OXFORD CENTRE FOR COLLABORATIVE APPLIED MATHEMATICS

Report Number 11/60

Bayesian data assimilation in shape registration

by

C.J. Cotter, S.L. Cotter, and F.-X. Vialard

Oxford Centre for Collaborative Applied Mathematics
Mathematical Institute
24 - 29 St Giles'
Oxford
OX1 3LB
England

Bayesian data assimilation in shape registration

C.J. Cotter[1], S.L. Cotter[2] and F.-X. Vialard[3]

E-mail: cotter@maths.ox.ac.uk
[1]Department of Aeronautics, Imperial College London, London SW7 2AZ United Kingdom
[2]Mathematical Institute, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, United Kingdom
[3]Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, London SW7 2PG, United Kingdom

Abstract. In this paper we apply a Bayesian framework to the problem of geodesic curve matching. Given a template curve, the geodesic equations provide a mapping from initial conditions for the conjugate momentum onto topologically equivalent shapes. Here, we aim to recover the well defined posterior distribution on the initial momentum which gives rise to observed points on the target curve; this is achieved by explicitly including a reparameterisation in the formulation. Appropriate priors are chosen for the functions which together determine this field and the positions of the observation points, the initial momentum p_0 and the reparameterisation vector field ν, informed by regularity results about the forward model. Having done this, we illustrate how Maximum Likelihood Estimators (MLEs) can be used to find regions of high posterior density, but also how we can apply recently developed MCMC methods on function spaces to characterise the whole of the posterior density. These illustrative examples also include scenarios where the posterior distribution is multimodal and irregular, leading us to the conclusion that knowledge of a state of global maximal posterior density does not always give us the whole picture, and full posterior sampling can give better quantification of likely states and the overall uncertainty inherent in the problem.

1. Introduction

Geodesics on shape space induced from diffeomorphisms are proving to be a powerful tool in the field of computational anatomy [1, 2]. Not only do they provide a notion of distance between topologically equivalent shapes, they also provide a linear characterisation of patches of shape space centred on a template image using the initial conditions of the conjugate momentum along a geodesic [3]. This permits the application of linear statistical techniques (the adaptation of principal component analyis known as principal geodesic analysis, for example [4]) since the momentum inhabits a linear cotangent space. Bayesian statistical techniques have been developed in this area in recent years, with advances in methods of determining the most likely template for a dataset as well as the most likely metric [5, 6, 7]. There are many different shape spaces: the shape of curves, landmarks, images, fibre bundles *etc.*, but there is a unifying paradigm since in all cases underlying velocity field that generates the diffeomorphisms that act on the shapes evolves according to the EPDiff equation [8].

In this paper, we shall address the following inverse problem: given noisy observations of points around a curve (which may have been obtained from a segmentation algorithm or human input), what is the initial momentum that generates this curve from a given template? We concentrate on the parameterisation-independent problem, in which we treat curves as being equivalent if they are related *via* reparameterisation. This is a challenging problem which has been addressed in a few different ways [9, 10], but here we shall make use of the approach of [11] in which a reparameterisation variable is included explicitly, and the commutation of the reparameterisation with the geodesic flow equations is exploited. In the inverse problem nomenclature, this allows us to define an observation operator that produces point values on the curve, at a cost of introducing this extra variable.

In this paper, we will first define this problem in the framework of a minimisation problem over solutions of a PDE. We will then show how this minimisation problem can be translated into a data assimilation problem on function spaces. At this point, we will be able to utilise a general data assimilation framework, such as that used in fluid mechanics problems [12, 13, 14], that allows us to sample directly from the Bayesian posterior distribution, allowing us to find the highly likely values of the functions, and quantify the inherent uncertainty. The method relates closely to the Lagrangian data assimilation application of Bayesian inverse problems [12, 14], in the sense that we are trying to find a flow field which has advected one curve into another, guided by noisy observations of the final curve. Thus we construct a very similar approach, once again framed on function space. We go on to prove results regarding the forward problem which then allow us to frame the ill-posed inverse problem as a well-posed Bayesian inverse problem on function spaces with well-defined posterior measures for which we have algorithms [15] with which we are able to draw samples. We will then go on to present numerical results from the algorithm, which does not degenerate if a finer discretisation for the functions of interest is chosen, unlike its finite dimensional more common ancestor.

In Section 2 we introduce the problem, and describe the equations of motion of the shape as it is deformed by the velocity field, and demonstrate how we can find geodesic paths in shape space between two shapes. We also show in this section that the deformed shape is Lipschitz continuous with respect to two functions, the initial momentum p_0, and the reparameterisation function ν.

In Section 4.1 we frame the inverse problem by defining the observations which we are trying to assimilate. In Section 4.2 we define the prior distributions on the unknown functions p_0 and ν, informed by the analysis in Section 2.

In Section 5 we describe the Markov chain Monte Carlo (MCMC) algorithm that we use to sample directly from the well-defined posterior density on the two unknown functions of interest. In Section 6 we briefly discuss how we numerically approximate the dynamics described in Section 2, before presenting some illustrative numerical examples in Section 7. We finish in Section 8 with some conclusions and discussion.

2. Description of the forward model

In this section, we review the equations of motion for curves in the plane acted on by geodesics in the diffeomorphism group, and explain how these equations can be used to provide a mapping from the space of periodic scalar functions (which turn out to be the normal component of a conjugate momentum variable) to topology-preserving nonlinear deformations of some chosen curve in the plane. We then provide some analytical results that are required for defining the associated Bayesian inverse problem.

In our approach, we parameterise an oriented curve in the plane as a continuous function q from a space S (such as the circle, S^1) into $\mathbb{R}^2$ *i.e.*, $q \in C^0(S^1, \mathbb{R}^2)$. Although in this paper we concentrate on the lower dimensional case of a curve in the plane, it may be extended to surfaces in $\mathbb{R}^3$, which provides many applications in medical imaging, for example. We parameterise an evolving curve as $q(s,t)$, where $s \in S$ is the parameter around the curve, and $t \in [0,1]$ is the time parameter. Following the methodology of [16, 17, 10] we constrain the motion of the curve $q(s,t)$ to the action of diffeomorphisms by requiring that

$$\frac{\partial}{\partial t} q(s,t) = u(q(s,t),t) \tag{1}$$

where $u(x,t)$ is a time-parameterised family of vector fields on $\mathbb{R}^2$. This guarantees that the topology of the curve is preserved (*i.e.* there are no overlaps or cavitations). We wish to study curve evolution from a template curve Γ^1 (parameterised by $q^1(s)$) to a target curve Γ^2 (parameterised by $q^2(s)$). However, since we only want information about the shape of the curve, and not the specific parameterisation, we do not wish to enforce that any specific point $q^1(s)$ gets mapped to any specific point on Γ^2. We instead use the following generalised boundary condition

$$q(s,0) = q^1(\eta(s)), \quad q(s,1) = q^2(s), \tag{2}$$

for arbitrary reparameterisations $\eta \in \mathrm{Diff}_+(S)$, the orientation-preserving subgroup of the diffeomorphism group $\mathrm{Diff}(S)$ on S. If the boundary conditions (2) are satisfied, we say that u describes a *path* between Γ^1 and Γ^2. We select a function space B for vector fields, and define the *distance* along the path as

$$\int_0^1 \frac{1}{2} \|u\|_B^2 \, \mathrm{d}t. \tag{3}$$

For simplicity we assume that B is a Hilbert space and that there exists an operator A such that

$$\|u\|_B^2 = \langle u, Au \rangle_{L^2}. \tag{4}$$

The *shortest path* between Γ^1 and Γ^2 is defined by minimising (3) over u and η subject to (1) and the boundary conditions (2). To obtain the equations of motion for the shortest path, we introduce Lagrange multipliers $p(s,t)$ (which we call the "momentum") to enforce (1), and seek extrema of the action

$$\mathcal{A} = \int_0^1 \frac{1}{2}\|u\|_B^2 + \langle p, \dot{q} - u(q)\rangle \,\mathrm{d}t.$$

