

## Diffusion Bridges for Stochastic Hamiltonian Systems and Shape Evolutions

Arnaudon, Alexis; van der Meulen, Frank; Schauer, Moritz; Sommer, Stefan

**DOI**

[10.1137/21M1406283](https://doi.org/10.1137/21M1406283)

**Publication date**

2022

**Document Version**

Final published version

**Published in**

SIAM Journal on Imaging Sciences

**Citation (APA)**

Arnaudon, A., van der Meulen, F., Schauer, M., & Sommer, S. (2022). Diffusion Bridges for Stochastic Hamiltonian Systems and Shape Evolutions. *SIAM Journal on Imaging Sciences*, 15(1), 293-323. <https://doi.org/10.1137/21M1406283>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

***Green Open Access added to TU Delft Institutional Repository***

***'You share, we take care!' - Taverne project***

**<https://www.openaccess.nl/en/you-share-we-take-care>**

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

## Diffusion Bridges for Stochastic Hamiltonian Systems and Shape Evolutions\*

Alexis Arnaudon<sup>†</sup>, Frank van der Meulen<sup>‡</sup>, Moritz Schauer<sup>§</sup>, and Stefan Sommer<sup>¶</sup>

**Abstract.** Stochastically evolving geometric systems are studied in shape analysis and computational anatomy for modeling random evolutions of human organ shapes. The notion of geodesic paths between shapes is central to shape analysis and has a natural generalization as diffusion bridges in a stochastic setting. Simulation of such bridges is key to solving inference and registration problems in shape analysis. We demonstrate how to apply state-of-the-art diffusion bridge simulation methods to recently introduced stochastic shape deformation models, thereby substantially expanding the applicability of such models. We exemplify these methods by estimating template shapes from observed shape configurations while simultaneously learning model parameters.

**Key words.** shape analysis, bridge simulation, conditional diffusion, hypoelliptic diffusion, landmark dynamics, guided proposals, shape matching

**AMS subject classifications.** 60J60, 65C05, 62F15

**DOI.** 10.1137/21M1406283

**1. Introduction.** We develop a general scheme for simulation of conditioned diffusion processes—diffusion bridges—for finite-dimensional landmark manifolds appearing in fields including shape analysis and fluid dynamics. We demonstrate several important properties of the scheme, including that the approach applies to more general classes of stochastic models than presently covered in the literature, that it effectively handles challenging aspects of shape spaces, nonlinearity, and partial observability, and that it generalizes the inexact matching scheme that is commonly used in applications to the stochastic setting. Importantly, the simulation scheme allows for statistical inference of properties of the evolution of the shapes. We exemplify this by estimating template shapes from observed shape configurations.

\*Received by the editors March 19, 2021; accepted for publication (in revised form) November 29, 2021; published electronically March 14, 2022.

<https://doi.org/10.1137/21M1406283>

**Funding:** The work of the authors was supported by the CSGB Centre for Stochastic Geometry and Advanced Bioimaging funded by the Villum Foundation (grant 00022924) and by the Novo Nordisk Foundation (grant NNF18OC0052000). The work of the first author was supported by EPSRC grant EP/N014529/1 supporting the EPSRC Centre for Mathematics of Precision Healthcare at Imperial. The work of the second author was supported by the Faculty of Science, University of Copenhagen (UCPH), through the program “Visiting scholars at SCIENCE” as well as by the Mathematics and Computer Science departments, as a substantial part of this research was carried out during a sabbatical leave of the second author at the UCPH.

<sup>†</sup>Department of Mathematics, Imperial College of London, London, SW7 2AZ, UK ([alexis.arnaudon@epfl.ch](mailto:alexis.arnaudon@epfl.ch)).

<sup>‡</sup>Delft Institute of Applied Mathematics, Delft University of Technology, 2628 CD, Delft, The Netherlands ([F.H.vanderMeulen@tudelft.nl](mailto:F.H.vanderMeulen@tudelft.nl)).

<sup>§</sup>Department of Mathematical Sciences, Chalmers University of Technology, University of Gothenburg, Chalmers Tvärgata 3, 41296, Gothenburg, Sweden ([smoritz@chalmers.se](mailto:smoritz@chalmers.se)).

<sup>¶</sup>Department of Computer Science, University of Copenhagen, Copenhagen E, DK-2100, Denmark ([sommer@diku.dk](mailto:sommer@diku.dk)).

**1.1. Background.** The geometry of landmark spaces and the landmark matching problem, exemplified by matching of finite sets of anatomical markers on medical images, have been extensively studied in shape analysis. Examples include the seminal work of Kendall [23] on landmark configurations modulo scaling and rotation, and the extensive interest of landmark dynamics for diffeomorphic shape models [21, 46]. The landmark matching problem has a geometric formulation in the language of geometric mechanics where diffeomorphisms of the underlying domain act to move the landmarks, and optimal trajectories between observed landmark configurations satisfy geodesic equations for a right-invariant metric on the diffeomorphism group [47]. These equations are also called Euler–Poincaré equations [17]. Landmark dynamics are intrinsically linked to soliton dynamics in fluid dynamics [18], e.g., soliton solutions of the Camassa–Holm equations [11], or more generally singular solutions of the Euler–Poincaré equation [16]. Stochastic models of landmark evolutions are therefore of interest in both shape analysis and fluid dynamics.

Several recent models include stochastic effects in landmark analysis [28, 42, 45, 29, 1, 2]. The interest appears for several reasons:

1. If  $v$  denotes the observed landmark configuration, the presence of noise in the observation implies that landmark analysis is classically performed in the inexact setting where the requirement that the flow  $q_t$  satisfy  $q_T = v$  at observation time  $t = T$  is essentially relaxed to  $q_T = v + \epsilon$  for some perturbation  $\epsilon$ . This thus amounts to adding noise at the endpoint  $q_T$  of the flow  $q_t$ . If the perturbation is not linked to observation noise but instead comes from intrinsic stochastic behavior of the shape or soliton evolution, it is more natural to make the noise intrinsically time-continuous, i.e., added to  $q_t$  for each  $t$  on an infinitesimal level.
2. The flow  $q_t$  can have multiscale behavior where the coarse-scale evolution is modeled by a deterministic flow, while the fine-scale behavior, which in a fluid system is closer to turbulence, can be summarized with stochastic terms.
3. In medical applications, it can be hypothesized that the shape evolution is not purely deterministic, and that the stochastic deformation happen continuously in time.
4. It is generally hard to construct natural families of probability distributions on geometric spaces without linearizing the spaces around a center point. Solutions to stochastic differential equations, however, provide such probability models [39].

In this paper, we focus on models defined by stochastic differential equations. In all of the above enumerated cases, it is common that observations are discrete in time. Constructing paths that are conditioned on such partial-in-time observations is called *bridge simulation* and the paths *bridges*. Bridge simulation is essential for likelihood-based inference as it provides a stochastic method for approximating the intractable likelihood. If possible, it enables methods such as sequential Monte Carlo, Markov chain Monte Carlo, stochastic expectation-maximization, or stochastic gradient descent. This in turn allows for estimation of parameters in the processes' dynamics and recovery of latent states at observation times. Moreover, in applications like landmark matching, there is intrinsic interest in bridge simulation as bridges appear as stochastic perturbations of landmark geodesics or as critical paths for a stochastically perturbed energy. Bridges therefore present a stochastic generalization of the geodesic paths usually considered when matching sets of landmarks.

Bridge simulation is generally a challenging problem. It has been addressed in the Euclid-

ean setting in a series of works, including [12, 7, 34, 9, 38]. For shape manifolds, previous papers have centered on simulation in landmark spaces. Manifolds of finite numbers of landmarks are generally of finite dimensionality, in contrast to spaces of continuous shapes, which are in essence infinite-dimensional. While bridge sampling on infinite-dimensional shape spaces remains an open question, bridge sampling for landmark spaces has been considered recently [1, 2, 40] using the approach of [12, 27]. The underlying diffusion arises from stochastic Euler–Poincaré equations [15] and, in a different setting, Brownian motion on the landmark space [40]. The difficulty in landmark bridge sampling arises from the high-dimensionality, the curvature and hence nonlinearity of the diffusion process, from the degenerate diffusion processes in the stochastic Euler–Poincaré case (vanishing eigenvalues in the diffusion matrix), and from observations being only in the position variable of the phase-space flow. The methods from [12] and [1, 2] apply only for certain classes of stochastic evolution, excluding other important stochastic landmark approaches such as proposed in [42, 29], due to the degeneracy of the flow. As stated in [29]: “Unfortunately, none of the known methods for diffusion bridges works with (2.1) to give computationally convenient algorithms. Without an efficient method for sampling the diffusion bridge, it is hard to formulate an MCMC method with good acceptance rates. Consequently, the generalised Langevin prior distribution is difficult to use in Bayesian statistics, and we now turn to simpler prior distributions, which arise by approximating the Langevin equation.” In the present paper, we show that resorting to such simpler prior distributions is unnecessary.

