CSGB
CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

www.csgb.dk

RESEARCH REPORT

2017

A. Arnaudon, D.D. Holm, A. Pai and S. Sommer

A Stochastic Large Deformation Model
for Computational Anatomy

A Stochastic Large Deformation Model for Computational Anatomy

A. Arnaudon[1], D.D. Holm[1], A. Pai[2] and S. Sommer[2]

[1]Dept. of Mathematics Imperial College London, UK
[2]Department of Computer Science (DIKU), University of Copenhagen, Denmark

Abstract

In the study of shapes of human organs using computational anatomy, variations are found to arise from inter-subject anatomical differences, disease-specific effects, and measurement noise. This paper introduces a stochastic model for incorporating random variations into the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework. By accounting for randomness in a particular setup which is crafted to fit the geometrical properties of LDDMM, we formulate the template estimation problem for landmarks with noise and give two methods for efficiently estimating the parameters of the noise fields from a prescribed data set. One method directly approximates the time evolution of the variance of each landmark by a finite set of differential equations, and the other is based on an Expectation-Maximisation algorithm. In the second method, the evaluation of the data likelihood is achieved without registering the landmarks, by applying bridge sampling using a stochastically perturbed version of the large deformation gradient flow algorithm. The method and the estimation algorithms are experimentally validated on synthetic examples and shape data of human corpora callosa.

Keywords: computational anatomy, stochastic processes, large deformations, LDDMM

1   Introduction

Computational anatomy (CA) concerns the modelling and computational analysis of shapes of human organs. In CA, observed shapes or images of shapes exhibit variations due to multiple factors, such as inter-subject anatomical differences, disease-specific effects, and measurement noise. These variations in anatomy occur naturally over all populations and they must be handled in cross-sectional studies. In addition, neurodegenerative diseases such as Alzheimer's disease can cause temporal shape changes as the disease progresses. Finally, image acquisition and processing algorithms for extracting shape information can cause measurement variation in the estimated shapes. While variation can be incorporated into shape models via different approaches, e.g. via random initial conditions as in the random orbit model [12]
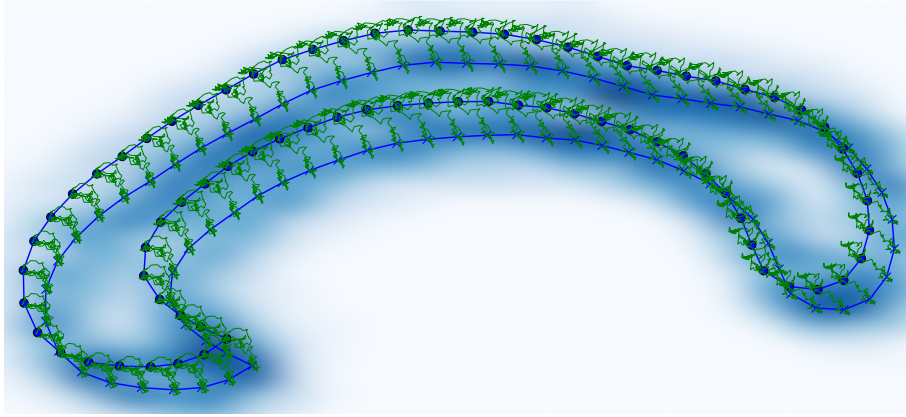
Figure 1: Sample from bridge process (green lines) conditioned on two corpus callosum shapes (blue). The landmarks are influenced by noise with spatial correlation structure determined by the noise fields σ_l. The bridge is constructed by integrating the perturbed gradient flow described in Section 4.3. The shaded background indicates the shape variations in the transition distribution of the stochastic model.

or in a mixed-effects setting [1], most models specify random variation relative to a base object, usually a template, and they involve some form of linearization. Here we will model these variations by inserting stochasticity directly into the nonlinear dynamics of shape transformations using Large Deformation Diffeomorphic Metric Mapping (LDDMM) [16]. In this paper, we use the approach of [2], based on the stochastic fluid models of [8], to introduce a model for incorporating stochastic variation into the shape deformation paths. This approach is particularly designed to be compatible with four geometric properties of LDDMM. Namely, (1) the noise is right-invariant in the same way as the LDDMM metric is right-invariant. (2) Evolution equations arise as extremal paths for a stochastic variational principle. (3) The Euler-Poincaré (EPDiff) equations in the deterministic case have stochastic versions. (4) Remarkably, the momentum maps arising via reduction by symmetry, and used to reduce dimensionality from the infinite-dimensional diffeomorphism group to finite-dimensional shape spaces, such as landmarks in deterministic LDDMM, still persist in this stochastic geometric setting. On top of that, our method permits a total control on the spatial correlation of the noise, independently of for example the data structure or the number of landmarks.

Plan. After a description of large deformation models in Section 2 and the LDDMM framework, we discuss the model of [2] in the CA context and use it to develop the stochastic equations for the landmark trajectories in Section 3. We then formulate two approaches for estimating the noise fields and initial momentum from prescribed data in Section 4. The first approach is a deterministic approximation of the Fokker-Planck equation with matching of moments of the probability distribution. The second is an Energy-Maximization (EM) algorithm which requires sampling of diffusion bridges to estimate data likelihood by sampling bridges without registering or matching the landmarks. From the bridge simulation scheme, we

obtain (see Fig. 1) a stochastic version of the large deformation gradient flow algorithm, see e.g. [16]. Finally, we validate the approach in experiments on synthetic and real data in Section 5 before.

