \psi. \end{cases} \quad (19)$$

The expression of  $ad_{(v, \delta)}(w, \eta)$  comes by differentiating  $Ad_{(\varphi, h)}(w, \eta)$  with respect to  $\varepsilon$ , where  $\varphi = \text{id} + \varepsilon v$  and  $h = \varepsilon \delta$ . This gives

$$ad_{(v, \delta)}(w, \eta) = (ad_v w, \langle \nabla h \mid w \rangle - \langle \nabla \eta \mid v \rangle).$$

The same operation on the dual yields

$$ad_{(v, \delta)}^*(m, \mu) = (ad_v^* m + \mu \nabla \delta, \langle \nabla \mu \mid v \rangle + \mu \text{div} v).$$

This provides, for example, the Jacobi evolution for a variation of the initial  $(v, \delta)$  by a perturbation in the direction of  $(w, \eta)$ . Let, for example,  $(\beta \circ \varphi_t, \rho + \langle \nabla h \circ \varphi \mid v \rangle)$  be the perturbation of  $(\varphi, h)^{-1}$ . We have, applying (4),

$$(\beta, \rho)_t = -Ad_{(\psi, -h \circ \varphi)} \hat{K} Ad_{(\psi, -h \circ \varphi)}^*((Lv(0), \eta/\sigma^2) + ad_{(\beta, \rho)}^*(Lv(0), \delta/\sigma^2)).$$

Other quantities, including  $Q^*(1)$ , adapt similarly. This opens the possibility of an implementation of metamorphoses using Jacobi fields. Denoting by  $(\varphi^{v(0), z(0)}, h^{v(0), z(0)})$  the solution of system (19), the problem consists in minimizing  $(Lv(0) \mid v(0) + \sigma^2 \|z(0)\|_2^2)$

under the constraint that  $h^{v(0),z(0)} = I_1 - I_0 \circ (\varphi^{v(0),z(0)})^{-1}$ . This can be done by minimizing

$$E(v, z) = \|I_1 - I_0 \circ (\varphi^{v,z})^{-1} - h^{v,z}\|_2^2$$

using gradient descent. One can furthermore use the fact that optimal solutions satisfy  $Lv(0) = \nabla I_0 z(0)$  and work only with the initial  $z$ .

It is also possible (maybe not feasible) to try to solve the equation  $I_1 - I_0 \circ (\varphi^{v,z})^{-1} - h^{v,z} = 0$  using Newton's algorithm, similarly to section 3.5.

Note that geodesic equations derived from the action of semi-direct products are related (with some important differences) to evolution equations derived in finite and infinite-dimensional mechanics [12].

**4. Application: parallel translation.** As a second application, we discuss an implementation of parallel translation in the group of diffeomorphisms and between images using Jacobi fields, with the following motivation.

There are situations when the objects of study are not shapes or images, but the relative positions of pairs (or  $m$ -tuples) of them. The usual point of view in computational anatomy [11] (and many other approaches in shape analysis) is template-based: before analysis, all the observations are placed in a "common coordinate system" by registering them to the template. This point of view is based on the expectation that objects with different shapes will end up having different representations relative to the template, as illustrated, for example, on our 2D simple shape database (Figure 2), where the target shape could be read from its momentum representation in the disc-centered coordinate system.

However, this representation is not adequate to easily address issues about the relative positions of the shapes. For example, one would be willing to decide whether a rectangle, relative to a square, is in some way in the same relation as an ellipse, relative to a disc. The same is true for real applications, such as studies of growth or other shape evolution.

A natural way to represent relative positions on a manifold is by using tangent vectors. They are of course first-order approximations of a variation, but can also represent large variations as initial velocities of geodesics, at least in the presence of a Riemannian structure. It is therefore natural to consider the issue of displacing a tangent vector at a given shape toward the tangent space at another shape. On Riemannian manifolds, this can be done using parallel translation.

In the following sections, we discuss parallel translation in the group of diffeomorphisms, then on shapes or image spaces, which are projections of the group obtained by the action on a template.