**1.2. Relation to image registration.** Landmark matching in the large deformation diffeomorphic metric mapping (LDDMM [46]) setting is closely related to image registration, both because landmarks often arise from annotations of images and because the LDDMM model allows one to generate deformations of the full image domain from matching of only a small number of landmarks. As such, landmark matching can partly be considered a reduced problem compared to matching of full images. In this paper, we solve a different, but related, problem compared to image registration: we aim to make possible statistical inference of properties of the evolution of shapes, specifically in the form of landmark configurations. The aim is thus not to find a good diffeomorphic matching between two images or shapes, but to infer statistical properties of stochastic transformations between shapes. Nevertheless, some of the stochastic models considered in the paper, particularly the Eulerian model, allows one to generate full domain deformations, as is the case for LDDMM landmark matching (this is called *lifting* of the trajectory from the landmark space to the space of diffeomorphisms). One can therefore envision making bridge sampling with data sizes—full images or large numbers of landmarks—comparable to what is currently used in image registrations algorithms. While this will be extremely interesting, it is also very challenging both from a computational perspective and theoretically due to the high or even infinite dimensionality of the resulting shape space. In the present paper, we aim to present one of the first applications of bridge sampling in shape analysis that nevertheless applies bridge sampling techniques to stochastic processes of much higher dimensionality than what is regularly seen in other areas of statistics. We will pursue extension to deformations of the full domain in future work.

**1.3. Contribution.** We extend the Euclidean simulation approach of [38, 8] to a general method for simulating stochastic landmark equations. We will pursue this under the stochastic

models for landmark and soliton dynamics introduced in [42], [29], and [1]. The presented method is the first in the literature to allow bridge sampling for the former schemes, and it improves upon previous sampling schemes by better incorporating nonlinearity in the drift and diffusion coefficients of the SDEs. For models with intrinsic noise, it is the first approach that provides full uncertainty quantification via samples of the posterior. The approach generalizes the notion of inexact shape matching as in [46]. We demonstrate how the sampling approach can be used to estimate the starting conditions of the diffusion and parameters of the diffusion dynamics. Particularly, we apply it for the problem of estimating a template shape given observed shape configurations. In this case, we show how using the Riemannian Manifold Metropolis Adjusted Langevin Algorithm (RMMALA) is crucial in obtaining efficient updates of the estimated template.

**1.4. Outline.** The paper starts in section 2 with a review of finite-dimensional shape analysis and its stochastic extensions. In section 3 we explain how guided proposals can be used for bridge simulation. Their specific implementation for the two stochastic landmark models considered in this work is specified in section 4.2. Markov chain Monte Carlo sampling schemes are detailed in section 5, and experimental evaluation on synthetic and empirical datasets are given in section 6. We end with a discussion section with suggestions for future work.

**2. Finite-dimensional shape analysis.** To provide the necessary background for the stochastic landmark models, we provide here a short review of landmark shape spaces with geometry inherited from a right-invariant metric on the diffeomorphism group. Landmark shape spaces are finite-dimensional in contrast to spaces of continuous curves and surfaces that are inherently infinite-dimensional. On a formal level, the geometric setting as outlined below is, however, equivalent. This makes the landmark case an example of how stochastics and bridge simulation can be used in a geometric setting, while we can postpone the intricacies of infinite dimensionality in other shape spaces to later works.

Shape analysis as pursued in the large deformation diffeomorphic metric mapping (LD-DMM [46]) framework starts with actions of the diffeomorphism group on shapes spaces. For landmarks, let  $q = (q_1, \dots, q_n)$  be a configuration of  $n$  distinct landmarks  $q_i \in \Omega$  in a domain  $\Omega \subset \mathbb{R}^d$ . Let  $\varphi$  be an element of the set  $\text{Diff}(\Omega)$  of diffeomorphism on  $\Omega$ , smooth invertible mappings with smooth inverses. Then  $\varphi$  acts on  $q$  by composition  $\varphi.q = (\varphi(q_1), \dots, \varphi(q_n))$ . For fixed  $q$ , the map  $\pi: \text{Diff}(\Omega) \rightarrow \mathcal{M}$ ,  $\pi(\varphi) = \varphi.q$  is denoted the action map.

The landmark space  $\mathcal{M} = \{(q_1, \dots, q_n) \in \Omega^n \mid q_i \neq q_j, i \neq j\}$  can be given the structure of a manifold by letting it inherit the differentiable structure from its embedding as an open subset of  $\mathbb{R}^{nd}$  ( $\mathbb{R}^{nd}$  except for the subset of points where landmark pairs coincide). It can furthermore be equipped with a Riemannian metric, which in turn defines the length and energy of path in the landmark space, and from this the matching energy (2.4) used below is defined. This happens with the following geometric structure.

(1) Let  $\mathcal{X}(\Omega)$  be the space of vector fields on  $\Omega$ . We equip subsets  $V$  of  $\mathcal{X}(\Omega)$  with an inner product using a reproducing kernel Hilbert space (RKHS) structure: For  $q_1, \dots, q_n \in \Omega$  let  $V$  be the completion of the set of vector fields on the form  $\sum_{i=1}^n K(\cdot, q_i)a_i$ ,  $a_i \in \mathbb{R}^d$ . Here  $K: \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$  is a matrix-valued map denoted a kernel.  $K$  is required to be symmetric

and positive definite in the sense that

$$\sum_{i,j=1}^n a_i^T K(q_i, q_j) b_j > 0 \quad \forall a_i, b_j \in \mathbb{R}^d, q_i, q_j \in \Omega, n \in \mathbb{N}.$$

In this case,  $V$  receives an inner product by completing the inner product

$$(2.1) \quad \langle K(\cdot, q_i) a_i, K(\cdot, q_j) b_j \rangle_V = \sum_{i,j=1}^n a_i^T K(q_i, q_j) b_j$$

to all of  $V$  by continuity. This makes  $V$  a Hilbert space.

(2) Since  $\mathcal{M} \subset \mathbb{R}^{nd}$ , tangent vectors  $v \in T\mathcal{M}$  can be represented as vectors in  $\mathbb{R}^{nd}$ . Let  $q \in \mathcal{M}$  be a landmark configuration. The LDDMM Riemannian metric is then the inner product

$$(2.2) \quad \langle v, w \rangle_q = \sum_{i,j=1}^n v_i^T K(q_i, q_j)^{-1} w_j$$

between vectors  $v, w \in T_q\mathcal{M}$ . Landmark dynamics are often described in terms of momenta, covectors in the dual bundle  $T^*\mathcal{M}$ . The corresponding inner product between momentum vectors  $a, b \in T_q^*\mathcal{M}$  (the cometric) is then

$$(2.3) \quad \langle a, b \rangle_q = \sum_{i,j=1}^n a_i^T K(q_i, q_j) b_j,$$

i.e., the cometric coincides with the inner product (2.1) on  $V$ .

(3) The Riemannian structure on  $\mathcal{M}$  described here has its geometric origin in the following facts: The tangent space  $T_{\text{Id}} \text{Diff}(\Omega)$  of the Lie group  $\text{Diff}(\Omega)$  is diffeomorphic to  $\mathcal{X}(\Omega)$ , and the Riemannian metric (2.2) is the Riemannian metric on  $\mathcal{M}$  that makes the action map  $\pi$  above a Riemannian submersion with respect to the right-invariant (invariant to the composition of diffeomorphisms on the right) Riemannian metric on  $\text{Diff}(\Omega)$  that the inner product (2.1) on  $V \subset \mathcal{X}(\Omega)$  specifies. In the current context, the main implication of this geometric argument is that the metric (2.2) and cometric (2.3) have their definitions rooted on a well-defined geometric structure.

The Riemannian metric defines the energy of a path  $\mathbf{q} = (q_t, t \in [0, T])$  of landmark configurations by

$$E(\mathbf{q}) = \int_0^T \langle \dot{q}_s, \dot{q}_s \rangle_q ds.$$

This allows matching of landmark configurations  $q$  and  $v$  by searching for a minimal energy path  $\mathbf{q}, q_t \in \mathcal{M}$  such that  $q_0 = q$  and  $q_T = v$ ,

$$(2.4) \quad \underset{\mathbf{q}, q_0=q, q_T=v}{\operatorname{argmin}} E(\mathbf{q}),$$

as pursued below. Paths realizing this energy are geodesics on  $\mathcal{M}$ .

