

## VARIATIONAL PROBLEMS ON FLOWS OF DIFEOMORPHISMS FOR IMAGE MATCHING

BY

PAUL DUPUIS (*LCDS, Division of Applied Mathematics, Brown University, Providence, RI*),

ULF GREANDER (*Division of Applied Mathematics, Brown University, Providence, RI*),

AND

MICHAEL I. MILLER (*Dept. of Electrical Engineering, Washington University, St. Louis, MO*)

**Abstract.** This paper studies a variational formulation of the image matching problem. We consider a scenario in which a canonical representative image  $T$  is to be carried via a smooth change of variable into an image that is intended to provide a good fit to the observed data. The images are all defined on an open bounded set  $G \subset \mathbf{R}^3$ . The changes of variable are determined as solutions of the nonlinear Eulerian transport equation

$$\frac{d\eta(s; x)}{ds} = v(\eta(s; x), s), \quad \eta(\tau; x) = x, \quad (0.1)$$

with the location  $\eta(0; x)$  in the canonical image carried to the location  $x$  in the deformed image. The variational problem then takes the form

$$\arg \min_v \left[ \|v\|^2 + \int_G |T \circ \eta(0; x) - D(x)|^2 dx \right], \quad (0.2)$$

where  $\|v\|$  is an appropriate norm on the velocity field  $v(\cdot, \cdot)$ , and the second term attempts to enforce fidelity to the data.

In this paper we derive conditions under which the variational problem described above is well posed. The key issue is the choice of the norm. Conditions are formulated under which the regularity of  $v(\cdot, \cdot)$  imposed by finiteness of the norm guarantees that the associated flow is supported on a space of diffeomorphisms. The problem (0.2) can

---

Received March 15, 1996.

1991 *Mathematics Subject Classification.* Primary 49J20, 49J27, 62G05.

The first author was supported in part by the National Science Foundation (NSF-DMS-9403820) and by the Army Research Office (DAAH04-93-G-0070).

The second author was supported by ONR N00014-91-J-1021, ARPA MDA972-93-1-0012, ARO DAAL03-92-G-0115, NSF BIR-9424264.

The third author was supported by the NIH grants RO1-MH525158-01A1 and NSF BIR-9424264, ARO-CIS DAAH04-95-1-0494.

be interpreted as a problem in optimal control, in which the superposition of the running cost  $\|v\|$  and the terminal cost determined by the data is to be minimized. We show that a minimizer  $\hat{v}$  exists, with the optimal smooth change of coordinates defined via (0.1). We also discuss an interpretation of the variational problem in the context of Bayesian estimation.

**1. Introduction.** We have been studying the variability of human and primate anatomy by generating smooth maps from a single template to a family of targets [1, 7, 11, 15]. In our approach the template represents the typical structure; variation is accommodated by the multitude of transformations that can be applied to the template. More precisely, if a function  $T(\cdot)$  defined on  $G$  is a canonical example of the structure of interest, then the family of possible targets are all functions of the form  $T(h(\cdot))$ , where  $h : G \rightarrow G$  is a smooth change of variable. By allowing all such maps a rich collection of possible targets is constructed. In particular, it is possible to accommodate the very finest details in variation of the individual anatomy through dilation, contraction, and warping of the underlying ideal coordinates of the template. If the change of coordinates that relates the template to a particular target is smooth then any detailed knowledge of the template can be exploited to quantitatively and qualitatively analyze corresponding geometric features of the target. For example, when the transformation is smooth connected sets remain connected, surfaces are mapped as surfaces, and the global relationships between structures are maintained.



FIG. 1. Showing two whole brain cryosection slices of the Macaque monkey brain taken from Van Essen and Coogan

Figure 1 shows the kinds of anatomical variation that motivate our mapping work. The left and right columns show sections through two macaques at roughly similar anatomical level. Notice the highly folded cortex, and the complex geometry of the deep nuclei and ventricle structures.

In the typical situation we are given data generated by an a priori unknown function. The data are usually a noise-corrupted vector-valued functional of the unknown. The problem is then to estimate the function. In a Bayesian approach to this problem, one would introduce a prior distribution on the family of possible targets, which is equivalent

to the mathematically simpler problem of imposing a prior distribution on the collection of smooth coordinate transformations. A Bayesian optimum is usually given by an extreme point of an energy potential composed of two terms, with the first term reflecting the “likelihood” of the coordinate transformation under the prior distribution, and with the second measuring the fidelity of the mapped template to the data. These two cost terms are minimized over all changes of variable  $h(\cdot)$  to determine an estimate  $T(h(\cdot))$  of the unknown target.

Such a Bayesian approach provides one motivation for the class of variational problems that will be studied in this paper. As we discuss at the end of this introduction, currently unresolved foundational issues prohibit the interpretation of our variational problem as a maximum a posteriori (MAP) solution for a Bayesian estimation problem (although a *minimum energy* interpretation in the sense of Mortensen’s deterministic estimator in nonlinear filtering [16, 12] seems possible). A longer term goal is an analysis of the MAP problem and its connection to the types of optimization problems considered in this paper. However, the variational problems are worthy of study even without the Bayesian interpretation, since they provide a method for systematically generating provably smooth maps that solve the “image registration” or “image matching” problem. Indeed, a number of approximate (i.e., suboptimal) computational methods based on the type of variational problem considered here have been implemented with some success, as for example in [5]. Further remarks on these approximate methods are given later in this introduction.