This leads to the following equations of motion in weak form:

$$\langle \delta u, Au\rangle_{L^2} - \langle p, \delta u(q)\rangle \qquad = 0, \tag{5}$$

$$\int_0^1 \left\langle \delta p, \frac{\partial q}{\partial t} - u(q) \right\rangle \mathrm{d}t \qquad = 0, \tag{6}$$

$$\int_0^1 \left\langle p, \frac{\partial \delta q}{\partial t} - (\nabla u(q))\delta q \right\rangle \mathrm{d}t = 0, \tag{7}$$

where δp and δq are space-time test functions, with

$$u, \delta u \in B, \quad p, \delta p \in L^2, \quad q, \delta q \in H^1.$$

If the minimisation is taken over all reparameterisations $\eta \in \mathrm{Diff}_+(S)$ then we obtain the condition

$$p \cdot \frac{\partial q}{\partial s} = 0, \quad \forall t \in [0,1]. \tag{8}$$

The condition states that the momentum p is normal to the shape. As described in [11], if this condition is satisfied, then the curve evolution is invariant under the transformation

$$(p,q) \mapsto (\bar{p},\bar{q}) = \left(p \circ \eta \frac{\partial \eta}{\partial s}, q \circ \eta\right) \tag{9}$$

for $\eta \in \mathrm{Diff}_+(S)$, which is the cotangent-lift of the transformation $q \mapsto q \circ \eta$. This means that if equations (5-7) are solved with initial conditions (p,q) and $(\bar{p},\bar{q})$ then $\bar{q} = q \circ \eta$ at time 1. As described in [18], solutions that satisfy Condition 8 are parameterised realisations of geodesics on the shape space $C^0(S^1,\mathbb{R}^2)/\mathrm{Diff}_+(S)$.

We define the time-1 flow map Ψ

$$\Psi(p|_{t=0}, q|_{t=0}) = q|_{t=1},$$

where p and q satisfy equations (5-7). Having fixed $q|_{t=0} = q^1$, we define a parameterisation-independent mapping between scalar functions $p_0 : S \mapsto \mathbb{R}$ (the normal component of the initial conditions of p) and topologically equivalent curves q^2 obtained by

$$q^2 = \Psi(p_0 n, q^1),$$

where n is the normal vector to the curve q^1. The power of this mapping is that it allows us to perform linear operations on the space of functions p_0, such as averages.

In this paper, our aim is to estimate the probability measure for p_0 given a set of observed points $(q_1,\ldots,q_n)$ from the curve q^2. These points may have been obtained from a digitised medical image, for example. These points are assumed to be sorted in order according to the orientation of the curve. We cannot directly solve the inverse problem of finding p_0 such that

$$\Psi(p_0 n, q^1)(s_i) = q_i, \qquad i = 1,\ldots,n,$$

since we do not know the values s_i of the curve parameter s that correspond to each q_i. To fix this, we introduce a reparameterisation variable $\eta \in \mathrm{Diff}_+(S^1)$, and seek (p_0, η) such that

$$\Psi\left((p_0 n)\circ\eta\,\frac{\partial\eta}{\partial s}, q^1\circ\eta\right)(s_i) = q_i, \quad i = 1,\ldots,n,$$

with $\{s_i\}_{i=1}^n$ is some chosen distinct sequence of points in S ordering according to orientation, *e.g.* equispaced points. This guarantees to preserve the ordering since the curve is evolved by a diffeomorphism. The introduction of the reparameterisation variable does not alter the shape of the obtained curve, just the particular parameterisation used.

In this paper, we construct reparameterisations from time-independent velocity fields according to

$$\frac{\partial\chi}{\partial t}(s,t) = \nu(\chi(s,t)), \quad \chi(s,0) = \mathrm{Id},\ \eta(s) = \chi(s,1), \tag{10}$$

where Id is the identity map, and ν is a scalar periodic function on I which we refer to as the generating vector field of the reparameterisation η. This is known as *Lie exponentiation*‡, and guarantees that η is an orientation preserving, smooth invertible map (for sufficiently smooth ν). This defines a reparameterisation map:

$$\mathcal{R}(p,q,\nu) = \left(\frac{(p_0 n)\circ\eta}{\frac{\partial\eta}{\partial s}}, q^1\circ\eta\right)$$

where η is obtained from (10). Hence we define the observation operator $\mathcal{G}$ by

$$\mathcal{G}(p_0,\nu) = \begin{pmatrix} \Psi\circ\mathcal{R}\,(p_0 n, q^1(s_1),\nu) \\ \vdots \\ \Psi\circ\mathcal{R}\,(p_0 n, q^1(s_n),\nu) \end{pmatrix} \tag{11}$$

The normal component of p_0 then characterises the shape of the target curve Γ^2 relative to the curve Γ^1, whilst the generator variable ν merely describes the reparameterisation of the target curve which is obtained at the minimum.

3. Properties of the Observation Operator

In this section, we show that the the observation operator $\mathcal{G}$, given in Equation (11), is Lipschitz in p_0 and ν.

To show that $\mathcal{G}$ is Lipschitz, we first prove existence, uniqueness and Lipschitz continuity for the maps Ψ and $\mathcal{R}$. This is then used to show that the observation operator is Lipschitz continuous with respect to the normal component p_0 of the initial conditions p and the generator variable ν.

Let B be the space of vector fields, Q be a Hilbert space and P be its dual. Here Q represents a space of curves with a particular norm, for example $Q = L^2(S_1,\mathbb{R})$ or $Q = H^1(S_1,R)$. We want to define the momentum map is defined by

$$J : (p,q) \in P\times Q \mapsto J(p,q) := \left(v \mapsto \langle p, v\circ q\rangle_{P,Q}\right) \in B^*\,. \tag{12}$$

‡ As described in [19], not all reparameterisations of the circle can be obtained from Lie exponentiation, and a more general approach would be to generate reparameterisations from geodesics in a way that is similar to how the deformations of the curves are obtained (known as Riemann exponentiation).

In this aim, we assume in addition that the map $B \times Q \mapsto Q$ defined by $(v, q) \to v \circ q$ is a Lipschitz map on Q for a fixed $v \in B$ namely there exists C_c s.t.

$$\|v \circ q_1 - v \circ q_2\|_Q \leq C_c \|v\|_B \|q_1 - q_2\|_Q \,. \tag{13}$$

As a consequence, since this composition map is linear on B, the map is Lipschitz on $B \times Q$. This assumption makes $J(p, q)$ well-defined and Lipschitz on $P \times Q$ for the dual norm on B^*.

Definition 1. *We say that B a Hilbert space of vector fields is embedded in $C^n(\Omega, \mathbb{R}^d)$ if there exists a constant C_e s.t. $\|v\|_{n,\infty} \leq C_e \|v\|_B$. In that case we also say for notation convenience that B is $n-$admissible.*

Proposition 2. *The assumption above is satisfied in the following situations*

- *B is $1-$admissible and $P = Q = L^2(S_1, \mathbb{R})$.*
- *B is $2-$admissible, $P = H^{-1}(S_1, \mathbb{R})$ and $Q = H^1(S_1, \mathbb{R})$.*

Proof. For both points, we use the integration formula for the path $q(t) = tq_1 - (1 - t)q_0$

$$f(q(1)) - f(q(0)) = \int_0^1 df(q(t))(q_1 - q_0)\, dt \,, \tag{14}$$

for $f \in C^1$. For the first point, for any $f \in C^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$,

$$\|f \circ q_1 - f \circ q_2\|_{L^2} \leq \|f\|_{1,\infty} \|q_1 - q_2\|_{L^2} \,, \tag{15}$$

which gives the result. For the second point, we note that $H^1(S_1, \mathbb{R})$ is Banach algebra, namely for any couple $(f, g) \in H^1(S_1, \mathbb{R})$ we have

$$\|fg\|_{H^1} \leq \|f\|_\infty \|g\|_{H^1} + \|g\|_\infty \|f\|_{H^1} \leq 4\|f\|_{H^1} \|g\|_{H^1} \,.$$

Therefore in the Formula (14), the term $df(q(t))(q_1 - q_0)$ can be understood as the product (generalized to $\mathbb{R}^d$) of two elements in $H^1(S_1, \mathbb{R}^d)$ provided that $df(q(t))$ belongs to H^1. Moreover H^1 is stable under the composition with a C^1 function bounded for

$$\|f \circ q\|_{H^1} = \|f\|_{1,\infty} \|q\|_{H^1} \tag{16}$$

The second point then follows from

$$\|f \circ q_1 - f \circ q_2\|_{H^1} \leq \|f\|_{2,\infty} \|q_1 - q_2\|_{H^1} \,, \tag{17}$$

□

We write the shooting equations as follows:

$$\dot{q} = u \circ q \tag{18}$$

$$\dot{p} = -\nabla u(q).p \tag{19}$$

$$u = J(p, q)^\sharp \,, \tag{20}$$

where $J(p, q)^\sharp$ denotes the corresponding element in B given by the Riesz theorem, i.e.

$$\langle J(p, q)^\sharp, v \rangle_B = J(p, q)(v) \,.$$

Equation (19) is a pointwise matrix multiplication when p is smooth. When $p \in H^{-1}$ for instance, the operation is defined by duality. Let us detail this point to make the notations clear:

Definition 3. *Let Q be a Banach algebra then the multiplication $Q^* \times Q \mapsto Q$ is defined by duality*

$$\langle q_1 . p, q_2 \rangle = \langle p, q_1 q_2 \rangle \tag{21}$$

Remark 4. *The multiplication on a Banach algebra is by definition bilinear continuous. As a consequence, the multiplication inherits the same properties and in particular, the Lipschitz property.*

We now prove that these shooting equations are a simple ODE in infinite dimensions i.e. $\dot{X} = F(X)$ with state variable X in a Banach space; This function F satisfies the classical Lipschitz assumption which enables to define the solutions via Picard's fixed point and it also ensures that the solutions are Lipschitz functions of the initial conditions.