**2.1. Inverse problem of landmark matching.** The interest in landmark matching arises from the case where a fixed sets of meaningful landmarks, for example, anatomical markers in medical images, can be identified. Shapes can then be analyzed by identifying differences between configurations of markers, in the medical case either between subjects or intrasubject when consecutive images are acquired over time. Interestingly, the LDDMM framework allows matching of landmarks as well as curves, surfaces, and even images on  $\Omega$  via the same framework with a metric on  $\text{Diff}(\Omega)$  descending to the particular shape spaces. However, we will restrict ourselves to the landmark case here for two primary reasons: The landmark manifold  $\mathcal{M}$  is finite-dimensional, and it has a natural embedding in  $\mathbb{R}^{nd}$  allowing us to write landmark dynamics in Euclidean coordinates.

Assume two landmark configurations  $q$  and  $v$  are given. A matching consists in solving the variational problem

$$(2.5) \quad \operatorname{argmin}_{\mathbf{q}} E(\mathbf{q}) \quad \text{such that} \quad q_0 = q, q_1 = v,$$

with the energy  $E$  from (2.4). This is known as exact matching and results in a  $t$ -dependent transformation of the configuration  $q$  to the configuration  $v$ . Due to the uncertainty often present in observations of landmarks, e.g., with manual annotations on medical images, inexact matching is instead often used in practice. The variational problem is now

$$(2.6) \quad \operatorname{argmin}_{\mathbf{q}} E(\mathbf{q}) + S(q_1, v) \quad \text{such that} \quad q_0 = q,$$

with the added term  $S$  measuring the dissimilarity between the two landmark configurations  $q_1$  and  $v$ . This term is often taken to be of the form  $S(q_1, v) = n^{-1} \|q_1 - v\|^2$  using the norm from the embedding of  $\mathcal{M}$  in  $\mathbb{R}^{nd}$ . Intuitively, we can think of Gaussian noise on the observations in  $\mathbb{R}^{nd}$  and  $S$  being proportional to a log-likelihood.

Note that the  $t$ -dependence of solutions to (2.5) and (2.6) allows us to think of  $q_1$  as an observation of landmark trajectories  $q_t$  at time  $t = 1$ . This will in particular relate to the conditioning of stochastic processes later, where we assume a fixed observation time  $T > 0$ . In the matching case, the notation implies  $T = 1$  without loss of generality.

**2.2. Deterministic landmark dynamics.** The energy (2.4) has a Hamiltonian formulation which involves the kernel  $K$ , where the Hamiltonian is

$$(2.7) \quad H(q, p) = \frac{1}{2} \langle p, p \rangle_q = \frac{1}{2} \sum_{i,j=1}^n p_i^T K(q_i, q_j) p_j.$$

Let  $x \in \mathbb{R}^{2dn}$  be the vector obtained by concatenating the tuples  $\{(q_i, p_i)\}_{i=1}^n$ . That is,  $x$  represents an element of the phase space  $T^*\mathcal{M}$  of the landmark space. Let  $K$  be a kernel as discussed above, and we assume it is scalar, i.e., of the form  $K(y) = k(y) \text{Id}_d$  for a real-valued function  $k$ . Hamilton's equations of motion are then explicitly given by

$$(2.8) \quad \begin{aligned} \frac{d}{dt} q_i &= \frac{\partial H}{\partial p_i} = \sum_{j=1}^n p_j k(q_i - q_j), \\ \frac{d}{dt} p_i &= -\frac{\partial H}{\partial q_i} = -\sum_{j=1}^n \langle p_i, p_j \rangle \nabla k(q_i - q_j), \end{aligned}$$



where  $p$  is the momentum of the flow and  $\nabla k(y)$  the gradient of  $k$ .

Solutions  $\mathbf{q}$  to (2.5) and (2.6) appear as the  $q$ -variable of solutions  $(\mathbf{q}, \mathbf{p})$  to (2.8). Therefore, the search space for optimal solutions of the optimization problems can be reduced to solutions of Hamilton’s equations, a reduction from an infinite to a finite-dimensional space.

**2.3. Stochastic landmark dynamics.** As already indicated, there are several reasons to introduce stochastic evolutions. We review here two different models of stochastic evolutions on landmarks which we will consider later. These are formulated in terms of stochastic differential equations. A good introduction to such equations is given in [24], whereas more advanced treatments are, for example, [33], [22], and [37].

**2.3.1. Stochastic forcing: Lagrangian and Langevin models.** The first stochastic perturbation of the dynamics landmarks described above has been introduced in [42, 45] and consists in an additive Lagrangian noise of the form

$$(2.9) \quad dq_i^\alpha = \frac{\partial H}{\partial p_i^\alpha} dt \quad \text{and} \quad dp_i^\alpha = -\frac{\partial H}{\partial q_i^\alpha} dt + \gamma_i dW_t^i,$$

where  $\gamma_i \in \mathbb{R}$  is a scalar noise amplitude given for each landmark. We will refer to this system as the *Lagrangian model*.

More recently, [29] generalized the Lagrangian model by adding a dissipative term to

$$(2.10) \quad dp_i^\alpha = -\lambda \frac{\partial H}{\partial p_i^\alpha} dt - \frac{\partial H}{\partial q_i^\alpha} dt + \gamma_i dW_t^i,$$

where  $\lambda > 0$  is a damping coefficient which implies existence of the Gibbs invariant measure for this SDE. This model in the following is referred to as the *Langevin model*.

**2.3.2. Transport noise: Eulerian model.** In addition to the Lagrangian noise of the previous model (where each Wiener process is associated to a landmark), [2] introduced an Eulerian noise, where the noise fields are functions of the domain  $\Omega$ . In general, any set of functions could be used, but in practice we will fix a family of  $J$  noise fields  $\sigma_1, \dots, \sigma_J$ , which are kernel functions centered at locations  $\{\delta_\ell\}$  of the form

$$(2.11) \quad \sigma_\ell^\alpha(q) = \gamma_\alpha \bar{k}_\tau(q - \delta_\ell),$$

where  $\bar{k}_\tau$  is a kernel with length-scale  $\tau$  and  $\gamma \in \mathbb{R}^d$  is the vector of noise amplitudes. Note that it is possible to choose both  $\gamma$  and  $\tau$  dependent on the location  $\delta_\ell$ , though we will not use this flexibility here.

The stochastic dynamics are then obtained as in the deterministic case to be a Hamiltonian system but in both the drift and the noise. It reads, in Stratonovich form,

$$(2.12) \quad \begin{aligned} dq_i &= \frac{\partial H}{\partial p_i} dt + \sum_{l=1}^J \sigma_l(q_i) \circ dW_t^l, \\ dp_i &= -\frac{\partial H}{\partial q_i} dt - \sum_{l=1}^J \frac{\partial}{\partial q_i} (p_i \cdot \sigma_l(q_i)) \circ dW_t^l. \end{aligned}$$

We will refer to this system as the *Eulerian model*.

**3. Bridge dynamics.** All models of the previous section can be written in the general Itô form

$$(3.1) \quad dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = x_0, \quad t \in [0, T],$$

where  $b: [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}^N$  and  $\sigma: [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}^{N \times N'}$  are the drift and diffusion coefficient, respectively.  $W_t$  is a Wiener process in dimension  $\mathbb{R}^{N'}$  with independent components, and the stochastic integrals are Itô.

In its most basic form, the previous problems correspond to simulating trajectories of  $X := (X_t, t \in [0, T])$ , where both  $X_0$  and  $X_T$  are partially observed. We will encode the partial observation with two possibly different projection operators,  $L_0$  and  $L_T$ , such that only the vectors  $L_0 X_0$  and  $L_T X_0$  are observed. As an example, if only landmark positions are observed at times 0 and  $T$ , then  $L_0 = L_T$  and  $L_0$  is such that  $L_0 X_t$  extracts only those components of the vector  $X_t$  that correspond to landmark positions, thus ignoring the momentum components.

To model *inexact matching*, the observed vectors have extrinsic noise, and the realizations of random vectors

$$(3.2) \quad V_0 \sim N(L_0 X_0, \Sigma_0) \quad \text{and} \quad V_T \sim N(L_T X_T, \Sigma_T)$$

are observed instead of  $L_0 X_0$  and  $L_T X_T$ . Notice that the *exact matching* limit is obtained upon taking  $\Sigma = \varepsilon^2 I$  and  $\varepsilon \rightarrow 0$ .