We (formally) identify  $V$  to the set of right-invariant vector fields on  $G_V$ , via the relation  $v \in V \leftrightarrow X^v : \varphi \mapsto v \cdot \varphi = v \circ \varphi$ . The bracket on  $V$  is such that  $[v, w]\varphi = -[X^v, X^w]_\varphi$ , where the last term is the Lie bracket between vector fields,  $[X, Y] = XY - YX$ .

#### 4.1. Parallel translation on $G$ .

4.1.1. *Covariant derivative.* Parallel translation along a curve  $\gamma$  being characterized by  $\nabla_{\dot{\gamma}_t} X = 0$ , we need the expression of the covariant derivative on  $G_V$ . If  $X$  is a vector

field on  $G_V$ , we let  $v_X$  be the function  $\varphi \mapsto X_\varphi \varphi^{-1} \in V$ . Therefore  $v_X$  is defined on  $G_V$  and takes values in  $V$ . We have the following proposition.

PROPOSITION 1.

$$(\nabla_Y X)_\varphi = (A_{v_Y} v_X + (Y v_X)) \varphi \quad (20)$$

with

$$A_v w = \frac{1}{2} \left( \text{ad}_v^T w + \text{ad}_w^T v - \text{ad}_v w \right). \quad (21)$$

*Proof.* This is proved in [1], [16], and derives from an application of the general formula for the Levi-Civita connection on a Riemannian manifold, namely,

$$2 \langle \nabla_Y X \mid Z \rangle = X \langle Y \mid Z \rangle_V + Y \langle X \mid Z \rangle_V - Z \langle X \mid Y \rangle_V + \langle [Y, X] \mid Z \rangle - \langle Y \mid [X, Z] \rangle - \langle X \mid [Y, Z] \rangle, \quad (22)$$

and on the identity (with notation from the proposition)  $v_{[X, Y]} = X v_Y - Y v_X + [v_X, v_Y]$ .  $\square$

4.1.2. *Parallel translation.* Let  $Y$  coincide with  $\dot{\gamma}$  along a curve  $\gamma$ . If  $v(t) = v_Y(\gamma(t))$  and  $w(t) = v_X(\gamma(t))$ , we obtain the equation for the parallel translation along  $\gamma$ :

$$\frac{dw}{dt} + A_v w = 0$$

or

$$\frac{dw}{dt} + \frac{1}{2} \left( \text{ad}_v^T w + \text{ad}_w^T v - \text{ad}_v w \right) = 0.$$

(In particular, for  $v = w$ , one retrieves the geodesic equation  $dv/dt + \text{ad}_v^T v = 0$ .)

The parallel translation in cotangent space can also be made explicit. Denote, as usual, by  $L$  the duality operator  $L : V \rightarrow V^*$  and let  $K = L^{-1}$ . Then, the covariant derivative of 1-forms is characterized by:

$$Y(M \mid X) = (\nabla_Y M \mid X) + (M \mid \nabla_Y X).$$

For  $\varphi \in G$ , define  $m(\varphi) \in V^*$  by  $(m(\varphi) \mid v) = (M \mid v_\varphi)$ : we will write  $M = m\varphi$ . Then one has  $(M \mid X) = (m \mid v_X)$ . This yields

$$(Ym \mid v_X) + (m \mid Yv_X) = (\nabla_Y M \mid X) + (m \mid Yv_X + A_{v_Y} v_X).$$

Therefore,

$$(\nabla_Y M \mid X) = (Ym \mid v_X) - (m \mid A_{v_Y} v_X).$$

We have  $(m \mid \text{ad}_v w) = (\text{ad}_v^* m \mid w)$ ,

$$(m \mid \text{ad}_w^T v) = \langle Km \mid \text{ad}_w^T v \rangle_V = \langle v \mid [w, Km] \rangle = -(\text{ad}_{Km}^* Lv \mid w)$$

and  $(m \mid \text{ad}_v^T w) = \langle w \mid \text{ad}_v(Km) \rangle$ . This implies that

$$2(m \mid A_{v_Y} v_X) = (-\text{ad}_{v_Y}^* m \mid v_X) - \langle \text{ad}_{Km}^*(Lv_Y) \mid v_X \rangle + \langle \text{ad}_{v_Y}(Km) \mid v_X \rangle,$$

which implies that

$$(\nabla_Y M)_g = \left( Ym + \frac{1}{2} (\text{ad}_{v_Y}^* m + \text{ad}_{Km}^*(Lv_Y) - L \text{ad}_{v_Y}(Km)) \right) g.$$