As noted previously, the object of our interest is the minimizing change of variable  $h$  in an as yet unspecified variational problem. Two issues immediately surface when a variational problem of this type is posed. (i) Are the functions that yield bounded cost legitimate changes of variable (i.e., are they homeomorphisms) and smooth? (ii) Can the variational problem accommodate large deformations while maintaining continuity and smooth deformation of the template? As demonstrated in this paper, it is possible to answer both questions in the affirmative if we define the coordinate transformation  $h$  indirectly as the output of a controlled flow. In this approach, which was first championed by Christensen and colleagues [6, 7], one considers the flow generated by a given velocity field in an Eulerian frame [13]. The variational cost is measured in terms of the velocity field, and consequently this function can be considered as a control. A very similar approach is taken in recent papers of Trouvé, such as [18]. The transformations are generated from a given velocity field  $v(\cdot, t) : G \rightarrow \mathbb{R}^3$  according to a partial differential equation that defines the dynamics of the controlled template:

$$h_t(x, t) = - [(\nabla_x h)^T(x, t)] v(x, t), \quad t \in [0, \tau], \tag{1.1}$$

where  $(\nabla_x f)^T(x)$  is the Jacobian matrix  $(\nabla_x f_1(x), \dots, \nabla_x f_3(x))^T$ . [A difficulty arises in defining a solution to (1.1) because we have not assumed (and for technical reasons do not wish to assume) regularity of  $v$  in  $t$ . This is avoided by replacing (1.1) by an alternative definition that coincides with (1.1) when  $v$  is smooth, and that is actually closer to the Eulerian description of the problem [cf. Section 2]. In this alternative formulation  $h(x, \tau)$  is defined to be the position of a particle that was initially at  $x$  and then subjected to the time-dependent forcing term  $v$ .]

Modeling issues from the Bayesian perspective and with regard to the selection of a prior distribution enter through the form of the dynamics, choice of boundary conditions, and selection of the cost for the velocity field. Particular examples whose penalty terms reflect the kinematics associated with elastic and fluid media are treated in [7, 11, 15].

Once a velocity field with sufficient regularity in  $x$  is fixed [cf. Section 2 for precise conditions], a candidate transformation for the variational problem is then given by  $h(\cdot, \tau)$ . Since we take  $v(\cdot, \cdot)$  as given and then construct  $h(\cdot, \tau)$ , the minimization should be done with respect to  $v$  rather than directly on  $h$ . Although this appears to complicate the problem, it seems to be necessary if one is to define a variational problem with the proper qualitative properties. In fact, it is because the penalty is applied to the *velocity* rather than to the deformation itself that large deformations are possible while preserving the diffeomorphic property. This should be compared to all quadratic penalties [3, 8, 15] based on small deformation assumptions [17] in which the restoring force to the template is proportional to the distance displaced, implying that large magnitude displacements are severely penalized. This inherent limitation was first observed in Christensen et al. [5, 4, 7, 6]. To deal with this, Christensen proposed using a sequence of local (in time) solutions of the global problem stated here. A large deformation flow of the form  $\{h(\cdot, t_i), i = 1, \dots, n\}$  was computed on the lattice of times  $\{t_i, i = 1, \dots, n\}$ , with  $\max_{i=1, \dots, n-1} (t_{i+1} - t_i)$  small. The mapping from  $h(\cdot, t_i) \rightarrow h(\cdot, t_{i+1})$  was obtained by integrating (1.1) with a constant velocity field  $v(\cdot, t_i)$  on the interval  $[t_i, t_{i+1}]$ . Each such constant velocity field  $v(\cdot, t_i)$  was generated as an approximate small time solution of the optimization problem stated below. In particular, the velocity field was chosen to minimize the sum of a cost that measured the distance between the deformed template at time  $t_{i+1}$  and the target data and the square of a norm on the velocity. The latter term was used to induce smoothness and took the form  $\|Lv\|^2$ , with the differential operator  $L$  chosen from fluid mechanics to be of the type  $L = -a\nabla^2 + b\nabla \cdot \nabla + cI$ .

The main focus of this paper is on the formulation of the global matching problem. An outline of the paper is as follows. In Section 2 we define the function  $h$  in terms of a flow generated by the velocity field  $v$ . We then state an integrability condition under which the velocity field generates diffeomorphisms and provide the proof. Section 3 starts with the formulation of a variational problem appropriate for the image registration problem for a given data model. We next apply the results of Section 2, and prove that under appropriate conditions a minimizing change of variable for this problem exists in the space of diffeomorphisms.

*Bayesian interpretation.* As remarked previously, one can formally obtain the class of variational problems considered in this paper as MAP estimators in a Bayesian framework. However, this connection is not rigorous and will require a careful development before a complete interpretation can be given. In this last part of the introduction we give a few remarks on the difficulties involved in establishing this connection.

A Bayesian model for the mapping problem involves a prior distribution and a data model. In general, the formulation of an appropriate data model is not too difficult. For an example, see Section 3. However, the determination of the prior distribution is much more involved. In the approach we use to model variation the different random "realizations" one might encounter in human and primate anatomy are all related to the

ideal template through a diffeomorphic change of variable. In other words, if the prior distribution is placed on the space of mappings from  $G \rightarrow G$  then we require that it be supported on the subspace of diffeomorphisms. In parallel with the introduction of a time variable in the formulation of the variational problem, it turns out to be quite useful in the construction of such a prior distribution to also add a time variable. In particular, a prior distribution that is guaranteed to take values in the set of diffeomorphisms on  $G$  can be obtained by using what Kunita [14] calls a *Brownian flow of diffeomorphisms*. These processes, which define stochastic flows on the space of diffeomorphisms, are substantially more involved than the flows generated by standard stochastic differential equations.