2 Background

This section briefly reviews large deformation shape modelling and LDDMM, referring to the monograph [16] for full details. A shape, defined as either a subset $s \subset \mathbb{R}^d$, $d = 2, 3$, or an image $I : \mathbb{R}^d \to \mathbb{R}$ can be modified by warping the domain $\mathbb{R}^d$ under the action of a diffeomorphism $\varphi : \mathbb{R}^d \to \mathbb{R}^d$. The diffeomorphism φ naturally acts on the left on a shape s by $\varphi.s = \varphi(s)$. For images, a left-action is obtained by letting $\varphi.I = I \circ \varphi^{-1}$. Changes in shape are then represented by an element $\varphi \in G_V \subset \mathrm{Diff}(\mathbb{R}^d)$ in the diffeomorphism group. The subset G_V is obtained as endpoints of flows

$$\partial_t \varphi_t(x) = v_t \circ \varphi_t(x), \ t \in [0, T] \tag{2.1}$$

where $v_t \in V$ is a time-dependent vector field contained in a space $V \subset \mathfrak{X}(\mathbb{R}^d)$. V is typically a reproducing kernel Hilbert space (RKHS) with inner product $\langle \cdot, \cdot \rangle_V$ defined in terms of a reproducing kernel $K_V : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ and with corresponding momentum operator $L_V : V \to L^2$. Under reasonable assumptions, the RKHS structure defines a Riemannian metric on G_V by right-invariance, i.e. by defining $\langle v_\varphi, w_\varphi \rangle_\varphi = \langle v_\varphi \circ \varphi^{-1}, w_\varphi \circ \varphi^{-1} \rangle_V$ for tangent vectors $v_\varphi, w_\varphi \in T_\varphi G_V$. The corresponding Riemannian metric is

$$d_{G_V}(\varphi, \psi)^2 = \min_{v_t, \varphi_0 = \varphi, \varphi_1 = \psi} \int_0^T \|v_t\|_{\varphi_t}^2 dt \tag{2.2}$$

where φ_t evolves according to (2.1). Elements of G_V act on shapes, and d_{G_V} descends to shape spaces by $d_S(s_1, s_2) = \min_{\varphi.s_1 = s_2} d_{G_V}(\mathrm{Id}, \varphi)$ and similarly for images. Shapes can be matched by finding a minimising φ in this equation. In inexact matching of shapes, the distance is weighted against a dissimilarity measure. For images, the inexact matching problem becomes $\varphi_{\min} = \mathrm{argmin}\, E(I_1, I_2, \varphi)$, where $E(I_1, I_2, \varphi) = d_{G_V}(\mathrm{Id}, \varphi)^2 + \lambda S(\varphi.I_1, I_2)$. Here $\lambda > 0$ controls the dissimilarity penalty, and S is the dissimilarity measure, e.g. the L^2 difference $S(I_1, I_2) = \int_{\mathbb{R}^d} |I_1(x) - I_2(x)|^2 dx$. Optimal paths in the LDDMM framework satisfy the EPDiff equation $\partial_t m = -\mathrm{ad}_u^* m$, where $m = L_V u$ and ad^* is the coadjoint action on the dual of the Lie algebra of vector fields on the plane, see [9] for more details.

For a set of N landmarks $\mathbf{q} \in \mathbb{R}^{dN}$ at positions $q_1, \ldots, q_N \in \mathbb{R}^d$ with momentum $\mathbf{p} = (p_1, \cdots, p_N) \in \mathbb{R}^{dN}$, the singular momentum map of [9] is given by $m(x, t) = \sum_i p_i(t) \delta(\mathbf{x} - \mathbf{q}_i(t))$. The corresponding deformation velocity can be written as $u(x, t) = K_V * m = \sum_i p_i K_V(x - q_i)$. The ODE for the positions and momentum are the canonical Hamiltonian equations $\dot{\mathbf{q}}_i = \frac{\partial h}{\partial \mathbf{p}_i}$ and $\dot{\mathbf{p}}_i = -\frac{\partial h}{\partial \mathbf{q}_i}$, with Hamiltonian $h(p, q) = \frac{1}{2} \sum_{i,j=1}^N p_i \cdot p_j K_V(q_i - q_j)$. In 1D, these equations go back to the dynamics of singular solutions of the Camassa-Holm equation [4].

3 Stochastic landmark dynamics

Stochastic differential equations (SDEs) have previously been considered in context of LDDMM by [14, 15, 5]. More recently [11] introduced a stochastic landmark model with dissipation in the momentum equation which corresponds to a Langevin equation. This allowed then to use the technology of statistical mechanics and Gibbs measure to study this stochastic system. We will use here the stochastic model of [2], based on the fluid models of [8]. The noise is introduced in the reconstruction relation (2.1) with J fields $\sigma_l : \mathbb{R}^d \to V$ as

$$d\varphi_t \circ \varphi_t^{-1}(x) = v_t(x)dt + \sum_{l=1}^{J} \sigma_l(x) \circ dW_t^l . \tag{3.1}$$