The equation for parallel translation along  $\gamma$  therefore is

$$\frac{dm}{dt} + \frac{1}{2} (ad_w^*p + ad_v^*m - Lad_v w) = 0 \tag{23}$$

with the notation  $\dot{\gamma} = v\gamma$ ,  $v = Kp$  and  $w = Km$ .

Again, for  $v = w$ , we retrieve the conservation equation

$$\frac{dm}{dt} + ad_v^*m = 0.$$

4.1.3. *Parallel translation along geodesics and Jacobi fields.* When  $\gamma$  is a geodesic on  $G$ , parallel translation along  $\gamma$  can be related to Jacobi fields. We indeed have the relation, in a Riemannian manifold,  $\Pi(t)$  holding for parallel transport along the geodesic,

$$\Pi(t)w(0) = J(t)/t + o(t).$$

Since  $\Pi(2t)w(0) = \Pi(t)\Pi(t)w(0)$ , we see that we can implement parallel translation by iterating the computation of normalized Jacobi fields over small periods of time. We describe this in more detail in the next section, first for a general manifold, then in the case of the group  $G$ .

4.2. *Discretization using Jacobi fields.*

4.2.1. *General algorithm.* For a point  $\varphi$  on a Riemannian manifold  $M$  and a tangent vector  $X$  at  $\varphi$ , denote by  $\gamma(t, \varphi, X)$  the geodesic starting at  $\varphi$  in the direction  $X$ . If  $Y$  is another vector tangent to  $M$  at  $\varphi$ , we want to compute the parallel translation of  $Y$  along this geodesic. We will denote this by  $P(t, Y)$ .

We consider the Jacobi field measuring the variation of the geodesic when the initial direction  $X$  is slightly perturbed by  $X \mapsto X + \varepsilon Y$ , namely,

$$J(t) = J(t, \varphi, X, Y) = \frac{d}{d\varepsilon} \gamma(t, \varphi, X + \varepsilon Y).$$

Using

$$P(t, Y) = \frac{1}{t} J(t, g, X, Y) + o(t),$$

we can use the following procedure to approximate  $P(1, Y)$  [2]: let  $\Delta = 1/N$  and  $Y_0 = Y$ ; then iterate

$$Y_{k+1} = \frac{1}{\Delta} J(\Delta, \varphi_k, X_k, Y_k)$$

until  $k = N$  ( $\varphi_k$  and  $X_k$  come from the discretization of the geodesic with step  $k$ ).

This algorithm is described in the appendix of [2]. One can enforce the conservation of norm and angles as follows. Let  $X_{k+1}$  be the velocity of the geodesic at time  $t = (k+1)/N$ . The value of  $Y_{k+1}$  previously obtained can be replaced by

$$\tilde{Y}_{k+1} = \alpha X_{k+1} + \beta Y_{k+1},$$

where  $\alpha$  and  $\beta$  are computed by solving the equations

$$\|\tilde{Y}_{k+1}\|_{\gamma_{k+1}}^2 = \|Y_{k+1}\|_{\gamma_{k+1}}^2 \quad \text{and} \quad \left\langle \tilde{Y}_{k+1} \mid X_{k+1} \right\rangle_{\gamma_{k+1}} = \left\langle Y_{k+1} \mid X_{k+1} \right\rangle_{\gamma_{k+1}}.$$

4.2.2. *Groups of diffeomorphisms.* We now adapt this algorithm to groups of diffeomorphisms, with our previously introduced notation. We consider a geodesic starting at the identity, and a vector  $w(0) \in V$ . The parallel translation along the geodesic starting in the direction  $v(0)$  can be computed as follows. Let  $\alpha(t) = R(t, v(0))w(0)$  be such that  $(d/d\varepsilon)\varphi^{v(0)+\varepsilon w(0)}(t) = \alpha(t) \circ \varphi^{v(0)}(t)$ . For a small  $\Delta = 1/N$ , iterate

$$w(k+1) = R(\Delta, v(k\Delta))w(k)/\Delta$$

until  $k = N$ . The parallel translation of  $w(0)$  is  $w(N) \circ \varphi(1)$ .