A standard method for assigning a precise MAP interpretation of a (conditional) distribution  $\mu$  defined on an infinite-dimensional space involves the calculation of the so-called Onsager-Machlup functional [9, 10]. In this approach there is a reference probability  $p_\epsilon$ , and given a point  $x$  in the range space one computes the limit  $a(x)$  of  $\mu\{y : \|y-x\| < \epsilon\}/p_\epsilon$  as  $\epsilon \rightarrow 0$ , where  $\|\cdot\|$  is a suitable norm. (In our case  $x$  would be a trajectory in the space of diffeomorphisms.) Because all calculations are done relative to the fixed sequence  $p_\epsilon$ ,  $-\log a(\cdot)$  can be taken as the quantity that should be minimized by the MAP estimator.

While this approach has intuitive appeal, it can be difficult to carry out in practice. The principal difficulty is the calculation of the required Onsager-Machlup functional. To the authors' knowledge, there are no existing results of this type for Brownian flows on diffeomorphisms, which are the natural class of processes for our applications. A less demanding approach is to consider a so-called *minimum energy* estimator along the lines of [12, 16], where it is developed in the context of nonlinear filtering for processes that take values in a Euclidean space. Although such an approach requires that large deviation estimates be available for Brownian flows on diffeomorphisms, it appears to be more feasible than the (more precise) approach based on the Onsager-Machlup functional.

**2. Velocity Fields that Generate Diffeomorphisms.**

2.1. *The Eulerian frame and ODE based flows.* We call the domain on which the template is defined  $G$ , and assume that  $G$  is an open bounded (not necessarily connected) subset of  $\mathbb{R}^3$ . The parameter  $\tau < \infty$  is fixed throughout. The results that we present below have analogues in  $\mathbb{R}^n$  for any  $n \in \{1, 2, \dots\}$ . However, the appropriate Sobolev norm depends on  $n$ . Since our principle application is to images in  $\mathbb{R}^3$ , results are presented for this case only.

This section establishes sufficient conditions for the vector fields  $v(\cdot, t) : G \rightarrow \mathbb{R}^3$  to generate diffeomorphisms. We first describe the space  $\mathcal{V}$  in which  $v(\cdot, t)$  will take values.  $\mathcal{V}$  is taken to be 3 products of the Sobolev space  $W_0^{3,2}(G)$ , which is the closure of  $C_0^\infty(G)$  with respect to the norm

$$\|f\|_{W_0^{3,2}(G)} = \left( \int_G \sum_{|\alpha| \leq 3} |D^\alpha f(x)|^2 dx \right)^{1/2} .$$

In the last equation  $\alpha$  stands for a multi-index. The norm on the product space  $\mathcal{V}$  will be denoted by  $\|\cdot\|_{[W_0^{3,2}(G)]^3}$ . A measurable function  $v : [0, \tau] \rightarrow \mathcal{V}$  will be called an *admissible velocity field* or an *admissible control* or *control process*.

At a ‘‘PDE’’ level, we can define the dynamics of the controlled template by

$$h_t(x, t) = - [(\nabla_x h)^T(x, t)] v(x, t), \tag{2.1}$$

where  $(\nabla_x f)^T(x)$  stands for  $(\nabla_x f_1(x), \dots, \nabla_x f_3(x))^T$ . This equation is to be solved together with the boundary conditions  $h(x, t) = x$  for  $x \in \partial G$  and  $t \in [0, \tau]$  and the initial condition  $h(x, 0) = x$  (corresponding to no deformation). Note that the dynamics are linear in  $h$  for a fixed  $v$ . Unfortunately, the lack of any regularity of  $v(\cdot, t)$  in the  $t$  variable makes the last equation somewhat awkward to work with.

Instead of using (2.1), we will adopt a trajectorial or particle description of the time evolution of the template. Accordingly, we introduce the ODEs

$$\dot{\eta}(s; t, x) = v(\eta(s; t, x), s), \quad \eta(t; t, x) = x, \tag{2.2}$$

where the dot stands for derivative with respect to  $s$ . The deformed template at time  $t$  is defined to be  $T(h(x, t))$ , where  $h(x, t) = \eta(0; t, x)$ . If  $h(x, \tau) = y$ , then a particle placed at  $y$  at time 0 and subject to the time and position dependent forcing term  $v$  ends at  $x$  at time  $\tau$ . Suppose we are only interested in deforming the template  $T(\cdot)$  to a prescribed configuration  $D(\cdot)$  at time  $\tau$ . Then the role of the vector field  $h$  will be to ‘‘carry’’ or ‘‘propagate’’ the template  $T$  to the data  $D$ :  $T(h(x, \tau)) = D(x)$ . The equality  $\eta(0; \tau, x) = h(x, \tau) = y$  tells us that the point  $y$  in the undeformed template at time 0 has been carried to the point  $x$  in the data at time  $\tau$ . With this description, although the template is being deformed in the forward time direction to match  $D(\cdot)$ , we are actually interested in finding the location  $y$  in the undeformed template that should be brought to  $x$ . Thus we take  $x$  as the given data, and solve (2.2) backward in time to locate the position  $y$  that should be mapped to  $x$ . The function  $T(h(\cdot, t))$  then evolves in a continuous fashion from the template  $T(\cdot)$  at time zero to  $T(h(\cdot, \tau)) \approx D(\cdot)$  at time  $\tau$ . See Figure 2.