The symbol d denotes stochastic evolution and $(\circ)$, when adjacent to the process dW_t^l, means stochastic integrals are interpreted in the Stratonovich sense. Given a realisation of the noise, we impose the stochastic dynamics (3.1) and seek the φ_t that minimises the cost functional in (2.2). A short computation gives the stochastic EPDiff equation $dm + \mathrm{ad}_v^* m dt + \sum_i \mathrm{ad}_{\sigma_i}^* m \circ dW_t^i = 0$. The landmarks can be introduced and their stochastic ODE dynamics are given by

$$
\begin{aligned}
d\mathbf{q}_i &= \frac{\partial h}{\partial \mathbf{p}_i}dt + \sum_{l=1}^{J} \sigma_l(\mathbf{q}_i) \circ dW_t^l \\
d\mathbf{p}_i &= -\frac{\partial h}{\partial \mathbf{q}_i}dt - \sum_{l=1}^{J} \frac{\partial}{\partial \mathbf{q}_i}\left(\mathbf{p}_i \cdot \sigma_l(\mathbf{q}_i)\right) \circ dW_t^l .
\end{aligned}
\tag{3.2}
$$

This is a Hamiltonian stochastic process in the sense of Bismut [3], where the stochastic extensions of the Hamiltonian, or stochastic potentials, are the momentum maps $\Phi_l(\mathbf{q}, \mathbf{p}) = \sum_i \mathbf{p}_i \cdot \sigma_l(\mathbf{q}_i)$ whose Hamilton equations generate the cotangent lift of the stochastic infinitesimal transformations of the landmark paths. In the following, we will only consider exact landmark matching, and refer to [2] for more details about the derivation of these equations which also includes the case of inexact matching. The noise can be seen as almost additive in the position equation, and it couples to the momentum with the gradient of the field $\sigma_l(\mathbf{q}_i)$ in the momentum equation. This last term ensures that the solution corresponds to a stochastic path in the diffeomorphism group. By comparison with previous works such as [14, 15, 11], this model has its noise amplitudes σ_l fixed to the domain, and not to each landmark. The noise is relative to a Eulerian frame and is right-invariant similar to the metric. Notice that the present approach is different from random initial conditions such as studied in [17].

3.1 The Fokker-Planck equation

For each time t, we will denote the transition probability density by $\mathbb{P}(\mathbf{q}, \mathbf{p}, t)$, a function $\mathbb{P} : \mathbb{R}^{2dN} \to \mathbb{R}$ which integrates to 1 and represents the probability of finding the stochastic process at a given position in the phase space. For a given initial distribution $\mathbb{P}_0$, it is possible to compute the partial differential equation

which governs the evolution of $\mathbb{P}$ in time. With the Hamiltonian bracket $\{F, G\}_{\text{can}} = \sum_i \frac{\partial F}{\partial q_i} \frac{\partial G}{\partial p_i} - \sum_i \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q_i}$, we compute the Fokker-Planck equation

$$\frac{d}{dt}\mathbb{P} = \{\mathbb{P}, h\}_{\text{can}} + \frac{1}{2} \sum_{l=1}^{J} \{\Phi_l, \{\Phi_l, \mathbb{P}\}_{\text{can}}\}_{\text{can}}. \tag{3.3}$$

The right-hand side of this equation has two parts, the first is an advection with first order derivatives of $\mathbb{P}$, arising from the deterministic Hamiltonian dynamics of the landmarks. The second part is a diffusion term with second order derivatives of $\mathbb{P}$ which arises only from the noise. This is the term which will describe the error in each landmark path subject to noise.

3.2 Equivalent formulation with 'Lagrangian noise'

The covariance between the stochastic displacements $d\mathbf{q}_i$, $d\mathbf{q}_j$ conditioned on the position of two landmarks $\mathbf{q}_i$, $\mathbf{q}_j$ may be computed as

$$\text{Cov}(d\mathbf{q}_i|\mathbf{q}_i, d\mathbf{q}_j|\mathbf{q}_j) = \sum_l \sigma_l(\mathbf{q}_i|\mathbf{q}_i)\sigma_l(\mathbf{q}_j|\mathbf{q}_j)\text{Var}(dW_t^l) ,$$

where $\text{Var}(dW_t^l)$ is the variance of the l-th Brownian motion dW_t^l. Thus, for a finite time discretization of the process in Itô form and an increment $\Delta = [t_1, t_2]$, $\text{Cov}(\Delta\mathbf{q}_i|\mathbf{q}_i, \Delta\mathbf{q}_j|\mathbf{q}_j) = \sum_l \sigma_l(\mathbf{q}_i|\mathbf{q}_i)\sigma_l(\mathbf{q}_j|\mathbf{q}_j)(t_2 - t_1)$. The stochastic differential $d\mathbf{q}$ can therefore formally be viewed as a Gaussian process on $\mathbb{R}^d$. The $dN \times dN$ matrix $\sigma^2(\mathbf{q}) = \sum_l [\sigma_l(\mathbf{q}_i)\sigma_l(\mathbf{q}_j)]_j^i$ is symmetric. If $\sigma^2(\mathbf{q})$ is positive definite, we let $K(\mathbf{q})$ be a square root, i.e. $\sigma^2(\mathbf{q}) = K(\mathbf{q})K(\mathbf{q})^T$. The following stochastic landmark dynamics is then equivalent to the original dynamics (3.2)

$$d\mathbf{q}_i = \frac{\partial h}{\partial \mathbf{p}_i}dt + \sum_{j,\alpha} K(\mathbf{q})_{j,\alpha}^i \circ dW_t^j$$
$$d\mathbf{p}_i = -\frac{\partial h}{\partial \mathbf{q}_i}dt - \sum_{j,\alpha} \frac{\partial}{\partial \mathbf{q}_i} \left(\mathbf{p}_i \cdot K(\mathbf{q})_{j,\alpha}^i \right) \circ dW_t^{j,\alpha} , \tag{3.4}$$

where j runs on the landmarks and α the spatial dimensions. Note that in (3.4), $W_t \in \mathbb{R}^{dN}$ as compared to $\mathbb{R}^J$ previously. With this approach, instead of starting with the spatial noise fields $\sigma_1, \dots, \sigma_J$, we can specify the stochastic system directly in terms of the matrix $K(\mathbf{q})$. One natural choice is to set $[K(\mathbf{q})]_j^i = \text{Id}_d k(\mathbf{q}_i - \mathbf{q}_j)$ for some scalar kernel k and where Id_d is the $d \times d$ identity matrix. The possible reduction in dimensionality of the noise from J to dN has computational benefits that we will exploit in the following.