Conservation of the norm and of the dot product with  $v$  can be enforced as before, keeping in mind that the norm that is used is  $\|v\|_V^2 = (Lv \mid v)$ .

### 4.3. Parallel translation in an orbit.

4.3.1. *General formulation.* Objects of interest, such as images and shapes, are not diffeomorphisms, but are acted upon by diffeomorphisms. To address this, we consider an action  $(G, M) \rightarrow M$  and the induced metric. The action is assumed to be transitive (or restricted to an orbit), and we fix  $m_0$ , a reference element in  $M$  (the template).

We define the projection  $\pi : G \rightarrow M$  by  $\pi(\varphi) = \varphi.m_0$ . For  $v \in V$  and  $m \in M$ , we use the notation  $\rho_m(v) = v.m$  for the infinitesimal action on  $m$  ( $v.m \in T_m M$ ), and let  $V_m = \{v : v.m = 0\}$ .

For  $m \in M$ , we let  $G_m = \pi^{-1}(m)$  be the fiber over  $m$ . For  $\varphi \in G_m$ ,  $T_\varphi G_m$  is the vertical space over  $m$ : it is equal to the right translation of  $V_m$ , denoted  $V_m \varphi$ . The horizontal space is orthogonal to  $V_m \varphi$  in  $T_\varphi$  and equal to  $V_m^\perp \varphi$  by right invariance.

Note that  $\rho_m$  restricted to  $V_m^\perp$  is a bijection onto  $T_m M$ . We will denote  $v^\xi$  for the inverse image (lifting) of  $\xi \in T_m M$  under this map:  $v^\xi \in V_m^\perp$  and  $v^\xi m = \xi$ . The metric on  $M$  is defined by

$$\|\delta\|_m = \|v^\delta\|_V = \inf \{\|v\|_V : v.m = \delta\}. \quad (24)$$

A vector field  $X$  on  $G$  is horizontal if  $X_g \in V_{g.m_0}^\perp$  for all  $g$ . Given a vector field  $\xi$  on  $M$ , there is a unique ‘‘basic’’ horizontal lift to a vector field, denoted  $\hat{\xi}$ , on  $G$  defined by  $\hat{\xi}_g = v^{\xi m}.g$  for  $m = \pi(g)$ .

It can easily be proved (cf. [20] or directly from the formula for the covariant derivative) that, when  $\xi$  and  $\eta$  are vector fields on  $M$ ,

$$(\nabla_\eta \xi) = d\pi(\nabla_{\hat{\eta}} \hat{\xi}).$$

This is therefore the projection of the covariant derivative of the lifted vector fields.

Letting  $v_\xi(\varphi) = v^{\xi \pi(\varphi)}$  and similarly for  $v_\eta$ , we have

$$(\nabla_{\hat{\eta}} \hat{\xi})_\varphi = (\hat{\eta} v_\xi + A_{v_\eta} v_\xi)_\varphi.$$

Consider a curve  $t \mapsto m(t)$  on  $M$ : this can be lifted in a unique way as a horizontal curve  $\varphi(t)$  on  $G$  with

$$\varphi_t = v^{m_t} \varphi(t).$$

In particular,  $m_t = \varphi_t m_0 = v^{m_t} m$ . Parallel translation along  $m$  is therefore characterized by (using  $d\pi(\varphi)(v.\varphi) = v.m$ )

$$\left( v_t^\xi + A_{v^{m_t}} v^\xi \right) m = 0.$$

Since  $\xi = v^\xi.m$ , we have  $\xi_t = v_t^\xi.m + v^\xi(v^{m_t}m)$ . We therefore get the evolution equation