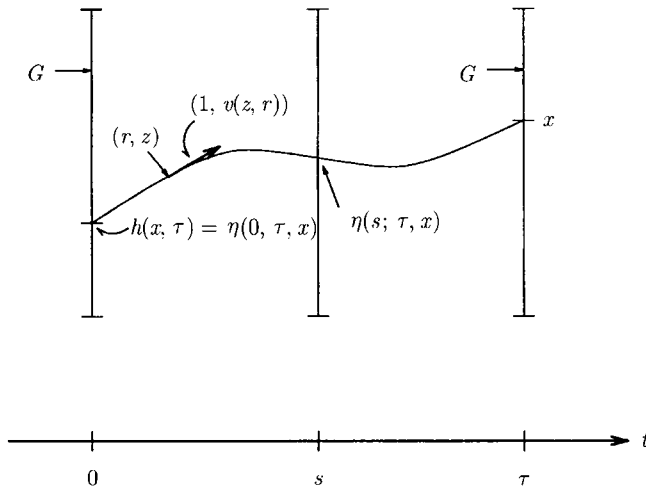


FIG. 2. Dynamics of the Flow

The definition of the dynamics just given is a “continuum mechanics” formulation of the problem in an Eulerian framework;  $\eta$  describes the path traced out by a particle under the forcing term  $v$  that ends at position  $x$  at time  $t$ , after having started at  $h(x, t)$  at time 0.

2.2. *The basic condition for diffeomorphisms.* Let  $L$  be a differential operator on  $[C_0^\infty(G)]^3$  that component-wise “dominates” the Sobolev norm we are using, in the sense that there is a  $c > 0$  such that if  $f \in [C_0^\infty(G)]^3$ , then

$$\|f\|_L^2 \doteq \sum_{i=1}^3 \int_G |(Lf)_i(x)|^2 dx \geq c \|f\|_{[W_0^{3,2}(G)]^3}^2. \tag{2.3}$$

In Section 3 we will need to consider limits of admissible controls and associated controlled templates. To guarantee the existence of subsequential limits, it will be necessary to verify a compactness condition. To simplify the exposition in Section 3 as much as possible, we will assume that  $\|f\|_L$  generates an inner product on  $[C_0^\infty(G)]^3$ , and that the Hilbert space  $H$  defined as the closure of  $[C_0^\infty(G)]^3$  with respect to this norm is separable. The main point of the comparison (2.3) is that  $\|f\|_L < \infty$  guarantees that each component of  $f$  will have the regularity properties of a function in  $W_0^{3,2}(G)$ .

We next state a condition on the velocity field that is sufficient for the field to generate a flow of diffeomorphisms. Since the condition works by imposing an integrability condition on Sobolev norms of the components of the velocity field, one may wonder why  $L$  is introduced at all. The reason is the role  $L$  plays in modeling issues. A natural approach to the selection of a prior distribution in the Bayesian setting is based on using principles from physics and continuum mechanics as a guide. For example, one could imagine that the prior distribution is obtained by driving a fluid or elastic media by a noise process. In such a case  $L$  is determined by the statistics of the noise and the constitutive laws of the media.

We next show that under the condition

$$\int_0^\tau \|v(\cdot, t)\|_L^2 dt < \infty, \tag{2.4}$$

the solution to (2.2) exists, is unique, and has nice differentiability properties with respect to  $x$ . Such a condition will imply the regularity of the mapping  $x \rightarrow h(x, \tau)$ . To do this, we first establish some basic properties of the ODE (2.2), under the condition (2.4).

LEMMA 2.1. There exists  $K < \infty$  and a modulus  $m$  such that for any  $f \in \mathcal{V}$ , any  $i = 1, 2, 3$ , any  $x$  and  $y$  in  $G$ , and any  $z \in \partial G$ ,

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq K \|f\|_L m(\|x - y\|),$$

$$\|f(x) - f(y)\| \leq K \|f\|_L \|x - y\|,$$

and

$$\nabla f_i(z) = f(z) = 0.$$

*Proof.* This follows from the existence of  $c > 0$  such that  $c\|f\|_{[W_0^{3,2}(G)]^3} \leq \|f\|_L^2$ , and a standard result on Sobolev functions. See, for example, Section 2.4 of [19]. These estimates form the basis for the results that follow, and any conditions that give analogous estimates will also work.

Let  $\bar{G}$  denote the closure of  $G$ .

**LEMMA 2.2.** Assume (2.4). Then for any  $t \in [0, \tau]$  and any  $x \in \bar{G}$ , a unique solution to (2.2) exists for  $s \in [0, \tau]$ .

*Proof.* Define  $v(x, t) = 0$  if  $x \notin G$ . If  $\|v(\cdot, t)\|_L < \infty$  for a given  $t$ , then the conclusion of Lemma 2.1 as applied to  $v(\cdot, t)$  is actually valid for all  $x, y$  with this extended definition of  $v$ . We will prove the conclusions of the lemma for any  $t \in [0, \tau]$  and  $s \in [0 \vee (t - \delta), \tau \wedge (t + \delta)]$ , where  $\delta > 0$  is independent of  $t$ . The lemma as stated follows by piecing together solutions over different time intervals. For simplicity, assume  $t \in [\delta, \tau - \delta]$ .