Simulating  $X$  conditional on  $(V_0, V_T)$  is challenging for several reasons:

1. Contrary to many papers on diffusion bridge simulations (e.g., [12, 7, 34, 9, 38]), the conditioning is not on the full state, but on a subspace given by  $L_0$  and  $L_T$ ;
2. the diffusion coefficient  $\sigma$  can be state-dependent, such as in the Eulerian model;
3. the dimension of the driving Brownian motion  $W$  may not equal the dimension of the state space of the diffusion; and
4. the regularity of the paths of landmark locations and momenta may not be the same (hypoellipticity), such as in the Lagrangian model.

We demonstrate how these challenges can be tackled with the approach of [38, 43, 8, 30], where new methods for simulating conditioned diffusions using *guided proposals* were presented. In addition to these works, the current setting needs special attention to obtain efficient numerical methods, due to the high dimension of the state space for a large number of landmarks (given as  $N = 2dn$ , where  $d$  and  $n$  denote the dimension and number of landmarks landmarks, respectively) and to the nonlinear nature of the problem (mostly in choice of parameters and a particular “auxiliary process,” described in section 4.2). Also, the structure of the landmarks configuration makes naive simple random-walk-type Metropolis–Hastings schemes practically useless. As we will show, adaptations that take the geometry properly into account, such as the Riemannian Adjusted Metropolis Adjusted Langevin Algorithm, can be exploited instead.

**3.1. Conditioned diffusions.** Throughout, we will assume that  $X$  admits smooth transition densities  $p$ , such that  $\mathbb{P}(X_t \in dy \mid X_s = x) = p(s, x; t, dy)$  for  $s < t$ . In this section, we consider the problem of simulating the diffusion process  $X$  conditioned on

$$(3.3) \quad X_0 = x_0 \quad \text{and} \quad V_T = v_T,$$

where  $V_T \sim N(L_T X_T, \Sigma)$  and  $x_0, v_T$  are given vectors. This condition is different from (3.2), as here the initial state is fully observed, which turns out to be an important step in the algorithms that we present.

The diffusion conditioned on the events specified in (3.3) is rather loosely referred to as a bridge process, as it bridges available observations on the state of the process at times 0 and  $T$ .

It is known that bridge processes satisfy an SDE similarly to that of the unconditioned process, but with an additional *guiding* term in the drift. This additional drift term can be obtained using Doob’s  $h$ -transform (see [37] or the theory of enlargement of initial filtration [20, 5]) to yield the bridge process  $X^* = (X_t^*, t \in [0, T])$  as the strong solution of the SDE

$$(3.4) \quad dX_t^* = b(t, X_t^*) dt + a(t, X_t^*)r(t, X_t^*) dt + \sigma(t, X_t^*) dW_t, \quad X_0^* = x_0,$$

where  $a = \sigma\sigma'$  (where  $'$  denotes matrix transpose),  $r(t, x) = \nabla_x \log \rho(t, x)$ ,  $t \in [0, T]$ ,  $x \in \mathbb{R}^N$ , and

$$(3.5) \quad \rho(t, x) = \int_{\mathbb{R}^N} p(t, x; T, \xi)\psi(v_T; L_T\xi, \Sigma) d\xi.$$

Here,  $\psi(x; \mu, \Sigma)$  denotes the density of the  $N(\mu, \Sigma)$ -distribution, evaluated at  $x$ , for any  $\mu \in \mathbb{R}^N$  and  $\Sigma \in \mathbb{R}^{N \times N}$ . Although we focus here on inexact matching, we remark that in the case of exact matching (where  $\Sigma \equiv 0$ ) the formula for  $\rho$  is more complex; see, for example, section 1.3.2 in [8].

*Example 1.* To help gain the reader some intuition on the term  $a(t, X_t^*)r(t, X_t^*)$ , consider the setting where  $b \equiv 0$ ,  $\sigma$  is constant, and the diffusion is fully observed at time  $T$  with  $N(0, \varepsilon^2 I)$ -noise. In that case, the SDE for the conditioned process is given by  $dX_t^* = (T + \varepsilon^2 - t)^{-1}(v_T - X_t^*) dt + \sigma dW_t$  (this follows by first computing  $\rho$  in (3.5) and subsequently  $r$ ). Note that for small values of  $\varepsilon$ , the guiding term is roughly “remaining distance to endpoint” divided by “remaining time” and that a value of  $\varepsilon$  that is strictly positive avoids dividing by zero in the drift at time  $T$ .

**3.2. Bayesian formulation for landmark matching.** We will follow the Bayesian paradigm for statistical inference in this paper. Within Bayesian statistics it is common to write the data-generating model in a hierarchical way. This hierarchical scheme (to be read from bottom upwards) is as follows:

$$\begin{aligned} v_T \mid x_T &\sim N(L_T x_T, \Sigma_T) , \\ v_0 \mid x_0 &\sim N(L_0 x_0, \Sigma_0) , \\ x_T \mid x_0 &\sim p(0, x_0; T, x_T) , \\ x_0 &\sim p(x_0) . \end{aligned}$$

Here,  $p(x_0)$  denotes the prior density on the initial state. The observed variables are  $v_0$  and  $v_T$ ; the unobserved variables are  $x_0$  and  $x_T$ . Bayesian inference is based on the distribution of unobserved variables, conditional on observed variables. Hence, using Bayesian notation

$$p(x_0, x_T \mid v_0, v_T) \propto p(v_T \mid x_T)p(v_0 \mid x_0)p(x_T \mid x_0)p(x_0)$$

(note that the third term on the right-hand side is the intractable transition density). In particular, this implies that

$$(3.6) \quad p(x_T | v_0, v_T, x_0) \propto p(v_T | x_T)p(x_T | x_0).$$

As shown in the appendix of [30], the solution of (3.4) with (3.5) implies that  $X_T^*$  has density

$$(3.7) \quad \xi \mapsto \eta(\xi; x_0, v_T) := \frac{p(0, x_0; T, \xi)\psi(v_T; L_T\xi, \Sigma) d\xi}{\int p(0, x_0; T, \xi')\psi(v_T; L_T\xi', \Sigma) d\xi'},$$

but this is just the density in (3.6). This reveals that sampling of  $x_T$ , conditional on  $(v_0, x_0, v_T)$ , can be done by sampling  $X_T^*$ , which in turn is obtained as the endpoint of a sampled bridge path.

#### 4. Guided proposals.

**4.1. Overview of guided proposals.** Since the transition densities of a diffusion are only available in closed form in very special cases,  $\rho$  in (3.5) is intractable. This prevents using a numerical discretization scheme on the SDE for  $X^*$ , where the process would be simulated forward on a fine grid (for example, using Euler's method). The key idea of [38] consists of instead simulating from a process  $X^\circ$  that is obtained after replacing  $p$  by the transition densities  $\tilde{p}$  of an auxiliary diffusion process  $\tilde{X}$  for which  $\tilde{p}$  is tractable. This is naturally the case for linear processes  $\tilde{X}$  whose dynamics are governed by the SDE

$$(4.1) \quad d\tilde{X}_t = \tilde{b}(t, \tilde{X}_t) dt + \tilde{\sigma}(t) dW_t,$$

where  $\tilde{b}$  is of the form

$$(4.2) \quad \tilde{b}(t, x) = \tilde{\beta}(t) + \tilde{B}(t)x.$$

Hence, instead of sampling from  $X^*$ , one samples from  $X^\circ$  defined by

$$(4.3) \quad dX_t^\circ = b(t, X_t^\circ) dt + a(t, X_t^\circ)\tilde{r}(t, X_t^\circ) dt + \sigma(t, X_t^\circ) dW_t, \quad X_0^\circ = x_0,$$

with  $\tilde{r}(t, x) = \nabla_x \log \tilde{\rho}(t, x)$ , where  $\tilde{\rho}$  is derived from  $\tilde{p}$  exactly as  $\rho$  is in terms of  $p$ . We call  $t \mapsto a(t, X_t^\circ)\tilde{r}(t, X_t^\circ)$  the *guiding term*, as it is a term that is superimposed on the drift of the original SDE for  $X$  to satisfy the imposed conditioning at time  $T$ . Draws from this *guided diffusion bridge proposal* can subsequently be accepted/rejected in a Metropolis–Hastings sampler, provided that

1. the law of  $X^*$  is absolutely continuous with respect to the law of  $X^\circ$ ;
2. the Radon–Nikodym derivative showing up in the acceptance probability is tractable.

A short introductory account on the Metropolis–Hastings algorithm is given in section A.