Theorem 5. *System (18-20) is an ODE on $Q \times P$ which satisfies the Lipschitz condition for the two following cases*

- *B is $2-$admissible, $P = L^2$ and $Q = L^\infty$*
- *B is $3-$admissible, $P = H^{-1}$ and $Q = H^1$.*

on bounded balls in $P \times Q$.

Remark 6. *The space $Q = L^\infty(S_1)$ is not a Hilbert space in this case but only a Banach space.*

Proof. In the two claimed cases, we have by Proposition 2 that $(p, q) \to J(p, q)^\sharp$ is Lipschitz for the norm on B. Then using the inequality (15), we get that $(p, q) \to J(p, q)^\sharp \circ q$ is Lipschitz as the composition of two Lipschitz functions. Thus, we have treated the first component (18) of the ODE.

Using Remark 4, it is then sufficient to prove that $(p, q) \in P \times Q \mapsto d[J(p, q)^\sharp](q) \in Q$ is a Lipschitz mapping. Denoting $v_i = J(p_i, q_i)^\sharp$ for $i = 1, 2$, use of the triangle inequality gives

$$\begin{aligned} \|dv_1 \circ q_1 - dv_2 \circ q_2\|_Q &\leq C_e \|v_1\|_B \|q_1 - q_2\|_Q + \|dv_1 \circ q_2 - dv_2 \circ q_2\|_Q \\ &\leq C_e \|v_1\|_B \|q_1 - q_2\|_Q + C_e \|v_1 - v_2\|_B \|q_2\|_q \,, \end{aligned} \tag{22}$$

where we rely on the inequality (17) for the H^1 case and (15) for the first case.

Up to this point, we only have proven a local Lipschitz condition. It remains to prove that for $(p_0, q_0) \in B(0, r) \subset P \times Q$ with r a positive real number, there exists a constant $M_{r,t}$ (the subindex indicating that this constant only depends on r and the time t) such that for all time $t \geq 0$ the system (18-20) is Lipschitz of Lipschitz constant M_r. In particular this implies that the solutions are defined for all times. To this end, we remark that the geodesic equations (18-20) are variational with the Hamiltonian is constant in time. This can be checked by a straightforward calculation. Since the Hamiltonian reads $H(p, q) = \left\| J(p, q)^\sharp \right\|_B^2$, the Lipschitz inequality (22) becomes

$$\|dv_1 \circ q_1 - dv_2 \circ q_2\|_Q \leq 3r \, C_e \, \|q_1 - q_2\|_Q \,. \tag{23}$$

For the equation (18), the inequality (13) implies

$$\|v_1 \circ q_1 - v_2 \circ q_2\| \leq r C_c \, \|q_1 - q_2\|_Q + \|v_1 - v_2\|_B \max\left(\|q_1\|_Q \,, \|q_2\|_Q\right) \tag{24}$$

but by direct integration of Inequality (13) we have that $\max\left(\|q_1\|_Q \,, \|q_2\|_Q\right) \leq C_c \, r \, t$ for any time $t > 0$. The same type of inequality is true for the momentum p, hence the geodesic equations (18-20) are Lipschitz for a constant $M_{r,t}$ that has polynomial growth in t and in r. □

We now claim the following result which is based on Grönwall's lemma:

Corollary 7. *Solutions of the geodesic equations (18-20) have a Lipschitz dependency w.r.t. their initial conditions for their respective norms.*

Theorem 8. *Let B be $2-$admissible and D be $1-$admissible. Then for $p_0 \in L^2$ and $q_0 \in W^{1,\infty}$, the observation operator $\mathcal{G}$ is Lipschitz continuous.*
Let B be $3-$admissible and D be $1-$admissible. Then for $p_0 \in H^{-1}$ and $q_0 \in W^{1,\infty}$, the observation operator $\mathcal{G}$ is Lipschitz continuous.

Proof. The observation operator $\mathcal{G}$ is the evaluation at points $(\eta(s_i))_{i=1}^n$ of the deformed template $q(1)$ obtained by the geodesic equations (18-20) without reparameterisation. In other words,

$$\mathrm{Ev}(p_0, q_0, \nu) = \begin{pmatrix} q(1) \circ \eta(s_1) \\ \vdots \\ q(1) \circ \eta(s_n) \end{pmatrix} \tag{25}$$

where $q(1)$ is the solution at time 1 of the geodesic equations (18-20) for initial data p_0, q_0 and η is the Lie exponential of ν. To prove that the observation operator is Lipschitz in those three variables, we observe that it is the composition of the flow of the geodesic equations (18-20) and the right composition with the reparameterisation by η. The first operation is Lipschitz continuous by Corollary (7). Now the right composition is also Lipschitz continuous since $q(1) \in W^{1,\infty}$ (indeed since B is $2-$admissible or $3-$admissible, the flow of the geodesic equations (18-20) preserves $W^{1,\infty}$) we have

$$\|q(1) \circ \eta_1 - q(1) \circ \eta_2\|_\infty \leq \|\nabla q(1)\|_\infty \|\eta_1 - \eta_2\|_\infty \leq M \|\nabla q(1)\|_\infty \|\nu_1 - \nu_2\|_D \,. \tag{26}$$

□

4. Bayesian Inversion

In this section we first review what Bayesian inversion is, and what kind of solution to the problem we are aiming to find. We will then focus specifically on the problem of finding the probability distributions on the pair of functions (p_0, ν) given observations of the target shape. This will involve using the results from Section 3 to identify an appropriate choice of prior distribution on the functions.

Uncertainty in this problem arises from errors incurred during observation. Observational noise is often modelled as being additive Gaussian. This uncertainty in the exact shape of the curve leads to uncertainty in the value of the pair of functions which would deform the reference shape to the observed shape. Understanding and quantifying that uncertainty is a key part of solving this problem, as naive least squares matching to the data would lead to an ill-posed problem.

Bayesian inversion is one way to regularise the problem, and convert this ill-posed inverse problem into a well posed one. Bayes' formula is central to this. Prior beliefs about the quantities of interest are blended with data from observations to give the posterior distribution. The finite dimensional version of this formula is

$$\mathbb{P}(X|\mathcal{D}) \propto \mathbb{P}(\mathcal{D}|X)\mathbb{P}(X),$$

where X denotes the quantities of interest, $\mathbb{P}(X)$ the probability distribution describing our prior beliefs about those quantities, and $\mathbb{P}(\mathcal{D}|X)$ the likelihood function, which returns the likelihood that we would make the observations $\mathcal{D}$ given a particular configuration X. If X is infinite dimensional, however, then we must use the analogous result regarding the Radon-Nikodym derivative of the posterior distribution with measure μ, with respect to the prior distribution on X with measure μ_0:

$$\frac{d\mu(p_0,\nu)}{d\mu_0} \propto \mathbb{P}(\mathcal{D}|X). \tag{27}$$

The likelihood function can still be calculated explicitly in this setting since the observations are finite dimensional. In order for this likelihood to be calculated, assumptions must be made about the form of the observations.

Note that this formula only holds if the posterior is indeed absolutely continuous with respect to the prior distribution, otherwise no such derivative is admitted. Therefore careful choice of prior distribution must be made in order for the problem to be well posed. One way to ensure that this absolute continuity condition holds is to choose a prior distribution which has full measure on a space of functions for which the likelihood function is Lipschitz continuous, and this is the motivation for the calculations is Section 3.