4 Estimation of Noise and Initial Conditions

We now aim for estimating a set of parameters for the model from N sets of observed landmarks $\mathbf{q}_1, \dots, \mathbf{q}_N$ at time T. The parameters can be both parameters for the

noise fields as described below and the initial landmark configuration $\mathbf{q}_0$ and momentum $\mathbf{p}_0$. The initial configuration $\mathbf{q}_0$ can be considered an estimated template of the dataset. The template will be optimal in the sense of reproducing the moments of the distribution or in being fitted by maximum likelihood. We use spatial noise fields $\sigma_1, \ldots, \sigma_J$ of the form

$$\sigma_l^\alpha(\mathbf{q}_i) = \lambda_l^\alpha k(\|\mathbf{q}_i - \delta_l\|), \tag{4.1}$$

where $\lambda_l \in \mathbb{R}^d$ is the spacial direction of the noise σ_l, δ_l its centre, and the kernel k is either Gaussian $k(x) = e^{-x^2/(2r_l^2)}$ or cubic B-spline $k(x) = S_3(x/r_l)$ with scale r_l. The Gaussian kernels simplify calculations of the moment equations. The B-spline representation has the advantage of providing a partition of unity when used in a grid. The model is not limited to this set of noise functions.

4.1 Matching of Moments

Our first method relies on the Fokker-Planck equation of the landmarks and aims at minimising the cost functional

$$C(\langle \mathbf{p} \rangle (0), \lambda_l) = \frac{1}{\gamma_1} \|\langle \mathbf{q} \rangle_T - \langle \mathbf{q} \rangle (1)\|^2 + \frac{1}{\gamma_2} \|\Delta_2 \langle \mathbf{qq} \rangle_T - \Delta_2 \langle \mathbf{qq} \rangle (1)\|^2, \tag{4.2}$$

where $\langle \mathbf{p} \rangle (0)$ is the initial mean momentum, $\langle \mathbf{q} \rangle (1)$ the final mean position, $\langle \mathbf{q} \rangle_T$ the target sample mean position estimated from the observed landmarks, $\Delta_2 \langle \mathbf{qq} \rangle_T$ the centred sample covariance of the observations, $\Delta_2 \langle \mathbf{qq} \rangle (1)$ the centered final covariance, and γ_1, γ_2 two parameters. The covariances are in components $\Delta_2 \langle q_i^\alpha q_j^\beta \rangle$ and are 2×2 matrices if $i = j$. For the norm, we used a normalised norm for each landmark such that they contribute equally in the sum. More explicitly, we have $\langle q_i^\alpha \rangle (t) := \int q_i^\alpha \mathbb{P}(\mathbf{q}, \mathbf{p}, t) d\mathbf{q} d\mathbf{p}$, and $\Delta_2 \langle q_i^\alpha q_j^\beta \rangle := \langle q_i^\alpha \rangle \langle q_i^\beta \rangle - \langle q_i^\alpha q_j^\beta \rangle$. This cost functional corresponds to the error in matching the mean and covariance of the final probability density $\mathbb{P}(\mathbf{q}, \mathbf{p}, T)$.

It is not possible to solve the Fokker-Planck equation; so we will derive a set of ODEs that approximates the evolution of the mean and covariance by applying the standard cluster expansion method of quantum mechanics to the probability distribution, see [2] for the equations. The expected values of the higher-order products are approximated by only the $\Delta_2 \langle q_i^\alpha q_j^\beta \rangle$ and $\langle q_i^\alpha \rangle$ variables, upon neglecting higher-order correlations such as $\Delta_3 \langle q_i^\alpha q_j^\beta q_k^\beta \rangle$. In order to capture the effect of the noise in the momentum equation, the other correlations such as $\Delta_2 \langle p_i^\alpha q_j^\beta \rangle$ and $\Delta_2 \langle p_i^\alpha p_j^\beta \rangle$ must be taken into account. Their equations of motion are directly computed by taking the derivative in the definition of the correlation. For simplicity, we have used Gaussian kernels for both the Hamiltonian and noise fields so that derivatives have simple forms, and we have approximated $\langle \sigma_l(\mathbf{q}) \rangle \approx \sigma_l(\langle \mathbf{q} \rangle)$. Higher-order corrections of this approximation are possible, but they would not result in significantly better results in practice. To avoid local minima in the minimization of (4.2), we use a global optimisation method based on genetic algorithms, the differential evolution method [13].