Consider any  $x \in \mathbb{R}^3$ . We claim that the mapping from

$$\{\phi \in C([t - \delta, t + \delta] : \mathbb{R}^3) : \phi(t) = x\}$$

into itself defined by

$$\Gamma(\phi)(s) = x + \int_t^s v(\phi(r), r) dr$$

is a contraction. Note that  $\Gamma$  is well defined by (2.4). By Lemma 2.1, for  $t - \delta \leq s \leq t + \delta$  and any  $(\phi_1, \phi_2) \in C([t - \delta, t + \delta] : \mathbb{R}^3)^2$

$$\begin{aligned} & \left\| \int_t^s [(v(\phi_1(r), r) - v(\phi_2(r), r))] dr \right\| \\ & \leq K \int_{t \wedge s}^{s \vee t} \|v(\cdot, r)\|_L \|\phi_1(r) - \phi_2(r)\| dr \\ & \leq K |s - t|^{\frac{1}{2}} \left[ \int_0^\tau \|v(\cdot, r)\|_L^2 dr \right]^{\frac{1}{2}} \|\phi_1 - \phi_2\|_{[t - \delta, t + \delta]}, \end{aligned}$$

where  $\|\phi_1 - \phi_2\|_{[t - \delta, t + \delta]} = \sup_{t \in [t - \delta, t + \delta]} \|\phi_1(t) - \phi_2(t)\|$ . Let

$$\delta^2 \leq \left( 2K^2 \left[ \int_0^\tau \|v(\cdot, r)\|_L^2 dr \right] \right)^{-1}.$$

Then  $\Gamma$  is a contraction on  $C([t - \delta, t + \delta] : \mathbb{R}^3)$ . By the contraction mapping theorem, a solution to

$$\eta(s; t, x) = \int_t^s v(\eta(r; t, x), r) dr + x$$

exists and is unique. Since this is equivalent to solving (2.2), the proof is complete.  $\square$

**LEMMA 2.3.** Assume (2.4). Then for any  $t \in [0, \tau]$ ,  $s \in [0, \tau]$  and  $x \in \bar{G}$ ,  $\eta(s; t, x)$  is continuously differentiable in  $x$  with the derivative given by the solution to

$$\dot{\Theta}(s; t, x) = \nabla_x v(\eta(s; t, x), s) \Theta(s; t, x), \quad \Theta(t; t, x) = I.$$



The proof is an adaptation of a standard argument [2]. We note that  $\nabla_x v(\cdot, \tau)$  is continuous by Lemma 2.1, and so the equation is not ambiguous.

LEMMA 2.4. Assume (2.4), and define  $\eta(s; t, x)$  by (2.2). Then for any  $t \in [0, \tau]$  and  $s \in [0, \tau]$  the mapping  $x \rightarrow \eta(t; s, x)$  is a 1-1 mapping from  $G$  onto itself, and moreover this mapping is a diffeomorphism.

*Proof.* This result is also elementary. Consider any  $x \in G$ . Since Lemma 2.1 states that  $v(x, t) = 0$  for  $t \in [0, \tau], x \in \partial G$ , we obviously have  $\eta(s; t, x) = x$  whenever  $x \in \partial G, t \in [0, \tau], s \in [0, \tau]$ . Because of the uniqueness proved in Lemma 2.2, solutions that start inside  $G$  cannot cross the boundary. Thus  $x \rightarrow \eta(t; s, x)$  maps  $G$  into itself. The uniqueness also implies that the mapping is 1-1, since if  $\eta(t; s, x_1) = \eta(t; s, x_2) = x$  for distinct  $x_1$  and  $x_2$  then there exist two different solutions that emanate from  $x$  at time  $t$ . To show that the mapping is onto, we use the existence guaranteed by Lemma 2.2. Given  $y \in G$  we can define  $x = \eta(t; s, y)$ , and simply note that  $y = \eta(s; t, x)$ . The diffeomorphism property now follows from Lemma 2.3.  $\square$

As noted above, if  $h(x, \tau)$  is associated to a controlled velocity field  $v$  that satisfies  $\int_0^\tau \|v(\cdot, t)\|_L^2 < \infty$  then  $h(\cdot, \tau)$  is automatically continuously differentiable, and furthermore  $h(x, \tau) = x$  if  $x \in \partial G$ . The density of the deformation at  $(x, t)$ , which is the inverse of the Jacobian, can be calculated as in Hunter [13, p. 30] to be

$$\begin{aligned} \rho(x, t) &= \exp \left[ - \int_0^t \sum_{i=1}^3 v_{i, x_i}(\eta(s; t, x), s) ds \right] \\ &\in (0, \infty). \end{aligned}$$

We can summarize the results above as follows.

THEOREM 2.5. Assume (2.4), and define the deformed template at time  $\tau$  by  $h(x, \tau) = \eta(0; \tau, x)$ , where  $\eta$  solves (2.2) with  $t = \tau$ . Then  $h(\cdot, \tau)$  is a diffeomorphism on  $G$ , with continuous derivatives given by  $\Theta(s; t, x)$  and Jacobian given by

$$\exp \left[ \int_0^\tau \sum_{i=1}^3 v_{i, x_i}(\eta(s; \tau, x), s) ds \right].$$

**3. Existence and Regularity Properties of the Solution to a Variational Problem.** We begin this section by formulating the variational problem that will be used in the image registration problem. As noted previously, this problem is motivated by the form of the MAP estimation problems, although such an interpretation has not yet been rigorously established.