An implicit assumption made throughout is that a strong solution to the SDE (4.3) exists, but, in addition, guided proposals only exist if the auxiliary process is chosen such that  $\tilde{X}$  admits transition densities; see section 2.1 in [8] for more details. Furthermore, a key result from [8] provides sufficient conditions for absolute continuity of  $\mathbb{P}^*$  with respect to  $\mathbb{P}^\circ$ , denoted by  $\mathbb{P}^* \ll \mathbb{P}^\circ$ , and an expression for the Radon–Nikodym derivative

$$(4.4) \quad \frac{d\mathbb{P}^*}{d\mathbb{P}^\circ}(X^\circ) = \frac{\tilde{\rho}(0+, x_0)}{\rho(0+, x_0)} \Psi(X^\circ),$$

Table 1

Summary of the notation of various bridges in this work. The rightmost three columns give the drift, diffusion coefficient, and measure on  $C([0, T], \mathbb{R}^N)$ , respectively.

$X$	original, unconditioned diffusion process, defined by (3.1)	$b$	$\sigma$	$\mathbb{P}$
$X^*$	corresponding bridge, conditioned on $v$ , defined by (3.4)	$b^*$	$\sigma$	$\mathbb{P}^*$
$X^\circ$	proposal process defined by (4.3)	$b^\circ$	$\sigma$	$\mathbb{P}^\circ$
$\tilde{X}$	linear process defined by (4.1) with transition densities $\tilde{p}$ appearing in the definition of $X^\circ$	$\tilde{b}$	$\tilde{\sigma}$	$\tilde{\mathbb{P}}$

with  $\rho$  defined as in (3.5) and

$$(4.5) \quad \Psi(X^\circ) = \exp \left( \int_0^T \mathcal{G}(s, X_s^\circ) ds \right),$$

where

$$(4.6) \quad \begin{aligned} \mathcal{G}(s, x) &= (b(s, x) - \tilde{b}(s, x))' \tilde{r}(s, x) \\ &\quad - \frac{1}{2} \text{tr} \left( [a(s, x) - \tilde{a}(s)] \left[ \tilde{H}(s) - \tilde{r}(s, x) \tilde{r}(s, x)' \right] \right) \end{aligned}$$

and  $\tilde{H}(s)$  is the negative of the Hessian matrix of  $x \mapsto \nabla_x \log \tilde{\rho}(s, x)$ , which turns out to be a constant on  $x$ .

Forward simulation of the guided proposal  $X^\circ$  in (4.3) requires evaluation of  $\tilde{r}$ . It turns out that in all algorithms of section 5 the intractable term  $\rho(0+, x_0)$ , appearing in (4.4) and depending on the transition densities of the process  $X$ , cancels. Ignoring this term, evaluating (4.4) requires evaluation of  $\tilde{r}$ ,  $\tilde{H}$ , and  $\tilde{\rho}(0+, x_0)$ . These quantities can be computed efficiently as they only depend on the tractable process  $\tilde{X}$ . There exist various expressions for this; full details are in [30]. For completeness we recap Theorem 2.4 in [30], which for  $t \in [0, T]$  gives

$$\begin{aligned} \tilde{r}(t, x) &= L(t)' M(t) (v(t) - \mu(t) - L(t)x), \\ \tilde{\rho}(t, x) &= \psi(v(t); \mu(t) + L(t)x, M^\dagger(t)), \\ H(t) &= L(t)' M(t) L(t). \end{aligned}$$

Here  $M(t) = [M^\dagger(t)]^{-1}$ , with  $(L(t), \mu(t), M^\dagger(t))$  being defined as solutions to the (backward) differential equations

$$(4.7) \quad \begin{aligned} dL(t) &= -L(t) \tilde{B}(t) dt, & L(T) &= L_T, \\ dM^\dagger(t) &= -L(t) \tilde{a}(t) L(t)' dt, & M^\dagger(T) &= \Sigma_T, \\ d\mu(t) &= -L(t) \tilde{\beta}(t) dt, & \mu(T) &= 0, \end{aligned}$$

and  $v(t) = v_T$  for  $t \in (0, T]$ .

As we assume that  $V_0 \sim N(L_0 X_0, \Sigma_0)$  (as in (3.2)), it follows again from Theorem 2.4 in [30] that for  $t = 0$

$$(4.8) \quad L(0) = \begin{bmatrix} L_0 \\ L(0+) \end{bmatrix}, \quad M^\dagger(0) = \begin{bmatrix} \Sigma_0 & 0 \\ 0 & M^\dagger(0+) \end{bmatrix}, \quad \mu(0) = \begin{bmatrix} 0 \\ \mu(0+) \end{bmatrix}, \quad v(0) = \begin{bmatrix} v_0 \\ v_T \end{bmatrix}$$

(with  $0$  denoting a vector or matrix of appropriate dimensions). Note that for fixed  $x$ ,  $t \mapsto \tilde{\rho}(t, x)$  is left-continuous.

*Remark 1.* From the discussion in Appendix B on the exact matching case, it is in general preferable to take  $\tilde{b}$  and  $\tilde{\sigma}$  such that

$$(4.9) \quad L_T b(T, X_T^0) = L_T \tilde{b}(T, X_T^0) \quad \text{and} \quad L_T a(T, X_T^0) L_T' = L_T a(T) L_T',$$

which is possible for both the Lagrangian and Eulerian models, as we show in the next section.

**4.2. Guided proposals for landmark models.** Guided proposals are specified by choice of the auxiliary process  $\tilde{X}$ . Here we detail this choice for the three landmark models considered. Recall (see also (2.8))

$$(4.10) \quad \frac{\partial H}{\partial p_i} = \sum_{j=1}^n p_j k(q_i - q_j) \quad \text{and} \quad \frac{\partial H}{\partial q_i} = \sum_{j=1}^n \langle p_i, p_j \rangle \nabla k(q_i - q_j).$$

**4.2.1. Choice of auxiliary process for the Langevin and Lagrangian models.** In the Langevin model we have

$$dq_i^\alpha = \frac{\partial H}{\partial p_i^\alpha} dt \quad \text{and} \quad dp_i^\alpha = -\lambda \frac{\partial H}{\partial p_i^\alpha} dt - \frac{\partial H}{\partial q_i^\alpha} dt + \gamma_i dW_t^i.$$

The matching conditions (4.9) suggest to match both the drift on the landmarks positions and the diffusivity at time  $T$ . In view of (4.10), we take the auxiliary process  $\tilde{X}$  as

$$d\tilde{q}_i^\alpha = \sum_{j=1}^n \tilde{p}_j k(q_i^T - q_j^T) dt \quad \text{and} \quad d\tilde{p}_i^\alpha = -\lambda \sum_{j=1}^n \tilde{p}_j k(q_i^T - q_j^T) dt + \gamma_i dW_t^i,$$

where  $q_i^T$  is the observed  $i$ th landmark position at time  $T$ . Because the kernels are evaluated on the final positions, these equations define a linear process, from which  $\tilde{B}$ ,  $\tilde{\beta}$ , and  $\tilde{\sigma}$  can be inferred directly.

**4.2.2. Choice of auxiliary process for the Eulerian model.** As the matching assumptions are formulated for an SDE in Itô form, we first need to transform the Eulerian model equations from Stratonovich to Itô form. To compute the additional term in the drift, we specify the noise kernels to be located at fixed positions. We suppose at location  $\delta_\ell$  kernel  $q \mapsto \sigma_\ell(q)$  with coordinates as specified in (2.11).

**Proposition 1.** *For the Eulerian model, the additional term in the drift when switching from Stratonovich to Itô form for  $q$  is given by*

$$(4.11) \quad \frac{1}{2} \sum_{\ell} z_\ell(q) \bar{k}_\tau(q - \delta_\ell) \gamma.$$

For  $p$  this term is given by

$$(4.12) \quad \frac{1}{2} \sum_{\ell} \langle p, \gamma \rangle (z_\ell(q) \nabla \bar{k}_\tau(q - \delta_\ell) - \bar{k}_\tau(q - \delta_\ell) \nabla z_\ell(q)).$$

Here, in the notation we have omitted dependence of  $(\tau, \delta, \gamma)$  on  $\ell$  and have denoted

$$z_\ell(q) = \langle \nabla \bar{k}_\tau(q - \delta_\ell), \gamma \rangle.$$

This proposition shows that there is a simple way to take the Stratonovich–Itô correction into account in the drift of the auxiliary process  $\tilde{X}$ : for landmark  $i$  one superimposes the terms (4.11) and (4.12) to the drift of the auxiliary process for the Lagrangian model, with the final positions  $q = q_i^T$  substituted.

**5. MCMC algorithms for landmark matching and template estimation.** Both the drift  $b$  and the diffusion coefficient  $\sigma$  in (3.1) may contain unknown parameters. Suppose the vector of unknown parameters is  $\theta$  and we will write, for example,  $b_\theta$  and  $\sigma_\theta$  to highlight the explicit dependence on these parameters. In this section, we present an MCMC (Markov chain Monte Carlo) algorithm for landmarks matching (section 5.1) and template estimation (section 5.2), including estimation of unknown parameters.