4.2 Maximum Likelihood

We derive a method for parameter estimation using maximum likelihood (ML). Upon assuming the landmark observations are independent, the likelihood for the set of unknown parameters θ satisfies $\mathcal{L}_k(\mathbf{q}_1, \ldots, \mathbf{q}_N; \theta) = \prod_{i=1}^{N} \mathbb{P}(\mathbf{q}_i, T; \theta)$ where $\mathbb{P}(\mathbf{q}, T; \theta)$ is the transition density at $\mathbf{q}$ marginalized over $\mathbf{p}$ and with parameters θ. We will denote by $P(\mathbf{q}, \mathbf{p}; \theta)$ the law of the corresponding process. An ML estimate of θ is then $\hat{\theta} \in \operatorname{argmax}_\theta \mathcal{L}(\mathbf{q}_1, \ldots, \mathbf{q}_N; \theta)$. The classic EM algorithm [7] finds parameter estimates θ_l converging to a $\hat{\theta}$ by iterating the following two steps:

(E) Compute expected log-likelihood using parameter estimate θ_{l-1}:

$$Q(\theta|\theta_{l-1}) = \sum_{i=1}^{N} \mathbb{E}_{P_{(\mathbf{q}, \mathbf{p}; \theta_{l-1})}} \left(\log \mathbb{P}(\mathbf{q}, \mathbf{p}; \theta | \mathbf{q}_i) \right). \tag{4.3}$$

(M) Update the parameter estimate $\theta_l = \operatorname{argmax}_\theta Q(\theta|\theta_{l-1})$.

We now describe a method for obtaining an approximation of equation (4.3).

4.3 Stochastically Perturbed Gradient Flow

The landmark trajectory $\mathbf{q}_i(t)$ for $t \in (0, T)$ and momentum $\mathbf{p}_i(t)$ for $t \in (0, T]$ can be considered the missing data for estimating the parameters θ. The conditional expectation in (4.3) amounts to marginalizing out the noise when finding the expected log-likelihood of the data. We approximate this by sampling diffusion bridges, i.e. producing sample paths conditioned on hitting $\mathbf{q}_i$ at time T. In [6, 10], a guidance scheme is constructed that modifies a general diffusion process for a generic variable $\mathbf{x}$ conditioned on hitting a point $\mathbf{v}$ at time T by adding a term of the form $\frac{\mathbf{x}-\mathbf{v}}{T-t} dt$ to the SDE. This scheme allows sample based approximation of $\mathbb{E}_{\mathbf{x}}(f(\mathbf{x})|\mathbf{v})$ for functions f of the stochastic process $\mathbf{x}$ by the formula $\mathbb{E}_{\mathbf{x}}(f(\mathbf{x})|\mathbf{v}) = C_{\mathbf{v}}\mathbb{E}_{\mathbf{y}}(f(\mathbf{y})\phi_{\mathbf{v}}(\mathbf{y}))$ for the modified process $\mathbf{y}$ with a $\mathbf{v}$-dependent constant $C_{\mathbf{v}}$ and a path-dependent correction factor $\phi_{\mathbf{v}}(\mathbf{y})$. We need to modify this scheme for our application because the diffusion field $\Sigma(\mathbf{q}, \mathbf{p})$ in (3.2) may not be invertible as required in the scheme of [6, 10]. A scheme based on the pseudo-inverse $\Sigma^\dagger$ can be derived, but it is computationally infeasible for high-dimensional problems. Instead, we construct the following landmark guidance scheme for the modified variables $\hat{\mathbf{q}}, \hat{\mathbf{p}}$

$$\begin{pmatrix} d\hat{\mathbf{q}} \\ d\hat{\mathbf{p}} \end{pmatrix} = \tilde{b}(\hat{\mathbf{q}}, \hat{\mathbf{p}})dt - \frac{\Sigma^2(\hat{\mathbf{q}}, \hat{\mathbf{p}})(\varphi_{T-t}(\hat{\mathbf{q}}, \hat{\mathbf{p}}) - \mathbf{v})}{T - t}dt + \Sigma(\hat{\mathbf{q}}, \hat{\mathbf{p}})dW. \tag{4.4}$$

$\tilde{b}$ is a bounded approximation of the drift term in (3.2) with the Itô correction; $\varphi_{T-t}(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ is an approximation of the landmark position at time T, given their position at time t; and $\Sigma^2 := \Sigma\Sigma^T$. The boundedness condition together with appropriate conditions on Σ is necessary to show that $\lim_{t \to T} \hat{\mathbf{q}} = \mathbf{v}$ but it is not needed in practice. As in [6], we can compute the correction factor $\phi_{\mathbf{v}}$ for the modified process (4.4). Because of the multiplication on Σ^2, the correction factor does not depend on the inverse or pseudo-inverse of Σ and the scheme is thus computationally much more efficient.

Interestingly, the forcing term of (4.4) is a time-rescaled gradient flow. This can already be seen in the original scheme of [6] by noticing that $(\mathbf{y} - \mathbf{v}) = \nabla_{\mathbf{y}} \frac{1}{2}\|\mathbf{y} - \mathbf{v}\|^2$ has the form of a gradient flow. Here, with appropriate conditions on the noise fields $\Sigma(\mathbf{q}, \mathbf{p})$, we can define an admissible norm $\|\mathbf{v}\|_{\Sigma}^2 := \langle v, \Sigma^2 v \rangle$. The forcing term in (4.4) is then a derivative of a gradient flow for the norm $\|\varphi_{T-t}(\hat{\mathbf{q}}, \hat{\mathbf{p}}) - \mathbf{v}\|_{\Sigma}^2$ in the $\mathbf{q}$ variable. The gradient is taken with respect to the predicted endpoint $\varphi_{T-t}(\hat{\mathbf{q}}, \hat{\mathbf{p}})$ and it couples to the $\mathbf{p}$ variable through Σ^2. The flow is time rescaled with $t \to \frac{1}{2}(T - t)^2$, $dt \to (t - T)\, dt$, $\Sigma \to \sqrt{T - t}\,\Sigma$ and $dW \to \sqrt{T - t}\, dW$. This rescaling slows down the time when the original time t is close to T. This makes sure that the system has enough time to converge and reach the target $\mathbf{v}$. Deterministic gradient flows are used in greedy matching algorithms using the LDDMM right-invariant metric as described in e.g. [16]. With the present stochastic flow, sampling (4.4) allows efficient evaluation of the Q-function in EM and a similar evaluation of the data likelihood. In both cases, no matching or registration of the data is needed.