We have previously introduced a dynamical model for the deformation as well as a cost of the form (2.4). The last part of the variational cost is the data term. Given a velocity field  $v$  and associated final coordinate transformation  $h(\cdot, \tau)$ , the term  $-\int_0^\tau \|v(\cdot, t)\|_L^2 dt$  formally represents the logarithm of the likelihood with respect to the prior distribution of this field. We will use this formal interpretation in the derivation of the variational problem.

The measured data  $\{D(i), i \in \mathcal{L}\}$  will be a Gaussian random field constructed as follows. Suppose that each component of the data represents a noisy measurement of a scalar (e.g., grey level) integrated over a subset of  $G$ . In particular, suppose that the data represent the output of CCD arrays with detectors centered at lattice sites  $\mathcal{L}$ , and with the  $i$ th detector measuring the integrated response from the voxel  $x_i \subset G$ , and with the full lattice tiling the entire volume:  $\bigcup_{i \in \mathcal{L}} x_i = G$ . We assume the high count limit and independence between distinct voxels. If  $h(\cdot, \tau)$  is the true change of variable that carries the template to the deformation, then the measurement  $D = \{D(i), i \in \mathcal{L}\}$  can be modeled by a Gaussian random vector of fixed isotropic variance and mean  $\int_{x_i} T(h(\sigma, \tau))d\sigma$  of the  $i$ th component.

The variational problem that corresponds to our formal Bayes posterior is now easy to derive. The density of  $D$ , given that the true change of variable is  $h(x, \tau)$ , is the product of

$$\exp\left(2 \sum_{i \in \mathcal{L}} D(i) \int_{x_i} T(h(\sigma, \tau))d\sigma - \sum_{i \in \mathcal{L}} \left| \int_{x_i} T(h(\sigma, \tau))d\sigma \right|^2\right)$$

and a function that depends only on  $D(i), i \in \mathcal{L}$ .

It is convenient to define data on the continuum by setting

$$D(x) = D(i), x \in x_i, i \in \mathcal{L} . \tag{3.1}$$

In this case we can write the variational problem in the convenient form

$$H(v) \doteq \int_0^\tau \|v(\cdot, t)\|_L^2 dt + \int_G |T(h(x, \tau)) - D(x)|^2 dx \tag{3.2}$$

where  $h(\cdot, \tau)$  is obtained from  $v$  through (2.2) and the definition  $h(x, \tau) = \eta(0; t, x)$ . We formally interpret  $-H(v)$  as the logarithm of a conditional likelihood.

In the rest of the section we will prove (under suitable conditions) that the minimizer in (3.2) exists, and moreover that any minimizer corresponds to a diffeomorphic deformation of the template.

The argument presented below will work in any dimension. However, the order of the Sobolev norm and the related restriction on the operator  $L$  will depend on the dimension. As in the last section, the details are only given for the case  $n = 3$ .

To establish the compactness conditions that will be needed to show the existence of a minimizer, we must work with a weaker topology than that induced by the norm in  $H$  (recall that  $H$  is the separable Hilbert space associated with the norm  $\|\cdot\|_L$ ). We begin by introducing the separable Hilbert space  $\mathcal{H} \doteq L^2([0, \tau] : H)$ , together with the norm  $\|w\|_{\mathcal{H}}^2 \doteq \int_0^\tau \|w(\cdot, t)\|_L^2 dt$  and the associated inner product.

We can consider this space with the weak topology, i.e.,  $v^n \rightarrow v$  if  $\langle v^n, w \rangle_{\mathcal{H}} \rightarrow \langle v, w \rangle_{\mathcal{H}}$  for all  $w \in \mathcal{H}$ . As usual, this topology is used because closed balls of finite radius are weakly compact. Hence if

$$\|v^n\|_{\mathcal{H}}^2 \leq M^2,$$

then for some subsequence (which we relabel by  $n$ ) there exists a weak limit  $v$ , and furthermore we have the desired lower semicontinuity of the running costs in the form

$$\liminf_{n \rightarrow \infty} \int_0^\tau \|v^n(\cdot, t)\|_L^2 dt = \liminf_{n \rightarrow \infty} \|v^n\|_{\mathcal{H}}^2 \geq \|v\|_{\mathcal{H}}^2 = \int_0^\tau \|v(\cdot, t)\|_L^2 dt.$$

We now consider such a weakly converging sequence, and examine the properties of the associated trajectories  $\eta^n(s; t, x)$ . Of course we are particularly interested in the case  $t = \tau$ , and want to show  $\eta^n \rightarrow \eta$ , where  $\eta$  solves

$$\dot{\eta}(s; \tau, x) = v(\eta(s; \tau, x), s), \quad \eta(\tau; \tau, x) = x. \tag{3.3}$$

Fix a point  $x \in G$ . We first show  $\{\eta^n(\cdot; \tau, x)\}$  is compact. Recall that  $G$  is bounded. If  $0 \leq s \leq t \leq \tau$  satisfy  $|s - t| \leq \delta$ , then by Lemma 2.1 there is a  $K < \infty$  such that

$$\begin{aligned} \|\eta^n(t; \tau, x) - \eta^n(s; \tau, x)\| &= \left\| \int_s^t v^n(\eta^n(r; \tau, x), r) dr \right\| \\ &\leq K \int_s^t \|v^n(\cdot, r)\|_L dr \\ &\leq K(t - s)^{\frac{1}{2}} \left[ \int_0^\tau \|v^n(\cdot, r)\|_L^2 dr \right]^{1/2} \\ &\leq (t - s)^{\frac{1}{2}} KM. \end{aligned}$$