The role of the prior in this respect is two-fold. The prior can contain prior information or beliefs about the form or structure of the solution, maybe taken from previous observations. However another important role of the prior is to regularise the problem, to make it well-posed. Indeed it plays a very similar role to the penalty term in a Tikhonov regularisation, an optimisation formulation of the inverse problem. We will explore these simalarities in Section 5.1, and exploit them in our numerical method.

The solution in a Bayesian inverse formulation is given by the posterior distribution itself. Once a suitable prior has been chosen in order to regularise the problem sufficiently to make it well-posed, the problem comes down to characterising the posterior probability distribution on the quantities or functions of interest.

4.1. The Observations

We assume that observations y of the quantity of interest are noisy in nature, satisfying:

$$y = \mathcal{G}(p_0,\nu) + \xi, \qquad \xi \sim \mathcal{N}(0,\Sigma),$$

where Σ is assumed to be known, and where $\mathcal{G}$ is defined to be the observation operator given by

$$\mathcal{G}(p_0,\nu) = \left\{\Psi \circ \mathcal{R}\left(p_0 n, q^1(s_i), \nu\right)\right\}_{i=1}^{n}.$$

That is, the function which returns to us what noiseless observations we would make if we were to deform the template shape and reparameterise using the function pair (p_0,ν). This assumption allows us to compute the likelihood that y was observed with a given (p_0,ν), through the probability density function of ξ:

$$\mathbb{P}(y|p_0,\nu) \propto \exp\left(-\frac{1}{2}\|\mathcal{G}(p_0,\nu) - y\|_\Sigma^2\right),$$

where $\|x\|_\Sigma^2 := x^T\Sigma^{-1}x$ is the covariance weighted norm.

4.2. Prior and Posterior Distributions on (p_0, ν)

By Bayes law the Radon-Nikodym derivative of the posterior with respect to the prior, assuming μ is μ_0-measurable, is given by a function proportional to the likelihood. That is, in our case, the measure with Radon-Nikodym derivative, assuming μ is μ_0-measurable, given by equation (27).

For the posterior distribution to be well defined, it must be absolutely continuous with respect to the choice of prior distribution on the two unknown functions of interest. This choice of prior distribution is informed by the properties of the forward model, and the observation operator $\mathcal{G}$.

Theorem 9. *If* $\mathcal{G} : H \to \mathbb{R}^N$ *is* μ_0*-measurable then the posterior measure* $\mu(dx) = \mathbb{P}(dx|y)$ *is absolutely continuous with respect to the prior measure* $\mu_0(dx)$ *and has Radon-Nikodym derivative given by (27).*

Proof. Proof given in Theorem 2.1 in [12]. □

Together with the results in Section 3, this theorem helps us to choose such appropriate prior distributions on the unknown functions. With this in mind, let us consider the Helmholtz operator

$$\mathcal{H} = \ell I - \Delta,$$

where $\ell \in \mathbb{R}$ defines the *length scale*. This is a positive definite operator with the same smoothing properties as the Stokes operator. However, unlike the Stokes operator its inverse is well-defined on constant functions. Let us consider the following result, concerning the regularity of draws from the distribution $v \sim \mathcal{N}(0, \delta\mathcal{H}^{-\alpha})$.

Lemma 10. *Suppose* $v \sim \mathcal{N}(0, \delta\mathcal{H}^{-\alpha})$ *for some* $\alpha > s + \frac{d}{2}$ *for some* $s \geq 0$*. Then* $v \in H^s$*.*

Proof. Suppose $v \sim \mathcal{N}(0, \delta\mathcal{H}^{-\alpha})$, then by the Karhunen-Loeve expansion,

$$v = \sum_k \sqrt{\lambda_k}\phi_k\xi_k, \qquad \xi_k \sim \mathcal{N}(0,1)\, i.i.d,$$

where $\phi_k = \exp(2\pi i k)$ are basis functions in $L^2([0,1])$, and $\lambda_k = \delta(\ell + 4\pi^2|k|^2)^{-\alpha}$ are the corresponding eigenvalues of $\delta\mathcal{H}^{-\alpha}$. This implies that

$$\begin{aligned}\mathbb{E}(\|v\|_s^2) &= \mathbb{E}\left(\sum_k |k|^{2s}|\lambda_k|\xi_k^2\right) \\ &= \sum_k |k|^{2s}|\lambda_k|.\end{aligned}$$

Here the bigger s is, the faster the λ_k must decay if the sum is to converge to a finite limit. Moreover, since

$$\mathbb{E}(\|v\|_s^2) < \infty \implies \|v\|_s < \infty \quad a.s.,$$

we have that $\mathbb{E}(\|v\|_s^2) < \infty$ is sufficient to show that $v \in H^s(\Omega)$ almost surely.

$$\begin{aligned}\mathbb{E}(\|v\|_s^2) &= \sum_k |k|^{2s}|\lambda_k| \\ &\leq C\sum_k |k|^{2s}|k|^{-2\alpha} \\ &= C\sum_k |k|^{2(s-\alpha)}.\end{aligned}$$

By comparison to integrals it follows that for v to be in H^s we require $\alpha > s + \frac{d}{2}$. □

The length scale parameter $\ell \in \mathbb{R}^+$ allows us to control at which scale the smoothing properties of the Laplacian take effect. With a larger value of ℓ, a larger value of $|k|$ is required before the effect of the Laplacian becomes dominant. Likewise, as $\ell \to 0$, $\mathcal{H} \to -\Delta$, meaning that the Laplacian is dominant on all scales. The choice of this value, however, does not affect the overall regularity of samples drawn from the distribution $\mathcal{N}(0, \mathcal{H}^{-\alpha})$.

Corollary 11. *Let B be 3-admissable, D be 1-admissable, and $\mu_0(p_0, \nu) = \mathcal{N}(0, \delta_1 \mathcal{H}_1^{-\alpha_1}) \times \mathcal{N}(0, \delta_2 \mathcal{H}_2^{-\alpha_2})$ for $\alpha_1 > -\frac{1}{2}$, $\alpha_2 > 3/2$, $\delta_1, \delta_2 \neq 0$, and where $\mathcal{H}_i = (\ell_i I - \Delta)$ where $\ell_i \in \mathbb{R}^+$ for $i \in \{1, 2\}$. Then $\mathcal{G}$ is measurable with respect to μ_0, and the posterior measure μ is absolutely continuous with respect to μ_0, with Radon-Nikodym derivative given by (27).*

Proof. Result follows by the Sobolev embedding theorem, Theorem 9, Lemmas 10, and Theorem 8. □

5. Characterising the Posterior Density

In this section we introduce the statistical methods which we use in order to characterise the posterior distribution of the function pair (p_0, ν) given a choice of prior distribution and a set of noisy observations of the target shape. Monte Carlo Markov chains (MCMC) are a set of tools which allow us to sample from a desired probability distribution. A chain of states is formed, which converges in distribution to the target density.

One aspect of using Markov chain Monte Carlo methods, is the way in which we initialise the chain. If we start the chain in the tails of the distribution which we are trying to characterise, then it takes a while before we can assume we are drawing samples from the distribution, as the chain has entered equilibrium. Therefore, often a "burn-in" is used, a period of time for which the MCMC chain is iterated, but where the results are ignored, until the chain has entered equilibrium.

However, since the forward model in this problem is reasonably costly, this "burn-in" period may be very expensive, and we may wish to use an alternative way of ensuring that the chain starts in equilibrium. By solving a related optimisation problem, we can start the chain in an area of higher probability density. The following subsection describes this optimisation problem.

5.1. Relationship with Tikhonov Regularisation

One approach to data assimilation is to try to find the minimiser of the following expression:

$$\min_u \|\mathcal{G}(u) - y\|_\Sigma^2.$$

Solutions to this problem can, however, be very rough and have undesirable properties, or the solution may not be unique or even exist. Therefore it may be appropriate to add a penalty term to this minimisation to ensure minimal amounts of regularity in the solution, so that we are trying to minimise J where

$$J(u) = \min_u \frac{1}{2}\|\mathcal{G}(u) - y\|_\Sigma^2 + \frac{1}{2}\|u\|_X^2$$

for some function space X. This is termed *Tikhonov* regularisation [20]. If X is the Cameron-Martin space for some Gaussian measure μ_0, for which the measure $\mu = \mathbb{P}(u|y)$ is absolutely continuous, then this problem is equivalent to trying to maximise the probability density from the Bayesian posterior we have previously defined. As such, the theorems in Sections 2 and 4.2 are also applicable to those wishing to attempt these types of variational approach, as they give insight into good choices of penalty term to ensure existence of global minima of J.