$$\xi_t + (A_{v^{m_t}}v^\xi)m = v^\xi(v^{m_t}m).$$

4.3.2. *Discretization using Jacobi fields.* When parallel translation is computed along a geodesic, we can again use Jacobi fields. Because the metric on  $M$  is the projection of the metric on  $G$ , the geodesic on  $M$  starting at  $m$  in the direction  $\xi$  simply is the projection of the geodesic on  $G$  starting at  $\text{id}$  in the direction  $v^\xi$ . In other terms, with our previous notation:

$$\gamma(t, m, \xi) = \varphi_{v^\xi}(t)m.$$

From this comes the fact that

$$J(t, m, \xi, \eta) = \frac{d}{d\varepsilon}\gamma(t, m, \xi + \varepsilon\eta) = (R(t, v^\xi).v^\eta).(\varphi.m).$$

In other terms, to compute the Jacobi field  $J(t, m, \xi, \eta)$  it suffices to lift  $\xi$  and  $\eta$  into  $v^\xi$  and  $v^\eta$ , then compute the corresponding Jacobi field  $w$  on  $G$ , and finally, to reproject it on  $M$ .

4.3.3. *Application to parallel transport with images.* Diffeomorphisms act on images by  $(\varphi, I) \rightarrow m \circ \varphi^{-1}$ . This implies that the infinitesimal action of a vector field on an image  $I$  is  $v.I = -\langle \nabla I \mid v \rangle$ . If  $\xi$  is a scalar field, its lift  $v^\xi$  is defined by the constrained minimization problem:

$$v^\xi = \operatorname{argmin}(\|v\|_V : \langle \nabla I \mid v \rangle = -\xi).$$

Introduce a space discretization by reducing the constraints to a finite grid constituted by points  $x_1, \dots, x_M$ . In this approximation, we therefore only require that, for all  $j$ :  $\xi(x_j) = -\langle \nabla_{x_j} I \mid v(x_j) \rangle$ . In such a case, we know that the solution takes the form

$$v(x) = \sum_{j=1}^M K(x, x_j)\alpha_j \nabla_{x_j} I$$

with  $\alpha_j \in \mathbb{R}$ . The coefficients  $\alpha_1, \dots, \alpha_M$  will be determined by the constraints, which yield (assuming that the kernel is scalar)

$$\xi(x_{j'}) = \sum_{j=1}^M K(x_{j'}, x_j) \langle \nabla_{x_{j'}} I \mid \nabla_{x_j} I \rangle \alpha_j, \quad j' = 1, \dots, M.$$

This is an  $M$ -dimensional linear system, which can be solved, for example, by conjugate gradient.

Another approach is possible when  $\xi$  is already given by  $-w.I$  for some vector field  $w$ . In this case, the constraint can be written  $\langle v - w \mid \nabla I \rangle = 0$ . Define the vector  $N_I$  to be  $\nabla I / |\nabla I|$  when  $\nabla I \neq 0$  and 0 if  $\nabla I = 0$ . The constraint can be rewritten  $\langle v - w \mid N_I \rangle = 0$ . Define the operator  $P_I$  by  $(P_I h)(x) = h(x) - \langle h(x) \mid N_I(x) \rangle N_I(x)$ . The problem we need to solve is equivalent to the minimization of  $|P_I h - w|_V^2$  with respect to  $h \in V$ . We have

$$\frac{1}{2}|P_I h - w|_V^2 = \frac{1}{2}(LP_I h \mid P_I h) - (Lw \mid P_I h) + |w|_V^2$$

so that the problem is (formally) equivalent to minimizing  $(1/2)(h \mid Ah) - (b \mid h)$  with  $A = P_I L P_I$  and  $b = P_I L w$  (using the fact that  $P_I$  is selfadjoint for the  $L^2$  inner product).