Thus for the given  $x$ ,  $\{\eta^n(\cdot; \tau, x)\}$  is compact in  $C([0, \tau] : \mathbb{R}^3)$ . Consider a convergent subsequence, which we temporarily label by  $\phi^n$  with limit  $\phi$ . We can write

$$\begin{aligned} &\int_t^\tau [v^n(\phi^n(s), s) ds - v(\phi(s), s)] ds \tag{3.4} \\ &= \int_t^\tau [v^n(\phi^n(s), s) ds - v^n(\phi(s), s)] ds \\ &\quad + \int_t^\tau [v^n(\phi(s), s) ds - v(\phi(s), s)] ds. \end{aligned}$$

The first term can be estimated by using Lemma 2.1:

$$\begin{aligned} &\int_t^\tau [v^n(\phi^n(s), s) ds - v^n(\phi(s), s)] ds \\ &\leq (\tau - t)^{1/2} \left[ \int_t^\tau \|v^n(\phi^n(s), s) - v^n(\phi(s), s)\|_L^2 ds \right]^{1/2} \\ &\leq \tau^{1/2} K \left[ \int_t^\tau \|v^n(\cdot, s)\|_L^2 \|\phi^n(s) - \phi(s)\|^2 ds \right]^{1/2} \\ &\leq \tau^{1/2} K \|\phi^n - \phi\|_{C([0, \tau] : \mathbb{R}^3)} \|v^n\|_{\mathcal{H}} \\ &\rightarrow 0. \end{aligned}$$

For the second, we have to use the weak convergence of  $v^n$  to  $v$ . This implies  $w^n \doteq v^n - v$  tends weakly to 0, with  $\|w^n\|_{\mathcal{H}} \leq 2M$  for all  $n$ . For  $x \in G$  and  $t \in [0, \tau]$  define the function

$$z^n(x, t) \doteq \int_t^\tau w^n(x, s) ds.$$

If  $x, y \in G$  and  $0 \leq s \leq t \leq \tau$ , then Lemma 2.1 implies

$$\begin{aligned} \|z^n(x, t) - z^n(y, s)\| &\leq \left\| \int_t^\tau [w^n(x, r) - w^n(y, r)] dr \right\| + \left\| \int_s^t w^n(y, r) dr \right\| \\ &\leq \tau^{1/2} K \|w^n\|_{\mathcal{H}} \|x - y\| + K(t - s)^{1/2} \|w^n\|_{\mathcal{H}}. \end{aligned}$$

Because  $z^n(x, \tau) = 0$  for all  $x \in G$ , we conclude that  $\{z^n\}$  is compact in  $C(G \times [0, \tau])^3$ . Since  $w^n$  tends weakly to zero, an argument by contradiction shows  $z^n \rightarrow 0$  in  $C(G \times [0, \tau])^3$ .

We thus conclude that

$$\int_t^\tau [v^n(x, s) - v(x, s)] ds \rightarrow 0$$

uniformly in  $x \in G$  and  $t \in [0, \tau]$ . If we were considering the particular case  $\phi(s) = x$  for all  $s \in [0, \tau]$  and some fixed  $x \in G$ , it would follow that the second term in (3.4) tends to zero. To deal with the general case, we can approximate the continuous function  $\phi$  by a piecewise constant function. The corresponding convergence will hold for the piecewise constant function, and the error introduced by this approximation can be made arbitrarily small by using the same argument used to estimate the first term in (3.4).

We conclude that for all  $t \in [0, \tau]$ ,

$$\int_t^\tau [v^n(\phi^n(s), s) - v(\phi(s), s)] ds \rightarrow 0,$$

and therefore for all such  $t$

$$\phi(t) = \lim_{n \rightarrow \infty} \phi^n(t) = \lim_{n \rightarrow \infty} \int_\tau^t v^n(\phi^n(s), s) ds + x = \int_\tau^t v(\phi(s), s) ds + x.$$

By the uniqueness proved in Lemma 2.2 and an argument by contradiction, we conclude that  $\eta^n(\cdot; \tau, x)$  converges uniformly on  $[0, \tau]$  to the unique solution to (3.3).

Before the discussion of lower semicontinuity and compactness can be completed the terminal cost must also be considered. We continue to work with the sequence along which  $v^n$  is converging weakly. Assume that  $T(\cdot)$  and  $D(\cdot)$  are uniformly bounded measurable functions on  $G$ , and that the set  $\Delta_T \doteq \{x : T(\cdot) \text{ is discontinuous at } x\}$  is a set of Lebesgue measure zero. Since  $h^n(x, \tau) = \eta^n(0; \tau, x) \rightarrow \eta(0; \tau, x) = h(x, \tau)$ , we have  $T(h^n(x, \tau)) \rightarrow T(h(x, \tau))$  for all  $x \notin h^{-1}(\Delta_T, \tau)$ . However, since  $h$  is a diffeomorphism this set is also a set of Lebesgue measure zero, and so by the Lebesgue Dominated Convergence Theorem

$$\int_G |T(h^n(x, \tau)) - D(x)|^2 dx \rightarrow \int_G |T(h(x, \tau)) - D(x)|^2 dx.$$

We can now state the main result, which is a consequence of the preceding discussion, Lemma 2.2, and Theorem 2.5.