5 Numerical examples

5.1 Synthetic Test Case

This synthetic test case addresses matching between two ellipses discretized by 5 landmarks. We used a low number of landmarks here in order to have readable pictures. (The algorithms scale well and are not limited by the number of landmarks.) For the noise fields σ_l, we use a grid of 4 by 4 Gaussian noise fields in the region $[-0.4, 0.4]^2$ with a fixed scale $r_l = 0.085$ corresponding to the distance between the grid points. For testing purposes, we let the bottom 8 of the noise fields have larger amplitude λ_l in the x direction, and the top 8 fields have larger amplitudes in the y direction. With these parameters, we produced 5000 sample landmark configurations from the model to estimate the sample mean and covariance of the landmarks at the final simulation time $T = 1$. On the left panel of Figure 2, we display the result of the moment algorithm. In black are shown the density and variance in black ellipses as well as the original fields σ_l with the correct parameters. The estimated parameters λ_l of the noise fields σ_l after running the differential evolution algorithm are shown in blue. The algorithm performs the estimation given only the final sample mean and covariance of the sampled landmarks. The result is in good agreement for the final variances and shows small differences for the estimation of the parameters of the fields σ_l. The differences arise from three sources: the approximation used in the moment evolution; errors in the estimation of sample mean and variance; and the error in the solution of the minimisation algorithm. The minimisation algorithm may not have found the global minimum, but it did converge, as shown on the right panel of Figure 2 where we display the value of the cost for each iteration of the genetic algorithm. Standard derivative-based algorithms would typically be caught in non-optimal local minima and thereby give worse results.
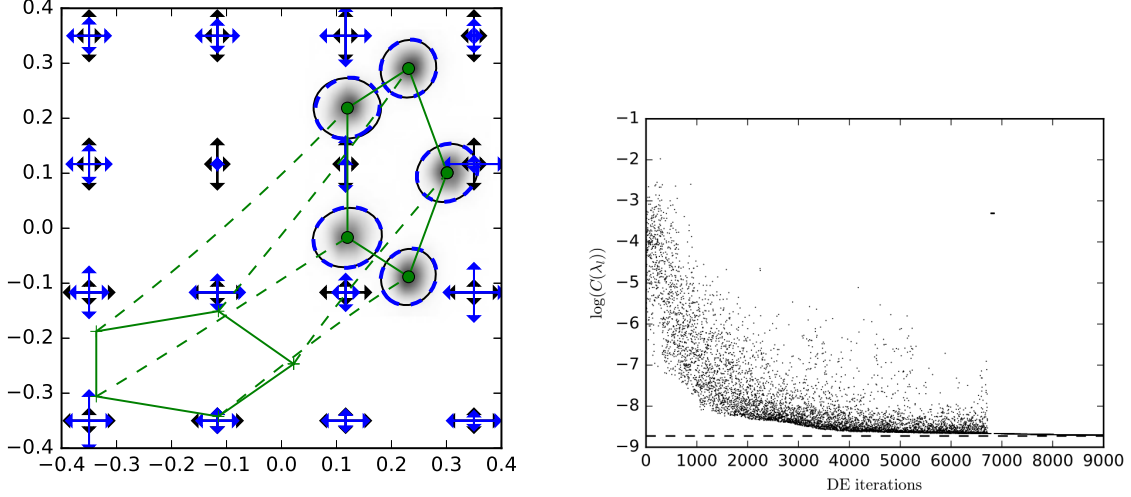
Figure 2: (left) Estimation of the noise amplitude (blue arrows) using only the final variance of the landmark (black ellipses). The landmarks trajectories (dashed green) have initial positions on the bottom left and the final mean position is indicated with green dots. The amplitudes λ_l of the noise fields are represented by arrows (black: original data, blue: estimated data). We used the parameters $\alpha = 0.4$, $r_l = 0.085$, $\Delta t = 0.001$. (right) The convergence of the genetic algorithm which minimises the cost functional. For the last 2000 points we used a gradient descent algorithm to improve the result.
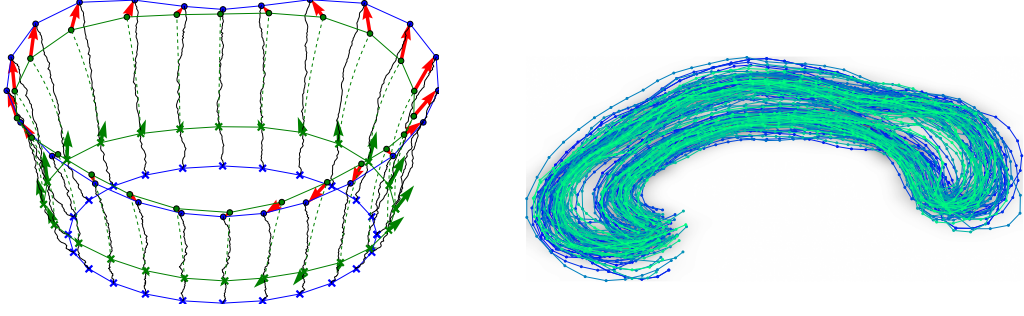
5.2 Stochastic Gradient Flows

We consider matching two ellipses using the stochastically perturbed gradient flow discussed in Section 4.3. In Figure 3 (left), the initial set of landmarks $\mathbf{q}_0$ is displayed along the stochastic path that is pulled by the gradient term to target set $\mathbf{v}$. The guidance is computed from the predicted endpoint $\varphi_{T-t}(\mathbf{q}(t), \mathbf{p}(t))$. The prediction and the guidance term is shown at $t = T/4$. The guidance attraction in (4.4) becomes increasingly strong with time as $t \to T$. This enforces $\mathbf{q}(t) \to \mathbf{v}$.