In this context, however, we wish to solve this optimisation problem in order to initialise our Markov chain for full posterior sampling. That is, finding the solution to:

$$\min_{p_0 \in L^2, \nu \in H^2} L(p_0, \nu) = \min_{p_0 \in L^2, \nu \in H^2} \frac{1}{2}\|\mathcal{G}(p_0, \nu) - y\|_\Sigma^2 + \frac{1}{2}\|(p_0, \nu)\|_{\mu_0}^2$$

where $\|(\cdot,\cdot)\|_{\mu_0}$ is the equivalent penalty term to that induced in the posterior measure by the choice of prior measure μ_0, or the *Cameron-Martin* norm corresponding to μ_0. In particular, if (as is the case for our purposes) $\mu_0(p_0, \nu) = \mathcal{N}(0, \delta_1 \mathcal{H}^{-\alpha_1}) \times \mathcal{N}(0, \delta_2 \mathcal{H}^{-\alpha_2})$, then

$$\|(p_0, \nu)\|_{\mu_0}^2 = \sum_k \delta_1 |p_k|^2 (\ell_1 - |k|^2)^{\alpha_1} + \delta_2 |\nu_k|^2 (\ell_2 - |k|^2)^{\alpha_2}$$

Various descent methods can be used to try to find local minima of this quantity. Most of these methods incorporate gradient information of some type to attempt to search for the minima in appropriate directions. These methods include steepest descent and conjugate gradient among others.

The method that we will utilise in our numerics in order to initialise our Markov chains is called the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [21, 22]. This method uses the gradient information from the last two states in the chain to approximate a Hessian matrix, which is then used to choose an appropriate direction in which to search for the local minimum.

5.2. The Random Walk Metropolis Hastings (RWMH) algorithm

For the purpose of the numerics, we use a version of the Random Walk Metropolis Hastings (RWMH) algorithm framed on function spaces. This method is superior to the usual vanilla random walk method in that it is well defined on function spaces, and as such is robust under refinement of any discretisation of the forward model [15]. This means that the rate of convergence of statistics in the Markov chains are independent of this discretisation, whereas the Metropolis Hasting algorithm with vanilla random walk proposal degenerates as the forward model's discretisation is refined. The following describes the RWMH sampler, where $\Phi = \frac{1}{2}\|\mathcal{G}(p_0, \nu) - y\|_\Sigma^2$:

1: $u_0 = (p_0, \nu)_0$
2: **for** $i = 1 : N$ **do**
3: Sample $v = (1 - \beta^2)^{1/2} u_{i-1} + \beta w$, where $w \sim \mu_0((p_0, \nu))$
4: $a(u_{i-1}, v) = \min\{1, \exp(\Phi(u_{i-1}) - \Phi(v)\}$
5: Sample $u \sim U([0, 1])$
6: **if** $u < a(u_{i-1}, v)$ **then**
7: $u_i = v$
8: **else**

9: $u_i = u_{i-1}$
10: **end if**
11: **end for**

This algorithm is almost identical to the usual finite dimensional version, whose proposal is given by:

$$v = u_{i-1} + \beta w, \qquad w \sim \mu_0((p_0, \nu)).$$

Using this proposal, the acceptance is altered, and is given by:

$$a(u_{i-1} = (p_0, \nu), v = \bar{p_0}, \bar{\nu}) = \min\left\{1, \exp(\Phi(u_{i-1}) + \frac{1}{2}|C_{p_0}^{-1/2} p_0|^2 + \frac{1}{2}|C_\nu^{-1/2}\nu|^2 - \Phi(v) - \frac{1}{2}|C_{p_0}^{-1/2}\bar{p_0}|^2 - \frac{1}{2}|C_\nu^{-1/2}\bar{\nu}|^2\right\},$$

where C_{p_0} and C_ν are the covariance operators of the priors on p_0 and ν respectively. These extra terms, concerned as they are with the regularity of the functions due to the prior, become bigger and bigger as the resolution of p_0 and ν are increased, and cause the average acceptance probability to become smaller and smaller. This in turn leads to the degeneration of the statistical algorithm, causing very slow convergence.

The choice of $\beta \in (0, 1)$, the step size of the proposal, is crucial in the efficiency of the algorithm. If too big a step size is chosen, then the proposal y will often be in the tails of the distribution, and will usually be rejected, meaning that the chain is not exploring the target density efficiently. Too small a step size will mean that many proposals are accepted, but the correlation between successive states is very high, and once again the exploration of the state space is slow. In the numerics that follow, β was tuned so that the acceptance rate of proposals was approximately 50%.

6. Numerical Approximation of $\mathcal{G}$

The observation operator consists of the concatenation of four maps: 1) the generation of the reparameterisation variable η from the vector field ν according to (10), 2) the application of the reparameterisation formula according to (9), 3) the time-1 flow map of the geodesic equations according to (5-7), and 4) the evaluation of the curve $q|_{t=1}$ at the finitely chosen points. Stages (1) and (3) are both numerically discretised using particle-mesh methods as described in [23]. These methods allow efficient evaluation of the velocity field at particle locations whilst preserving variational structure and making it possible to compute the exact adjoint equations of the discrete system which are used in the implementation of the deterministic burn-in. Stages (2) and (4) use cubic B-spline interpolation, which also make computing the adjoint tractable. Details of the numerical approximation procedure, along with numerical tests that show that the numerical scheme is second-order convergent in space, are given in [24].

As described in [13], since the MCMC algorithm used has been described on function space, as the approximation of model is refined, the sampled posterior converges to the true posterior; furthermore, the number of samples required for convergence of statistics remains bounded.

7. Numerical Results

We present numerical results to show that the algorithm is successfully drawing samples from the posterior distribution, and then go on to study the posterior distribution in certain data scenarios. The first thing to address is our choice of template shape Γ^1 and the parameterisation $q^1(s)$ that we use for this shape. Since we are only considering closed curves, it seems natural to choose a circle for this to keep things as simple as possible. We also wish to choose a nice smooth parameterisation for this shape, centred in the middle of our domain $\mathbb{T}^2 = [0, 2\pi)^2$ so we pick

$$q^1(s) = (\cos(s) + \pi, \sin(s) + \pi), \qquad s \in [0, 2\pi). \tag{28}$$

We will engage in simulation studies in which the data is itself produced by employing the numerical simulation of a forward PDE model. In all the numerics, we assume that the noise ξ through which we make the observations

$$y = \mathcal{G}(p_0, \nu) + \xi, \qquad \xi \sim \mathcal{N}(0, \Sigma),$$

has a diagonal covariance matrix Σ, which we describe for each experiment.

In terms of the approximation of the forward model in the algorithm, 50 time steps are used in which to deform the parameterisation of the shape from time $t = 0$ to $t = 1$. The curves themselves are approximated by 100 points, and with a 64×64 grid approximating the underlying velocity field. The values of s_j at which observations are made are given by $\left\{\frac{2\pi s_j}{N}\right\}_{j=0}^{N-1}$. These parameters are used in the model to create the data, and in the implementation of the statistical algorithm.

The prior distributions on p_0 and ν are $\mathcal{N}(0, \delta_1\mathcal{H}^{-\alpha_1})$ and $\mathcal{N}(0, \delta_1\mathcal{H}^{-\alpha_1})$ respectively, with $\alpha_1 = 0.55$, $\alpha_2 = 1.55$, $\delta_1 = 30$ and $\delta_2 = 0.05$. Note that these parameters are sufficient to ensure that Corollary 11 holds.

In the following sections we plot normalised histograms for particular Fourier modes of the two functions of interest, from converged Markov chains which are invariant with respect to the posterior distribution as described in Section 5.