**THEOREM 3.1.** Assume that  $T(\cdot)$  and  $D(\cdot)$  are uniformly bounded measurable functions on  $G$ , and that the set  $\Delta_T \doteq \{x : T(\cdot) \text{ is discontinuous at } x\}$  is a set of Lebesgue measure zero. Assume also that the operator  $L$  satisfies the condition (2.3). Then the following conclusions hold.

- (1) Let  $\{v^n\} \subset \mathcal{H}$  be any sequence of controls with uniformly bounded costs:

$$\|v^n\|_{\mathcal{H}} = \int_0^\tau \|v^n(\cdot, t)\|_L^2 dt \leq M < \infty.$$

Then  $\{v^n\}$  is weakly compact in  $\mathcal{H}$ . If  $n$  denotes a weakly convergent subsequence with limit  $v$ , then the running costs satisfy the lower semicontinuity

$$\liminf_{n \rightarrow \infty} \int_0^\tau \|v^n(\cdot, t)\|_L^2 dt \geq \int_0^\tau \|v(\cdot, t)\|_L^2 dt.$$

- (2) Suppose that  $\{v^n\}$  converges weakly to  $v$ , and consider the trajectories associated with these controls:

$$\dot{\eta}^n(t; \tau, x) = v^n(\eta^n(t; \tau, x), t), \quad \eta^n(\tau; \tau, x) = x,$$

and

$$\dot{\eta}(t; \tau, x) = v(\eta(t; \tau, x), t), \quad \eta(\tau; \tau, x) = x.$$

Then each of these equations has a unique solution for all  $x \in G$ , and furthermore for all  $x \in G$ ,  $\eta^n(\cdot; \tau, x) \rightarrow \eta(\cdot; \tau, x)$  uniformly on  $[0, \tau]$ . In particular, we have

$$h^n(x, \tau) = \eta^n(0; \tau, x) \rightarrow \eta(0; \tau, x) = h(x, \tau),$$

and also the convergence of the terminal costs:

$$\int_G |T(h^n(x, \tau)) - D(x)|^2 dx \rightarrow \int_G |T(h(x, \tau)) - D(x)|^2 dx.$$

- (3) A minimizing control  $v^*$  for the variational functional (3.2) exists, and it defines a deformation  $h^*(\cdot, \tau)$  that is a diffeomorphism on  $G$ , whose derivatives are as given in Theorem 2.5.

### REFERENCES

- [1] Y. Amit, U. Grenander, and M. Piccioni, *Structural image restoration through deformable templates*. J. American Statistical Association **86** (414), 376–387 (June 1991)
- [2] V. I. Arnold, *Ordinary Differential Equations*, MIT Press, Cambridge, Massachusetts, 1973
- [3] R. Bajcsy and S. Kovacic, *Multiresolution Elastic Matching*, Computer Vision, Graphics, and Image Processing **46**, pp. 1–21, 1989
- [4] G. Christensen, *Deformable Shape Models for Anatomy*, Ph.D. Dissertation, Department of Electrical Engineering, Sever Institute of Technology, Washington University, St. Louis, MO, August 1994
- [5] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, *Deformable templates using large deformation kinematics*, IEEE Transactions on Image Processing, **5**(10), 1996, pp. 1435–1447.
- [6] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, *A deformable neuroanatomy textbook based on viscous fluid mechanics*, In Jerry Prince and Thordur Runolfsson, editors, *Proceedings of the Twenty-Seventh Annual Conference on Information Sciences and Systems*, pp. 211–216, Baltimore, Maryland, March 24–26, 1993. Department of Electrical Engineering, The Johns Hopkins University.
- [7] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, *3D brain mapping using a deformable neuroanatomy*, Physics in Medicine and Biology **39**, 609–618 (1994)
- [8] R. Dann, J. Hoford, S. Kovacic, M. Reivich, and R. Bajcsy, *Evaluation of Elastic Matching Systems for Anatomic (CT, MR) and Functional (PET) Cerebral Images*, Journal of Computer Assisted Tomography **13**(4), 603–611 (July/August 1989)
- [9] O. Zeitouni and A. Dembo, *A maximum a-posteriori estimator for trajectories of diffusion processes*, Stochastics **20**, 211–246 (1987) Erratum, p. 341
- [10] O. Zeitouni and A. Dembo, *An existence theorem and some properties of maximum a-posteriori estimation of trajectories of diffusion*, Stochastics **22**, 197–218 (1988)

- [11] U. Grenander and M. I. Miller, *Representations of knowledge in complex systems*, Journal of the Royal Statistical Society B **56**(3), 549–603 (1994)
- [12] O. Hijab, *Minimum Energy Estimation*, Ph.D. Dissertation, University of California, Berkeley, 1980
- [13] S. C. Hunter, *Mechanics of Continuous Media*, Ellis Horwood Limited, Chichester, England, 1976
- [14] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, 1990
- [15] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander, *Mathematical textbook of deformable neuroanatomies*, Proceedings of the National Academy of Science **90**(24) (December 1993)
- [16] R. E. Mortensen, *Maximum-likelihood recursive nonlinear filtering*, JOTA **2**(6), 386–394 (1968)
- [17] S. Timoshenko, *Theory of Elasticity*, McGraw-Hill, New York, 1934
- [18] A. Trouvé, *Habilitation à diriger les recherches*, Technical Report, University Orsay, 1996
- [19] W. P. Ziemer, *Weakly Differentiable Functions*, Springer-Verlag, New York, 1989