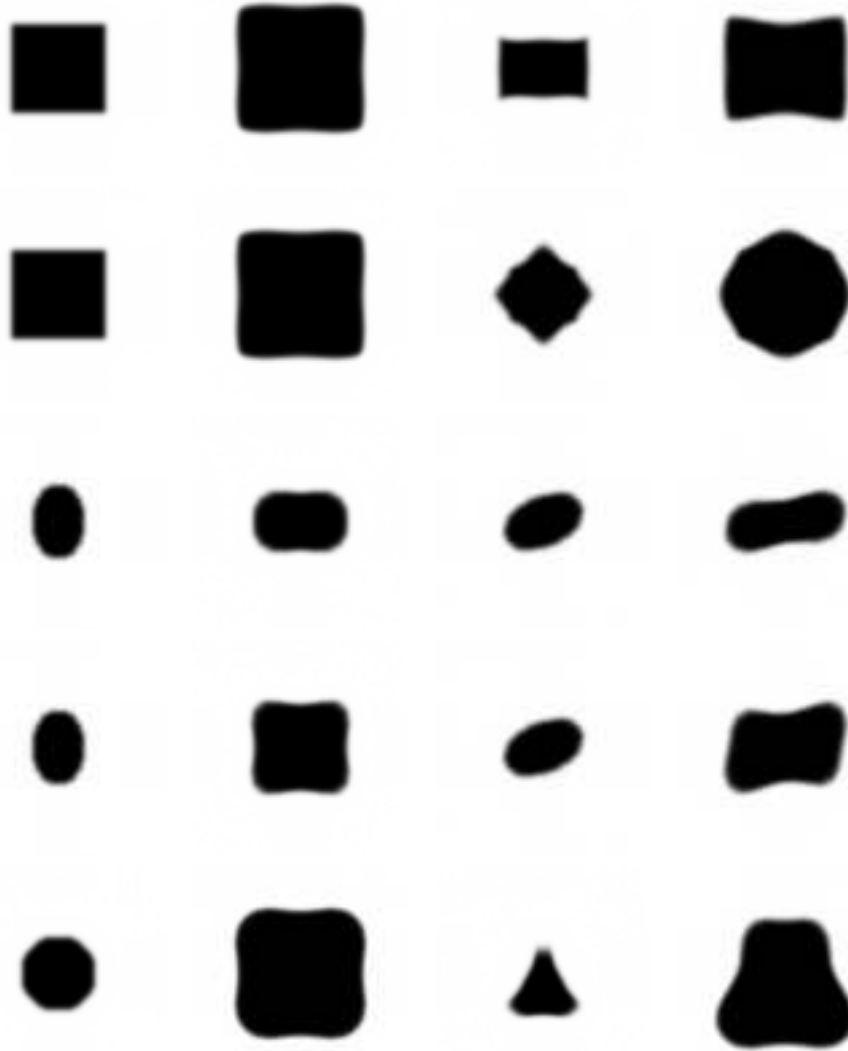


FIG. 4. Using parallel translation in the shape database. Image 4 is, relative to Image 2, in the same position as Image 3 relative to Image 1.

This can again be implemented by using conjugate gradient. This second approach seemed more robust in our experiments.

Figure 4 provides examples of transformations resulting from parallel translation using our basis of simple shapes. The experiment answers the following question: given three images,  $I_1, I_2, I_3$ , which image  $I_4$  is in the same situation relative to  $I_2$  as is  $I_3$  relative to  $I_1$ . This is done by computing the initial momentum of the geodesic between  $I_1$  and  $I_3$ ,

then parallel transporting it along the geodesic between  $I_1$  and  $I_2$ , and finally shooting from  $I_2$  with the obtained momentum.