7.1. Posterior Consistency

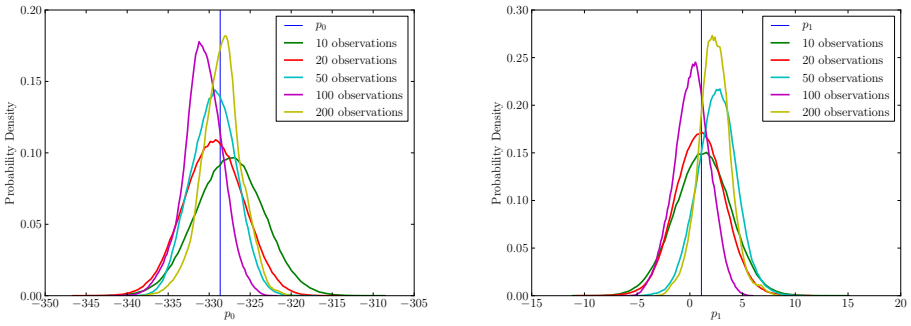
It seems reasonable to expect that as we increase the amount of informative data that we are using in our inference, the closer our posterior mean will be to the functions that created the data, and that at the same time the uncertainty in that estimation will decrease. In this set of numerical experiments, we take a draw (p_0, ν) from the prior measure, and using our approximation of the forward model, create data y such that

$$y = \left\{q^2\left(\frac{2\pi n}{N}\right)\right\}_{n=0}^{N-1} + \xi, \qquad \xi \sim \mathcal{N}(0, \Sigma),$$

with increasing N.

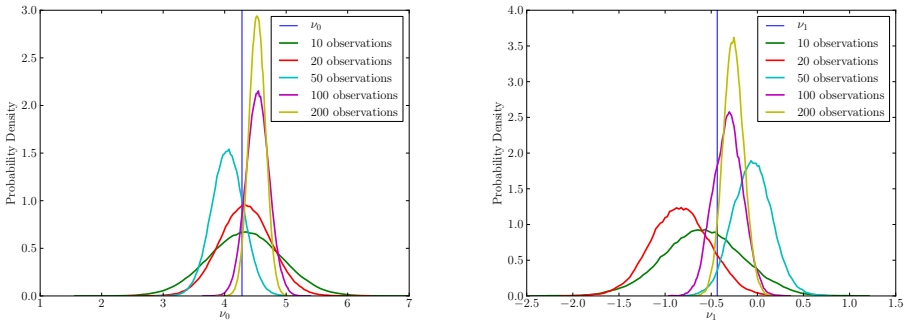
In this experiment, the data is simulated using the numerical approximation of the forward model. The initial momentum and reparameterisation functions that created the data were drawn from the prior distributions. Mean-zero Gaussian noise with covariance matrix $\Sigma = \sigma^2 I$ with $\sigma = 10^{-2}$ was added, and this same covariance matrix was used for the function $\Phi = \frac{1}{2}\|\mathcal{G}(p_0, \nu) - y\|_\Sigma^2$ in the likelihood functional.

In the following graphs we look at the marginal distributions for the two lowest frequency Fourier modes of both p_0 and ν as estimated by our MCMC method, for $N = 10, 20, 50, 100$, and 200. We choose the low frequencies as these are the Fourier modes which are most informed by our observations. Since we simulate the data ourselves, using the same approximation of the forward model as that used within the statistical algorithm, we know the values of the functions that we are trying to recover.



(a) (b)

Figure 1. *Plot of marginal distributions on (a) the lowest, (b) the second lowest wave number Fourier mode of the initial momentum function p_0, given varying numbers of observations. The vertical blue lines denote the value of this Fourier mode in the function which was used to create the data.*



(a) (b)

Figure 2. *Plot of marginal distributions on (a) the lowest, (b) the second lowest wave number Fourier mode of the reparameterisation function ν_0, given varying numbers of observations. The vertical blue lines denote the value of this Fourier mode in the function which was used to create the data.*

7.2. Effect of lengthscale on multimodality

One might ask what the advantage of full posterior sampling is over other less computationally expensive optimisation approaches. In this section we show that certain data scenarios can cause

the posterior distribution on the initial momentum p_0 to be complex with many local maxima of probability density. In this case the solution from an optimisation approach may not be unique, depending on the initial state of the solver. Being able to characterise the whole of the distribution in this case allows us to identify several different possibilities for the function p_0, and to get a better idea of the uncertainty in the problem. One such data scenario that exhibits this posterior multimodality is where we have features in the data whose lengthscale is smaller than the filtering lengthscale α in the metric for the diffeomorphism, as we demonstrate in the following example. Let us once again set the initial shape Γ^1 to be that defined by the parameterisation $p^A(s)$ given in (28). We then define $\Gamma^2(r)$ for $r \in [0, 1]$, such that it consists of a square of length 2, centered at $[\pi, \pi]$, with quarter-circles of radius r continuously added on each corner.

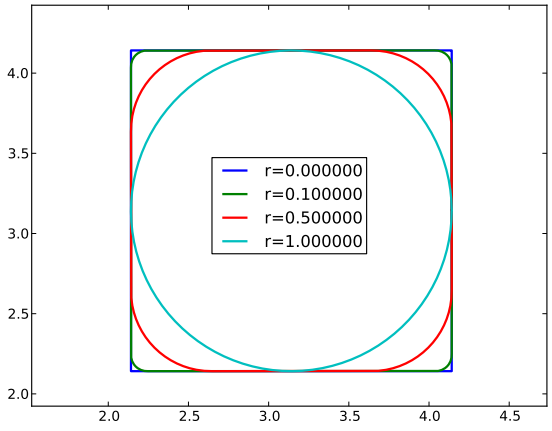
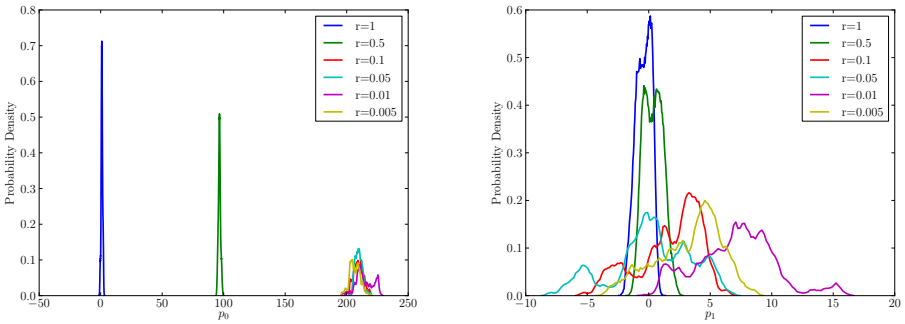


Figure 3. *Plot of the target shape with varying r, the radius of the quarter circles on the corners of the square.*

Figure 3 shows the target shape for a range of radii r. Note that for $r = 0$, the shape is exactly a square of length 2, and for $r = 1$ the shape is a circle of radius 2, exactly overlapping the template shape (28). The variable r defines the minimum lengthscale of features in the data. Once r becomes smaller than the maximum lengthscale than can be resolved by the model, we would expect to see multimodality creeping into the posterior distribution on p_0. Since the template shape cannot be bent to imitate the features on such a small scale, near to this feature only some of the observation points can be well matched for any given p_0. This problem does not cause multimodality in the posterior distribution on the reparameterisation function ν since the observation points on the template shape do not need to move in order to minimise the distance between themselves and the observations for any of the possible p_0 configurations in the multimodal posterior on the initial momentum.

In the following experiments we look at what happens to the posterior distribution on (p_0, ν) as r starts at 1 and steadily decreases. For each value of r, we consider the shape $\Gamma^2(r)$ as described above, and place 10^3 observation points along the shape, evenly spaced in terms of arc length. A large number of observations are used so that the data can resolve small lengthscales. We use these observation points as our data, with no noise added, to isolate the cause of the multimodality

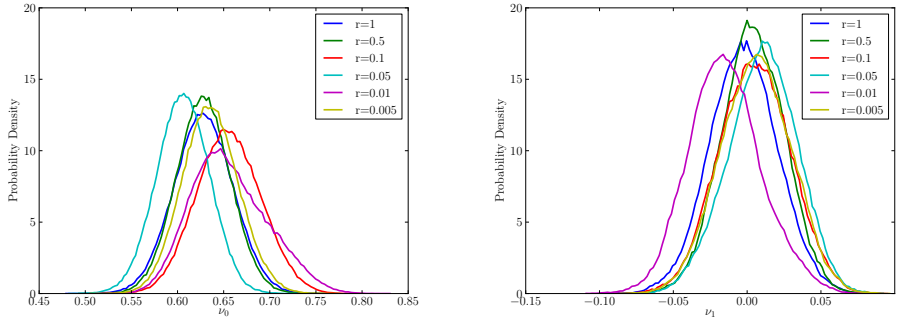
purely to the lengthscale of the features in the data. For the function $\Phi = \frac{1}{2}\|\mathcal{G}(p_0, \nu) - y\|^2_\Sigma$ in the likelihood functional, we pick $\Sigma = \sigma^2 I$ with $\sigma = 10^{-2}$.