Compared to the scheme of [6], the use of the function φ_{T-t} to predict the endpoint from $(\mathbf{q}(t), \mathbf{p}(t))$ removes much of the coupling between the momentum $\mathbf{p}$ and the guidance term. Without φ_{T-t}, the scheme would guide based solely on the difference $\mathbf{v} - \mathbf{q}(t)$. However, this would result in the scheme overshooting the target and producing samples of very low probability. The Σ^2 term on the guidance makes the scheme computationally feasible since the inverse of Σ^2 is not needed in the computation of the correction factor.

5.3 Corpus Callosum Variation and Template Estimation

From a dataset of 65 corpus callosum shapes represented by 77 2D landmarks $\mathbf{q}_{i,k}$, $i = 1, \ldots, 77$, $k = 1, \ldots, 65$ evenly distributed along the shape outline, here we aim at estimating the template $\mathbf{q}_0$ and noise correlation represented in this case by a correlation matrix K. The parameters of the model are $\theta = (\mathbf{q}_0, (\lambda_l^0, \lambda_l^1))$. For the MLE, we use the 'Lagrangian' scheme with components of the spatial correlation matrix $k(\mathbf{q}_i, \mathbf{q}_j) = \text{diag}(\lambda_j^0, \lambda_j^1) S_3(\|\mathbf{q}_i - \mathbf{q}_j\|^2 / r)$. The range r is $r = \text{mean}_{ijk} \|\mathbf{q}_{i,k} - \mathbf{q}_{j,k}\|^2$.

9

Figure 3: (left) Stochastic gradient flow with guidance terms. At each time step t with landmarks $\mathbf{q}(t)$ (green line+crosses), a prediction $\varphi_{T-t}(\mathbf{q}(t), \mathbf{p}(t))$ of the endpoint configuration is computed (green line+dots). From the prediction, the difference $\mathbf{v} - \mathbf{q}(t)$ to the target $\mathbf{v}$ (red arrows) is multiplied on Σ^2 and back transported to $\mathbf{q}(t)$ giving the correction term (green arrows). (right) The set of 65 corpora callosa shapes. Landmarks are evenly distributed along the outline of each shape.
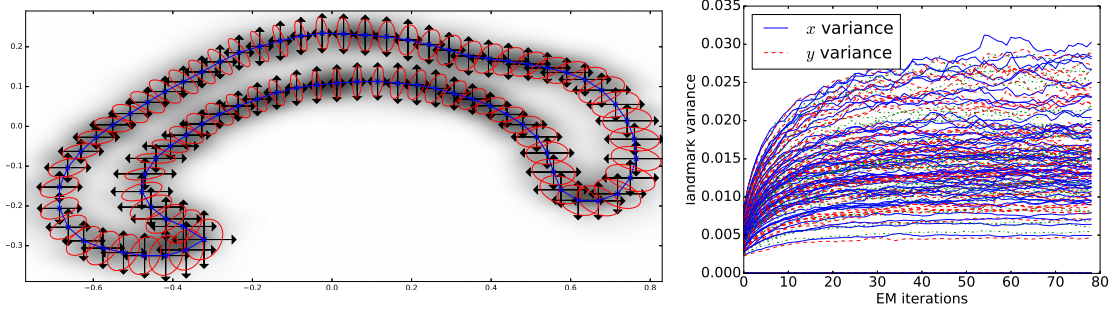
Figure 4: (left) Estimated template $\mathbf{q}_0$ from the corpus callosum shapes with variance magnitude given by arrow length at each landmark. Sample noise covariances of each landmark of the original dataset plotted over each landmark (ellipsis). The background is shaded with a smoothed histogram of the observed landmark positions. (right) Convergence of all variance parameters $(\lambda_l^0, \lambda_l^1)$ is shown as a function of EM iterations.

We initialize $\mathbf{q}_0$ with the Euclidean mean and run the EM algorithm for estimation of θ until convergence. The evolution of the variances $(\lambda_l^0, \lambda_l^1)$ is plotted in the right panel of Figure 4. On the left panel of Figure 4 shows the estimated template $\mathbf{q}_0$. The variance magnitude $(\lambda_l^0, \lambda_l^1)$ is plotted with arrows on each landmark and can be compared with the landmark-wise empirical variance from the dataset. The variance specified by $(\lambda_l^0, \lambda_l^1)$ is axis-aligned, and results in differences when the eigenvalues of the empirical variance are not axis-aligned.