**5.1. Bridging two landmark configurations.** The Bayesian approach to the problem of landmarks matching boils down to sampling from  $(X, \theta)$  conditional on  $(v_0, v_T)$  (the observed two landmarks configurations). It is well known that in the case of unknown parameters in the diffusion coefficient, a Gibbs sampler that successively updates  $\theta \mid (X, v_0, v_T)$  and  $X \mid (\theta, v_0, v_T)$  leads to an invalid (reducible) scheme (see, for example, [36]). To circumvent this problem, we will update the Wiener increments  $W$  instead of the process  $X$  itself. Then, as we have made the assumption that a strong solution to the SDE (4.3) exists, there is a measurable map  $\mathcal{GP}_\theta$  such that  $X^\circ = \mathcal{GP}_\theta(x_0, W)$ , where  $W$  is the driving Wiener process in  $\mathbb{R}^{N'}$  ( $\mathcal{GP}$  being an abbreviation of guided proposal). Note that we could have written  $\mathcal{GP}_\theta(x_0, W, v_T)$  instead of  $\mathcal{GP}_\theta(x_0, W)$  to highlight that the guided proposal always depends on the state observed with error at time  $T$ . To reduce notational overhead we have chosen not to do so.

Let  $x_0 = (q_0, p_0)$  and  $x_T = (q_T, p_T)$  denote the latent states. We propose a Gibbs sampler for updating  $(W, \theta, p_0)$  and assume that the noise on  $v_0$  is negligible so that it can be assumed that  $q_0 = v_0$ . Then for each sample of  $(W, \theta, p_0)$  we obtain a sample of the guided proposal via  $\mathcal{GP}_\theta((q_0, p_0), W)$ , which henceforth also yields a sample of  $x_T$ .

After initialization of  $(W, \theta, p_0)$  the sampler cycles over the following steps:

1. sample  $W$  conditional  $(x_0, \theta, v_T)$  using guided proposals, with pCN (preconditioned Crank–Nicolson) updates on the Wiener increments with Algorithm 5.1;
2. sample  $p_0$  conditional on  $(q_0, \theta, W, v_T)$  using MALA (Metropolis adjusted Langevin algorithm), with gradients obtained using automatic differentiation with Algorithm 5.2;
3. sample  $\theta$  conditional on  $(x_0, W, v_T)$  with Algorithm 5.3.

In these algorithms, we denote prior densities on  $\theta$ ,  $q_0$ , and  $p_0$  generically by  $\pi$  (Bayesian notation). In each of the steps, we keep track of the changes to  $X$  induced by updating either  $W$ ,  $p_0$ , or  $\theta$ . The acceptance probabilities in Algorithms 5.1, 5.2, 5.3, and 5.4 follow from the computations in section 4.1 of [30]. For readers less familiar with MCMC algorithms, we refer the reader to section A, where we also derive the acceptance probability for a simplified version of Algorithm 5.1.

---

**Algorithm 5.1** Update  $(W, X)$ , conditional on  $(x_0, \theta, v_T)$ .

---

- 1: Choose a persistence parameter  $\eta \in [0, 1]$ .
- 2: Compute  $L_\theta(t)$ ,  $M_\theta^\dagger(t)$  and  $\mu_\theta(t)$  for  $t \in [0, T]$  by solving (4.7).
- 2: Sample a Wiener process  $Z$ , independently of  $W$ , and set

$$W^\circ = \eta W + \sqrt{1 - \eta^2} Z.$$

- 3: Compute  $X^\circ = \mathcal{GP}_\theta(x_0, W^\circ)$ .
- 4: Compute

$$A = \Psi_\theta(X^\circ) / \Psi_\theta(X).$$

- 5: Draw  $U \sim \mathcal{U}(0, 1)$ .
  - 6: If  $U < A$ , set  $X = X^\circ$  and  $W = W^\circ$  else keep previous  $X$  and  $W$ .
- 

**Algorithm 5.2** Update  $(p_0, X)$ , conditional on  $(q_0, \theta, W, v_T)$ .

---

- 1: Choose a stepsize  $\delta > 0$ .
- 2: Sample  $Z \sim N_{dn}(0, I)$  (with  $d$  the dimension of a landmark).
- 3: Set  $x_0 = (q_0, p_0)$  and propose

$$p_0^\circ = p_0 + \frac{\delta}{2} \mathcal{L}_\theta(x_0, W) + \sqrt{\delta} Z,$$

where

$$\mathcal{L}_\theta(x_0, W) = \nabla_{p_0} \left( \log \Psi_\theta(\mathcal{GP}_\theta(x_0, W)) + \log \tilde{\rho}_\theta(0, x_0) \right).$$

Here, the gradient is obtained by automatic differentiation.

- 4: Set  $x_0^\circ = (q_0, p_0^\circ)$  and compute  $X^\circ = \mathcal{GP}_\theta(x_0^\circ, W)$  and  $\mathcal{L}_\theta(x_0^\circ, W)$ .
- 4: Compute

$$A = \frac{\Psi_\theta(X^\circ) \tilde{\rho}_\theta(0, x_0^\circ) \pi(p_0^\circ) \psi(p_0; p_0^\circ + \delta \mathcal{L}_\theta(x_0^\circ, W)/2, \delta I)}{\Psi_\theta(X) \tilde{\rho}_\theta(0, x_0) \pi(p_0) \psi(p_0; p_0 + \delta \mathcal{L}_\theta(x_0, W)/2, \delta I)}.$$

- 5: Draw  $U \sim \mathcal{U}(0, 1)$ .
  - 6: If  $U < A$ , set  $X = X^\circ$  and  $p_0 = p_0^\circ$  else keep previous  $X$  and  $p_0$ .
- 

**5.2. Template estimation.** In the template estimation problem, we assume  $I$  landmark configurations are observed with error. We assume that each configuration is a stochastic deformation of a fixed template configuration  $q_0$ . Fix an initial momentum vector  $p_0$  and denote  $x_0 = (q_0, p_0)$ . Let  $\{W^i, i = 1, \dots, I\}$  be independent Wiener processes on  $[0, T]$ . We assume that the  $i$ th landmark configuration is a sample from  $v_T^i = L_T X_T^i + N(0, \Sigma_T)$ , where



---

**Algorithm 5.3** Update  $(\theta, X)$ , conditional on  $(x_0, W, v_T)$ .

---

- 1: Choose an irreducible Markov kernel  $q$  on the domain of  $\theta$ .
- 2: Sample  $\theta^\circ$  from a kernel  $q(\cdot | \theta)$ .
- 3: Compute  $X^\circ = \mathcal{GP}_{\theta^\circ}(x_0, W)$ .
- 4: Compute

$$A = \frac{\Psi_{\theta^\circ}(X^\circ) \tilde{\rho}_{\theta^\circ}(0, x_0) \pi(\theta^\circ) q(\theta | \theta^\circ)}{\Psi_\theta(X) \tilde{\rho}_\theta(0, x_0) \pi(\theta) q(\theta^\circ | \theta)}.$$

- 5: Draw  $U \sim \mathcal{U}(0, 1)$ .
  - 6: If  $U < A$ , set  $X = X^\circ$  and  $\theta = \theta^\circ$  else keep previous  $X$  and  $\theta$ .
- 

$X^i$  has dynamics governed by the SDE of the Lagrangian, Langevin, or Eulerian model with initial state  $x_0$  and driven by the Wiener process  $W^i$ .

For simplicity we will assume the initial momentum to be zero, though it would not pose a problem to relax this assumption using Algorithm 5.2. A Gibbs sampler to sample from the posterior distribution of  $(W^1, \dots, W^I, \theta, q_0)$  is obtained by iterating the following steps:

1. sample each  $W^i$  conditional on  $(x_0, \theta, v_T^i)$  independently using Algorithm 5.1;
2. sample  $\theta$  conditional on  $(x_0, W^1, \dots, W^I, v_T^1, \dots, v_T^I)$  using a slight adaptation of Algorithm 5.3, where the loglikelihood contributions from different  $W^i$ 's are added;
3. sample  $q_0$  conditional on  $(W^1, \dots, W^I, v_T^1, \dots, v_T^I)$ .

For updating  $q_0$ , with a large number of landmarks simple random-walk updates will perform terribly as these do not respect the geometry of the landmark configuration. For that reason we propose to use the RMMALA (Riemannian Manifold Metropolis Adjusted Langevin Algorithm) as introduced in section 5 of [13]. Details are given in Algorithm 5.4.