(a) (b)

Figure 4. *Plot of marginal distributions on (a) the lowest, (b) the second lowest wave number Fourier mode of the initial momentum function p_0, given varying value of r, the radius of the quarter circles in the target shape.*

Figure 4 shows the marginal distributions on the first two Fourier modes of the initial momentum p_0 as the radius of the quarter circles in the data is reduced. Notice that for $r = 1$, where the data is simply a circle, the distributions are smooth, monomodal and bell-shaped. However, as r is reduced, the distributions become increasingly irregular, and in some cases multimodal. Moreover, these later examples take a great deal of time to converge as the posterior distribution is more complex. Simply finding the local maxima of probability density for these examples would not be a good description of the nature of the distribution as a whole, not would it help us in determining the uncertainty in the problem.



(a) (b)

Figure 5. *Plot of marginal distributions on (a) the lowest, (b) the second lowest wave number Fourier mode of the reparameterisation function ν, given varying value of r, the radius of the quarter circles in the target shape.*

Similarly, Figure 5 shows the marginal distributions on the first two Fourier modes of the reparameterisation function ν as the radius of the quarter circles in the data is reduced. Note that the distributions on ν are monomodal in nature, in contrast to those on p in this data scenario.

7.3. Partial observations

Another area of interest to many practitioners is algorithms which try to recover information in the case where our observations of the target shape are only partial. We may only observe part of the shape, or our observations in some regions of the shape may be essentially destroyed by excessive noise. In this case we would expect to find many different possible shapes which could fit with our data equally well, meaning that the uncertainty quantification through full sampling of the Bayesian posterior is informative.

Let us consider the scenario where we have 1000 fairly evenly spaced observations of the target shape, where the template shape is given by (28) and where the functions (p_0, ν) that create the target shape and the positions of the observation points are draws from the prior. Suppose now that the observation noise is no longer i.i.d. around the entire shape, but instead we have two regions with different covariance structure. In one region, covering approximately three quarters of the shape's circumference, the noise is relatively small, distributed as $\mathcal{N}(0, \Gamma)$ with $\Gamma = \sigma_D^2 I$ with $\sigma_D = 0.0001$. In the final quarter of the domain we add noise with much larger variance, with $\sigma_L = 0.1$. This value is so large that most of this data is rendered useless, in terms of extrapolating information about the value of the functions p_0 and ν.

In this data scenario, we must also be careful in our choice of functional $\Phi(p_0, \nu) = \frac{1}{2}\|\mathcal{G}(p_0, \nu) - y\|^2_{\Gamma_{\text{PO}}}$ in the likelihood. If we were to set $\Gamma_{\text{PO}} = \sigma_D^2$ over the whole domain, then we would be attempting to fit the shape very closely to the very noisy observations in the final third, which would cause all manner of problems. On the other hand, if we were to take $\Gamma_{\text{PO}} = \sigma_L^2$, we would be assuming a lot more uncertainty in the majority of the observations than is actually the case. For the numerics that follow we take $\Gamma_{\text{PO}} = \sigma^2_{\text{PO}}$ where $\sigma^2_{\text{PO}} = \frac{3}{4}\sigma_D^2 + \frac{1}{4}\sigma_L^2$ is the average variance of the observational noise over the whole data set.

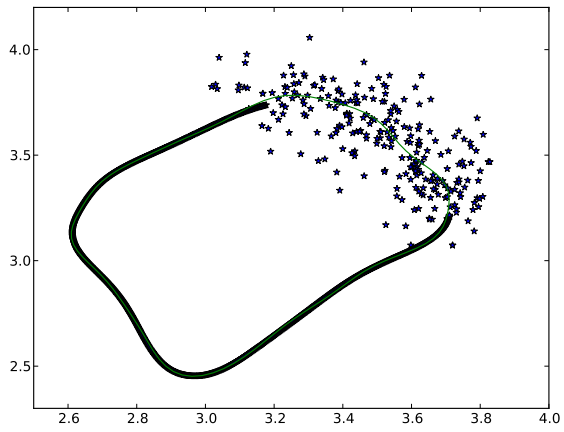
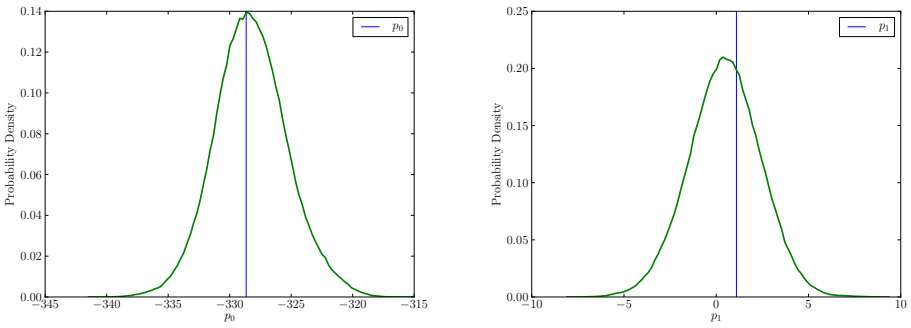


Figure 6. *Plot of partial data, due to large noise variance in one region of the observed shape. Green line denotes the converged solution of the corresponding optimisation problem.*

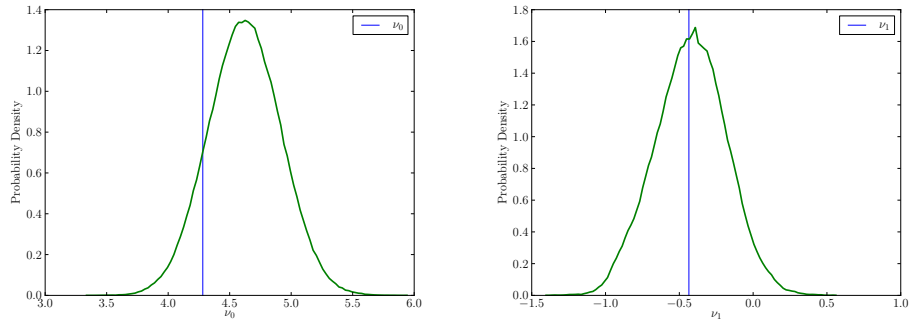
Figure 6 shows an example of such a data scenario. The stars denote data points, and the green curve denotes the converged solution of the corresponding optimisation problem as described in subsection 5.1. However, this single curve does not tell us the whole story, since there are many

curves that could satisfy the data (almost) as well as this curve.



(a) (b)

Figure 7. *Plot of marginal distribution on (a) the lowest, (b) the second lowest wave number Fourier mode of the initial momentum function p_0, given the data set shown in Figure 6.*



(a) (b)

Figure 8. *Plot of marginal distribution on (a) the lowest, (b) the second lowest wave number Fourier mode of the reparameterisation function ν, given the data set shown in Figure 6.*

Figures 7 and 8 show the marginal distributions for the two lowest wave number Fourier modes for the the initial momentum p_0 and reparameterisation function ν respectively. Note that in comparison with the distributions in Figures 1 and 2 with 1000 observations, the variances in these distributions are much higher. With a region with a much higher signal to noise ratio, the uncertainty in the system is much greater, and as such there is a much bigger region of the state space with similar probability density. The approach we have used in this paper allows us to quantify this uncertainty as well as finding states of maximal likelihood.

8. Conclusions

This paper presents a Bayesian approach to shape registration which takes into account the fact that observations of a shape are not exact, and gives a distribution on the shortest distance in shape space between the template shape to the observed shape. We concentrate on the case of a

finite number of observations of points from the target curve, and our approach makes use of the use of explicit reparameterisation variable as described in [11, 24]. The sampling method is framed on function space and so is robust under refinements of discretization [15].

By careful analysis of the forward problem, we have been able to formulate a well-posed Bayesian inverse problem regarding shape matching of a curve with a set of noisy observations of another. We have shown that the likelihood function is Lipschitz continuous on a space which has full measure with respect to a specified choice of Gaussian prior measure. Using this, we have shown how to draw samples from well defined posterior distributions using the RWMH MCMC sampler on function space. This choice of algorithm prevents slower convergence of the statistical algorithm as the discretisation of the functions of interest is refined. We have then implemented this algorithm, and presented briefly some illustrative numerics.