This experiment is repeated in Figure 5 with moment matching algorithm but with several differences of the model. First, the noise fields are the original spatially fixed Gaussian σ_l represented by red arrows. We also allow λ_l to be non-axis aligned and we placed them at the position of every 4 landmarks. We then applied the genetic algorithms using only the landmarks at the same position as the noise fields, fixing the initial momenta to 0 and the initial position to the mean positions of

10
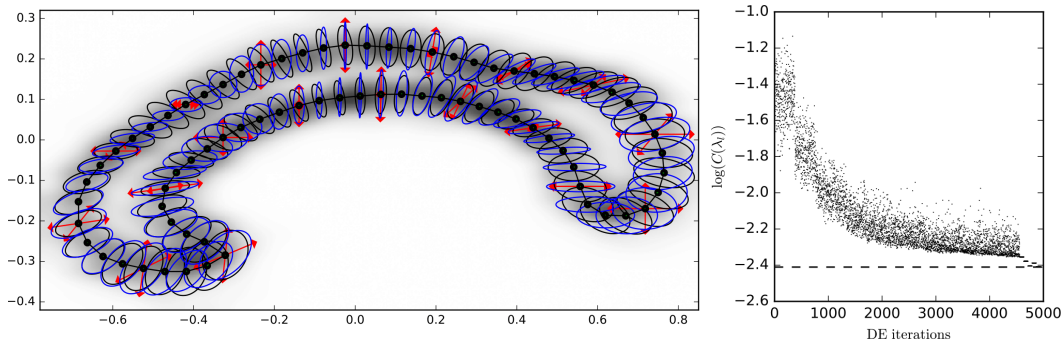
Figure 5: (left) Corpus Callosum estimation with moment matching and setup as in Figure 2 with parameters: $\alpha = 0.1$, $r_l = 0.1$, $\Delta t = 0.01$. We divided by 4 the number of landmarks and placed the σ at the same position. We ran the genetic algorithm to find the best set of parameters λ_l which reproduce the variance for this small set of landmarks. We used all the landmarks to compare the final estimated variance in blue with the observed variance in black. The σ fields are represented by red arrows. (right) Convergence of the differential evolution algorithm, as in figure (2).

the landmarks. In addition to its rapid convergence, as shown in the right panel of Figure 5, the global minimisation algorithm also gave a reliable estimate of the final variance for all the landmarks, even when fewer landmarks and noise fields were used.

6   Conclusion and Open Problems

We have presented and implemented a model for large deformation stochastic variations in computational anatomy. The extremal paths are governed by stochastic EPDiff equations which arise from a right-invariant stochastic variational principle, and admit reduction from the diffeomorphism group to lower dimensional shape spaces (landmarks). We have shown that accurate estimation of the noise fields of the model can be achieved either by approximating the Fokker-Planck equations with a finite set of moments, or by Monte Carlo sampling and EM-estimation. We have derived and implemented the methods in both cases. The second approach introduces the concept of stochastically perturbed gradient flows for data likelihood evaluation. It can be hypothesised that shape evolution of human organs under the influence of diseases do not follow smooth geodesics as in conventional models used in computational anatomy, but rather it exhibits stochastic variations in shape as the disease progresses. The approaches presented enable modelling of such variations. We expect to test this hypothesis on additional shape spaces and with multiple time point longitudinal shape datasets in future work.

References

[1] Allassonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(1), 3–29 (2007)

[2] Arnaudon, A., Holm, D., Sommer, S.: Stochastic Models of Large Deformations. in preparation (2017)

[3] Bismut, J.M.: Mécanique aléatoire. In: Tenth Saint Flour Probability Summer School—1980 (Saint Flour, 1980), Lecture Notes in Math., vol. 929, pp. 1–100. Springer, Berlin-New York (1982)

[4] Camassa, R., Holm, D.D.: An integrable shallow water equation with peaked solitons. Phys. Rev. Lett. 71(11), 1661–1664 (1993)

[5] Cotter, C.J., Cotter, S.L., Vialard, F.X.: Bayesian data assimilation in shape registration. Inverse Problems 29(4), 045011 (2013)

[6] Delyon, B., Hu, Y.: Simulation of conditioned diffusion and application to parameter estimation. Stochastic Processes and their Appl. 116(11), 1660–1675 (2006)

[7] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. JRSS, Series B 39(1), 1–38 (1977)

[8] Holm, D.D.: Variational principles for stochastic fluid dynamics. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 471(2176), 20140963 (2015)

[9] Holm, D.D., Marsden, J.E.: Momentum maps and measure-valued solutions (peakons, filaments, and sheets) for the epdiff equation. In: The breadth of symplectic and Poisson geometry, pp. 203–235. Springer (2005)

[10] Marchand, J.L.: Conditioning diffusions with respect to partial observations. arXiv (2011), arXiv: 1105.1608

[11] Marsland, S., Shardlow, T.: Langevin equations for landmark image registration with uncertainty. arXiv preprint arXiv:1605.09276 (2016)

[12] Miller, M., Banerjee, A., Christensen, G., Joshi, S., Khaneja, N., Grenander, U., Matejic, L.: Statistical methods in computational anatomy. Statistical Methods in Medical Research 6(3), 267–299 (1997)

[13] Storn, R., Price, K.: Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optim. 11(4), 341–359 (1997)

[14] Trouvé, A., Vialard, F.X.: Shape splines and stochastic shape evolutions: a second order point of view. Quarterly of Applied Mathematics 70(2), 219–251 (2012)

[15] Vialard, F.X.: Extension to infinite dimensions of a stochastic second-order model associated with shape splines. Stochastic Processes and their Applications 123(6), 2110–2157 (2013)

[16] Younes, L.: Shapes and Diffeomorphisms. Springer (2010)

[17] Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: International Conference on Information Processing in Medical Imaging. pp. 37–48. Springer (2013)