**5.3. Computational aspects.** The computational cost of the proposed algorithms manifests itself in two ways. First, if parameter estimation is included (Algorithm 5.3), then the backward ODEs in (4.7) have to be recomputed if the parameters of the proposal change. Note, however, that in state-of-the-art competing methods the parameter is fixed, and if we choose to do likewise, then the ODEs need to be computed only once. Once solved on a fine temporal grid, at each gridpoint the Cholesky decomposition of  $M_t^\dagger$  needs to be computed, which scales cubically in the number of landmarks. With a large number of landmarks the computational costs can be drastically decreased by either applying an ensemble backward filter or enforcing sparsity. For details of these approaches we refer the reader to sections 6 and 7.3 in [30].

Second, both Algorithms 5.2 and 5.4 rely on gradients obtained by automatic differentiation. The runtime of the algorithms therefore depends on the computational efficiency of the automatic differentiation algorithm used. With large numbers of landmarks we expect reverse-mode automatic differentiation to be more efficient. Note that this is an implementational issue and not a deficit of the proposed methods. Finally, we remark that the Eulerian model is computationally more demanding than the Langevin model, as its diffusivity is nonconstant. This incurs some extra cost in simulating the guided proposal but, more importantly, also

---

**Algorithm 5.4** Update  $q_0$ , conditional on  $(\theta, W^1, \dots, W^I, v_T^1, \dots, v_T^I)$ .

---

- 1: Choose a stepsize  $\delta > 0$ .
- 2: Sample  $Z \sim N_{dn}(0, K(q_0))$ , where  $K(q)$  is the matrix with blocks  $K(q)_{ij} = K(q_i, q_j)$ ,  $1 \leq i \leq n$ .
- 3: Propose

$$q_0^\circ = q_0 + \frac{\delta}{2} K(q_0) \mathcal{L}(q_0) + \sqrt{\delta} Z,$$

where

$$\mathcal{L}(q_0) := \nabla_q \left( \log \tilde{\rho}_\theta(0, (q_0, p_0)) + \sum_{i=1}^I \log \Psi_\theta(g((q_0, p_0), W^i)) \right).$$

Here, the gradient is obtained by automatic differentiation. Set  $x_0^\circ = (q_0^\circ, p_0)$ .

- 4: For  $i = 1, \dots, I$ , compute  $X^{\circ, i} = \mathcal{GP}_\theta(x_0^\circ, W^i)$  and  $\mathcal{L}_\theta(q_0^\circ)$ .
- 5: Compute

$$A = \frac{\tilde{\rho}_\theta(0, x_0^\circ) \pi(q_0^\circ) \psi(q_0; q_0^\circ + \delta K(q_0^\circ) \mathcal{L}(q_0^\circ)/2, \delta K(q_0^\circ))}{\tilde{\rho}_\theta(0, x_0) \pi(q_0) \psi(q_0; q_0 + \delta K(q_0) \mathcal{L}(q_0)/2, \delta K(q_0))} \prod_{i=1}^I \frac{\Psi_\theta(X^{\circ, i})}{\Psi_\theta(X^i)}.$$

- 6: Draw  $U \sim \mathcal{U}(0, 1)$ .
  - 7: If  $U < A$ , set  $X^i = X^{\circ, i}$  for  $i = 1, \dots, I$  and  $q_0 = q_0^\circ$  else keep previous  $X$  and  $q_0$ .
- 

increased computational cost in computing  $\Psi(X^\circ)$  (cf. (4.5)). Contrary to the Lagrangian model, for the Eulerian model also the second term on the right-hand side of (4.6) needs to be evaluated.

From a practical point of view, we have noticed that tuning of stepsizes in the algorithm can be delicate. This can potentially be resolved by adaptive tuning at early iterations and fixing the stepsizes at subsequent iterations.

## 6. Numerical examples.

**6.1. Settings.** Before presenting examples, we discuss precise settings in the numerical experiments.

**6.1.1. Specification of Hamiltonian kernel.** We take a Gaussian kernel  $K$  of the form  $K(x) = k(x) \text{Id}_d$  with

$$(6.1) \quad k_a(x) = c \exp(-\|x\|^2/(2a^2)).$$

Note that  $\nabla k_a(x) = -c a^{-2} k(x) x$ . Without loss of generality we take  $c = 1$ .

*Remark 2.* For the Lagrangian model specified in (2.9), we have, by writing  $\tilde{H} = cH$ ,

$$(6.2) \quad dq_i^\alpha = c \frac{\partial \tilde{H}}{\partial p_i^\alpha} dt \quad \text{and} \quad dp_i^\alpha = -c \frac{\partial \tilde{H}}{\partial q_i^\alpha} dt + \gamma_i dW_t^i,$$

which, upon rescaling  $\gamma_i \rightarrow \gamma_i\sqrt{c}$  and  $t \rightarrow t/c$  (and  $T \rightarrow T/c$ ), has the form of (2.9) with  $c = 1$ . This implies that the parameters  $\gamma$  and  $c$  are not mutually independent. A similar consideration applies to the Eulerian model.

**6.1.2. Specification of intrinsic noise.** For the Lagrangian model, we take the noise on all landmarks the same, scaled by  $\sqrt{n}$ . That is,  $\gamma_i = \gamma/\sqrt{n}$  for all  $i \in \{1, \dots, n\}$ . In the experiments we take  $\gamma = 0.1$ .

For the Eulerian model, let  $\tau > 0$ . We take the noise fields centered at points that are both horizontally and vertically separated by a distance that is an integer multiple of  $2\tau$ . Denote the locations of the noise fields by  $\{\delta_j\}$ . At each location  $\delta_j$  we take noise fields to be

$$\frac{2}{\pi}\gamma k_\tau(x - \delta_j) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \frac{2}{\pi}\gamma k_\tau(x - \delta_j) \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

This kernel is the same as the Hamiltonian kernel, but does not have to be so in general. The scaling by  $2/\pi$  is chosen so that if  $\gamma = 1$ , the noise on the landmark positions is close to 1 uniformly in space (close to a decomposition of unity).

**6.1.3. Specification of extrinsic noise.** In each example we choose  $\Sigma_T = \epsilon^2 I$  with  $\epsilon = 0.01$  (cf. (3.5)), except in the examples of sections 6.2.1 and 6.3.2, where we took  $\epsilon = 0.001$ . From a practical point of view, as landmarks are often annotated manually, it is a realistic assumption that  $\epsilon$  is strictly positive.

**6.1.4. Specification of parameter  $\theta$ .** In the simulations, we fix the scaling parameter  $\tau$  of the noise-kernels, as well as their positions. We estimate the Hamiltonian kernel  $a$ , but fix the size (amplitude) of the noise ( $\gamma$ ). Hence,  $\theta = a$ .

**6.1.5. Prior on the landmark positions and momenta.** In the problem of bridging two landmark configurations we use a prior on the landmark momenta, similarly to the prior suggested in section 3 of [29]. More specifically, we take, for  $\kappa_{\text{mom}}, \kappa_{\text{pos}} > 0$ ,

$$\begin{aligned} \pi(q_1, \dots, q_n, p_1, \dots, p_n) &= \pi(p_1, \dots, p_n \mid q_1, \dots, q_n)\pi(q_1, \dots, q_n) \\ &= \psi(p_1, \dots, p_n; 0, \kappa_{\text{mom}}K(q)^{-1}) \prod_{i=1}^n \psi(q_i; 0, \kappa_{\text{pos}}). \end{aligned}$$

As we assumed that the landmark positions are observed, we make the shortcut where we assume  $(p_1, \dots, p_n) \sim N(0, \kappa_{\text{mom}}K(q_0)^{-1})$ . Here,  $\kappa_{\text{mom}}$  is a parameter that we fix to a large value. In the experiments we take  $\kappa_{\text{mom}} = 100$ .

**6.1.6. Tuning parameters for the MCMC sampler.** As can be seen from Algorithms 5.1, 5.2, and 5.3, the proposed MCMC algorithm requires specification of

- the persistence parameter  $\eta \in [0, 1]$  (Algorithm 5.1);
- the stepsize  $\delta > 0$  (Algorithm 5.2);
- the Markov kernel  $q$  for updating  $\theta$  (Algorithm 5.3).

Naturally, these choices affect the efficiency of the algorithm to explore the support of the target distribution. We target acceptance rates of about 50%. For the Markov kernel  $q$  as  $\theta$  we use updates of the form  $\log \theta_i^\circ \mid \theta_i \sim N(\log \theta_i, \sigma_\theta^2)$  to ensure  $\theta_i^\circ > 0$ .