In these numerics, we first showed that the posterior distribution shows consistency as the number of observations of a recoverable shape are increased, and the posterior distributions on the functions of interest become increasingly peaked on the functions which were used to create the data. We also showed an example, where the lengthscale of features in the data is small, in which the posterior distributions exhibit multimodal behaviour, indicating that full characterisation of the posterior can give us more information than the solution of an equivalent optimisation problem. Since the shape space is a Riemannian manifold with regions of positive and negative curvature, there may be more than one geodesic between two points in shape space, which may also lead to multimodality.

Finally we showed that this method can give us a range of different possible solutions in the case of lost or destroyed data for some part of the target shape.

Since we already have at our disposal an implementation of the adjoint problem so that we can calculate the gradient of the observation operator (which we use in the BFGS method for the deterministic burn-in), we can very simply adapt the MCMC method from random walk to a gradient method, such as the Metropolis Adjusted Langevin Algorithm (MALA). This would increase the efficiency of the algorithm markedly. Further analytical results would be needed, however, to ensure that the gradient of the observation operator is continuous on a space which has full measure with respect to an appropriately chosen prior distribution.

Another extension of this work would be to also make the problem translation, rotation and scale invariant, so that any misalignment of the imaging equipment has a negligible effect on the results. This would involve adding parameters into the state space to allow for these types of operations. One might also consider the case where the template shapes themselves are only noisily observed, so that we are trying to find a distribution on the length of geodesic paths in shape space between two noisily observed shapes. The problem could also be extended to the case in which the ordering of the observed points around the curve is not known, in which case the discrete ordering would also have to be learned as part of the inverse problem. This could be done by combining the present approach with a Gibb's sampler for the ordering. This would result in a Bayesian approach to the segmentation process, in which points on the boundary between two materials might first be estimated from a bitmapped image but their ordering is not known with 100% confidence. The result would be a probability distribution on the boundary curve with some prescribed topology.

[1] M. I. Miller, A. Trouvé, and L. Younes. On the metrics, euler equations and normal geodesic image motions of

computational anatomy. In *Proceedings of the 2003 International Conference on Image Processing*, volume 2, pages 635–638. IEEE, 2003.

[2] M. I. Miller, A. Trouvé, and L. Younes. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–222, 2006.

[3] M. Vailliant, M. I. Miller, L. Younes, and A. Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 13:161–169, 2004.

[4] P.T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

[5] S. Allassonnire, Estelle Kuhn, , and Alain Trouv. Construction of bayesian deformable models via stochastic approximation algorithm: A convergence study. *Bernoulli*, 16(3):641–678, 2010.

[6] S. Allassonnire, Estelle Kuhn, and Alain Trouv. Map estimation of statistical deformable template via nonlinear mixed effect models : Deterministic and stochastic approaches. *Proc. of Mathematical Foundations of Computational Anatomy*, 2008.

[7] J. Ma, M.I. Miller, A. Trouv, and L. Younes. Bayesian template estimation in computational anatomy. *NeuroImage*, 42(1):256–261, 2008.

[8] D. D. Holm, J. T. Ratnanather, A. Trouv, and L. Younes. Soliton dynamics in computational anatomy. *NeuroImage*, 23, Supplement 1:S170 – S178, 2004.

[9] J. Glaunes, A. Trouvé, and L. Younes. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub manifolds matching. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[10] Marc Vaillant and Joan Glaunes. Surface matching via currents. In *IPMI*, pages 381–392, 2005.

[11] C.J. Cotter and D.D. Holm. Geodesic boundary value problems with symmetry. *J. Geom. Mech.*, 2:51–68, 2010.

[12] S.L. Cotter, M. Dashti, J.C. Robinson, and A.M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25, 2009.

[13] S. L. Cotter, M. Dashti, and A. M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM J. Numer. Anal.*, 48(1):322–345, 2010.

[14] S. L. Cotter, M. Dashti, and A. M. Stuart. Variational data assimilation using targetted random walks. *International Journal for Numerical Methods in Fluids*, pages n/a–n/a, 2011.

[15] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White. Mcmc methods for functions: modifying old algorithms to make them faster. *In preparation*, 2011.

[16] M. T. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41:61–84, 2001.

[17] J. Glaunes, A. Trouvé, and L. Younes. Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 712–718, 2004.

[18] F. Gay-Balmaz and T.S. Ratiu. Clebsch optimal control in formulation in mechanics. *J. Geom. Mech.*, 3(1), 2011.

[19] B. Khesin and R. Wendt. *Geometry of infinite-dimensional groups*, volume 51 of *Ergebnisse der Mathematik und Grenzgebiete 3.Folge*, chapter 1. Springer-Verlag, 2008.

[20] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, chapter 2. Springer, 2004.

[21] C.G. Broyden. The convergence of a class of double-rank minimization algorithms. *J. Inst. Maths Applics*, 6:76–90, 1970.

[22] D.F. Shanno. On Broyden-Fletcher-Goldfarb-Shanno method. *J. Optimiz. Theory Appl.*, 46, 1985.

[23] C. J. Cotter. The variational particle-mesh method for matching curves. *J. Phys. A: Math. Theor.*, 41:344003, 2008.

[24] C.J. Cotter, A. Clark, and J Peiró. A reparameterisation-based approach to geodesic shooting for curve matching. Submitted, 2011.

Acknowledgements The research leading to these results has received funding from the European Research Council under the *European Community's* Seventh Framework Programme (*FP7/2007-2013*)/ ERC *grant agreement* No. 239870. This publication was based on work supported in part

by Award No KUK-C1-013-04, made by King Abdullah University of Science and Technology (KAUST), and the results were obtained using the Imperial College High Performance Computing Centre cluster. SLC would also like to thank St Cross College Oxford for support via a Junior Research Fellowship.

RECENT REPORTS

37/11	Frequency jumps in the planar vibrations of an elastic beam	Neukirch Frelat Goriely Maurini
38/11	Ice-lens formation and con nement-induced supercooling in soils and other colloidal materials	Style Cocks Peppin Wettlaufer
39/11	An asymptotic theory for the re-equilibration of a micellar surfactant solution	Griffiths Bain Breward Chapman Howell Water
40/11	Higher-order numerical methods for stochastic simulation of chemical reaction systems	Székely Burrage Erban Zygalakis
41/11	On the modelling and simulation of a high pressure shift freezing process	Smith Peppin Ángel M. Ramos
42/11	An efficient implementation of an implicit FEM scheme for fractional-in-space reaction-diffusion equations	Burrage Hale Kay
43/11	Coupling fluid and solute dynamics within the ocular surface tear film: a modelling study of black Line osmolarity	Zubkov Breward Gaffney
44/11	A prototypical model for tensional wrinkling in thin sheets	Davidovitch Schroll Vella Adda-Bedia Cerda
45/11	A fibrocontractive mechanochemical model of dermal wound closure incorporating realistic growth factor	Murphy Hall Maini McCue McElwain
46/11	A two-compartment mechanochemical model of the roles of transforming growth factor β and tissue tension in dermal wound healing	Murphy Hall Maini McCue McElwain
47/11	Effects of demographic noise on the synchronization of a metapopulation in a fluctuating environment	Lai Newby Bressloff
48/11	High order weak methods for stochastic differential equations based on modified equations	Abdulle Cohen Vilmart

51/11	Wound healing angiogenesis: the clinical implications of a simple mathematical model	Du Gunzburger Lehoucq Zhou
52/11	Image Inpainting based on coherence transport with Adapted distance functions	März
53/11	Surface growth kinematics via local curve evolution	Moulton Goriely
54/11	A multiple scales approach to evaporation induced Marangoni convection	Hennessey Münch
55/11	The dynamics of bistable liquid crystal wells	Luo Majumdar Erban
56/11	Real-Time Fluid Effects on Surfaces using the Closest Point Method	Auer Macdonald Treib Schneider Westermann
57/11	Isolating intrinsic noise sources in a stochastic genetic switch	Newby
58/11	Riemann-Cartan Geometry of Nonlinear Dislocation Mechanics	Yavari Goriely
59/11	Helices through 3 or 4 points?	Goriely Neukirch Hausrath

Copies of these, and any other OCCAM reports can be obtained from:

Oxford Centre for Collaborative Applied Mathematics
Mathematical Institute
24 - 29 St Giles'
Oxford
OX1 3LB
England
www.maths.ox.ac.uk/occam