## REFERENCES

- [1] V. I. Arnold, *Les méthodes mathématiques de la mécanique classique*, MIR, Moscow, 1976. MR0474391 (57:14033a)
- [2] ———, *Mathematical methods of classical mechanics*, Springer, 1978, Second Edition: 1989. MR0997295 (90c:58046)
- [3] V. I. Arnold, *Sur un principe variationnel pour les écoulements stationnaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaires*, J. Mécanique **5** (1966), 29–43.
- [4] B. Avants and J. Gee, *Shape averaging with diffeomorphic flows for atlas creation*, Proceedings of ISBI 2004, 2004.
- [5] R. Bajcsy and C. Broit, *Matching of deformed images*, The 6th international conference in pattern recognition, 1982, pp. 351–353.
- [6] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, Int. J. Comp. Vis. **61** (2005), no. 2, 139–157.
- [7] R. Camassa and D. D. Holm, *An integrable shallow water equation with peaked solitons*, Phys. Rev. Lett. **71** (1993), 1661–1664. MR1234453 (94f:35121)
- [8] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, 1992. MR1138207 (92i:53001)
- [9] E. G. Christensen, D. R. Rabbitt, and I. M. Miller, *Deformable templates using large deformation kinematics*, IEEE Trans. Image Proc. (1996).
- [10] P. Dupuis, U. Grenander, and I. M. Miller, *Variational problems on flows of diffeomorphisms for image matching*, Quarterly of Applied Math. **LVI** (1998) no. 3, 587–600. MR1632326 (99f:49006)
- [11] U. Grenander and I. M. Miller, *Computational anatomy: An emerging discipline*, Quarterly of Applied Mathematics **LVI** (1998), no. 4, 617–694. MR1668732 (2000e:92011)
- [12] D. D. Holm, J. E. Marsden, and T. S. Ratiu, *The Euler–Poincaré equations and semidirect products with applications to continuum theories*, Adv. in Math. **137** (1998), 1–81. MR1627802 (99e:58070)
- [13] D. D. Holm, T. J. Ratnanather, A. Trouvé, and L. Younes, *Soliton dynamics in computational anatomy*, Neuroimage **23** (2004), S170–S178.
- [14] S. Joshi, *Large deformation diffeomorphisms and Gaussian random fields for statistical characterization of brain sub-manifolds*, Ph.D. thesis, Sever Institute of Technology, Washington University, 1997.
- [15] J. E. Marsden, and T. S. Ratiu, *Introduction to mechanics and symmetry*, Springer, 1999. MR1723696 (2000i:70002)
- [16] P. Michor, *Some geometric equations arising as geodesic equations on groups of diffeomorphisms and spaces of plane curves including the hamiltonian approach*, Lecture Course, University of Wien.
- [17] I. M. Miller, A. Trouvé, and L. Younes, *On the metrics and Euler–Lagrange equations of computational anatomy*, Annual Review of Biomedical Engineering **4** (2002), 375–405.
- [18] ———, *Geodesic shooting for computational anatomy*, J. Math. Image and Vision (2005).
- [19] I. M. Miller and L. Younes, *Group action, diffeomorphism and matching: a general framework*, Int. J. Comp. Vis. **41** (2001), 61–84 (Originally published in electronic form in: *Proceeding of SCTV 99*, <http://www.cis.ohio-state.edu/~szhu/SCTV99.html>).
- [20] B. O’Neill, *The fundamental equations of a submersion*, Michigan Math. J. (1966), no. 13, 459–469. MR0200865 (34:751)
- [21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C, the art of scientific computing*, Cambridge University Press, second edition, 1992. MR1201159 (93i:65001b)
- [22] J.-P. Thirion, *Image matching as a diffusion process: An analogy with Maxwell’s demons*, Medical Image Analysis **2** (1998), no. 3, 243–260.
- [23] ———, *Diffusing models and applications*, Brain Warping (A. W. Toga, ed.), 1999, pp. 144–155.
- [24] A. Trouvé and L. Younes, *Local geometry of deformable templates*, SIAM J. Math. Anal. **37** (2005) no. 1, 17–59. MR2176922 (2006g:58010)
- [25] A. Trouvé and L. Younes, *Metamorphoses through Lie group action*, Found. Comp. Math. **5** (2005), 173–198. MR2149415 (2006a:68161)

- [26] Alain Trouvé, *Infinite dimensional group action and pattern recognition*, Tech. report, DMI, École Normale Supérieure, 1995. C. R. Acad. Sci. Paris Sér. I. Math. **321** (1995), no. 8, 1031–1034. MR1360567 (96j:58019)
- [27] ———, *Diffeomorphism groups and pattern matching in image analysis*, Int. J. of Comp. Vis. **28** (1998), no. 3, 213–221.