**6.1.7. Time discretization.** We discretize guided proposals in time by taking a regular grid on  $[0, 1]$  with mesh-width 0.01, followed by applying the map  $s \mapsto s(2 - s)$  to this grid. The latter ensures the discretization is finer near the endpoint  $T$ , where the guiding term can become large (see also section 5 on [43] for discretization of guided proposals, where the choice of maps  $s \mapsto s(2 - s)$  is motivated).

For computing the guiding term and the right-hand side of (4.4), we solve the system in (4.7), where the differential equation for  $L(t)$  was solved with an implicit Euler scheme and both  $M^\dagger$  and  $\mu$  were approximated using the trapezoid rule (all on a regular grid with mesh-width 0.01).

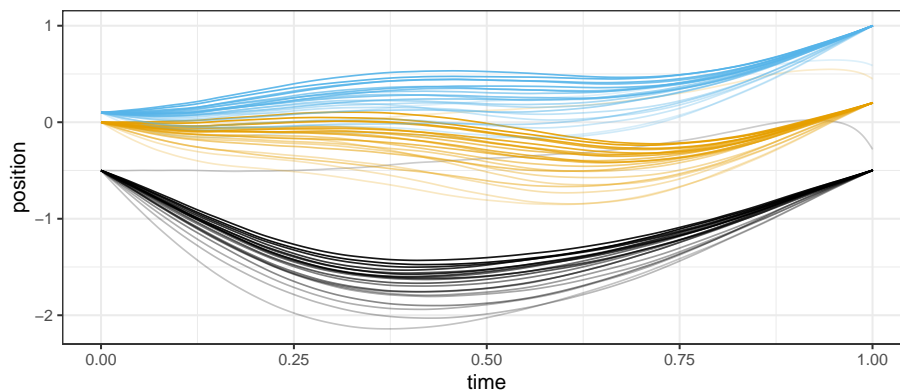
**6.1.8. Software, hardware.** The source code of the numerical examples is available together with a Julia package `BridgeLandmarks` for shape analysis with stochastic landmark dynamics, [44]. Automatic differentiation was implemented via [35]. All computations were done on a MacBook Pro, with a 2 GHz Quad-Core Intel Core i5 with 16 GB RAM.

## 6.2. Examples for landmark matching.

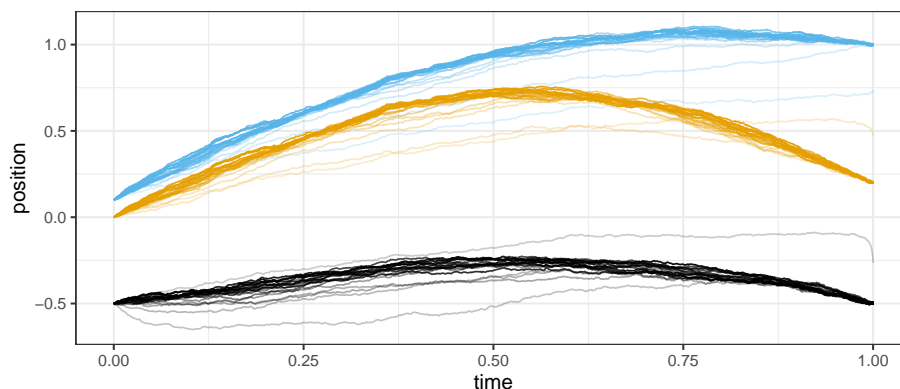
**6.2.1. One-dimensional landmarks.** We start with an example that is meant to illustrate the developed methodology in a one-dimensional setup. We consider a problem with just 3 landmarks. Here, the initial locations of the points are given by  $-0.5$ ,  $0.0$ , and  $0.1$  and we condition on  $-0.5$ ,  $0.2$ , and  $1.0$  respectively. We consider both the Lagrangian and Eulerian models and fixed the Hamiltonian kernel parameter to be 1.0. For the Eulerian model, 6 noise sources were positioned on the domain  $[-2.5, 2.5]$  with  $\tau = 0.5$ . For the Lagrangian model we fixed  $\gamma = 1/\sqrt{n} \approx 0.58$ , and for the Eulerian model we took  $\gamma = 0.1$ . Hence, we assume all parameters are known, except for the initial momenta and bridges. We consider landmark matching without noise, which is forced by assuming that the final position is observed with additive  $N(0, 10^{-6})$ -noise ( $\varepsilon = 0.001$ ; cf. section 6.1.3).

We ran an algorithm composed of alternating one step of Algorithm 5.1 (bridge updating) and one step of Algorithm 5.2 (initial momenta updating) for 20,000 iterations. We computed the guided proposal on a grid with mesh-width 0.001, subsequently mapped to a nonequidistant grid as explained in subsection 6.1.7. We initialized with zero momentum for all of the three points at time 0.

Sampled bridges for the Lagrangian and Eulerian models are shown in Figures 1 and 2, respectively. Clearly, upon initialization the bridges do not match the conditionings at time  $t = 1$ . This is resolved, however, within a couple of thousand iterations. In Figures 3 and 4 we show for both models traceplots of the momenta, both at the initial time and final time (note that we have matched the colors in the four figures). This foremost serves as a way to detect whether or not the chain has reached its stationary regime. Second, it is interesting to observe that for the Lagrangian model, the initial momenta are quite high and that for the upper two curves (blue and orange) the initial momenta are a posteriori in opposite directions. This can be explained by the value of the Hamiltonian kernel parameter (taken equal to 1), which is deliberately chosen relatively large, forcing the orange and blue paths to remain close. However, as we condition the paths at distant locations (0.2 and 1.0), the opposite initial momenta ensure the conditionings at time 1 are satisfied.



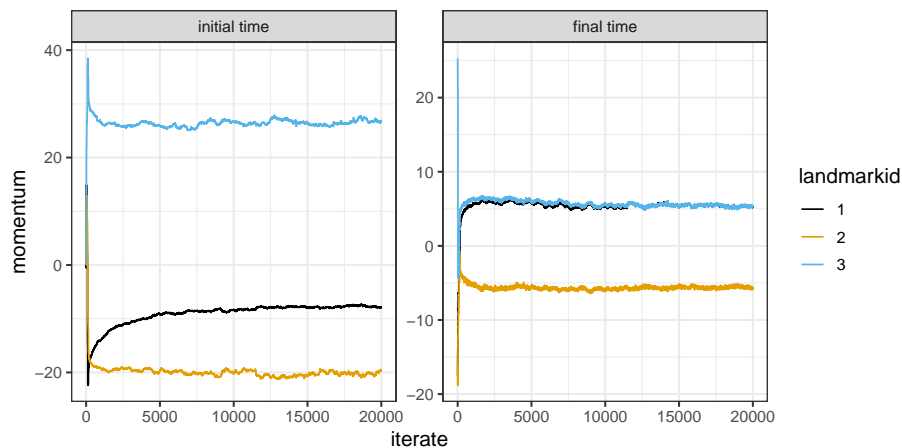
**Figure 1.** Bridges for the one-dimensional landmarks example, Lagrangian model. The sampler ran for 20,000 iterations. Every 1,000th iteration is shown. “Early” iterations are depicted less thick. Note that the initial trajectory does not satisfy the conditioning.



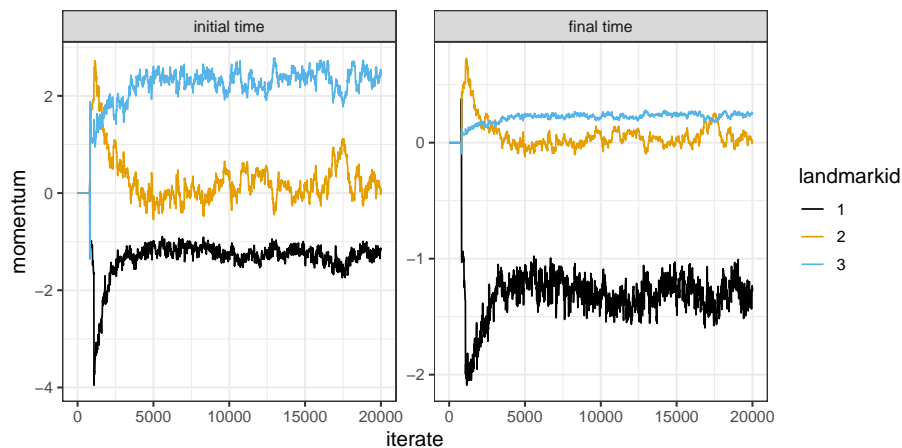
**Figure 2.** Bridges for the one-dimensional landmarks example, Eulerian model. The sampler ran for 20,000 iterations. Every 1,000th iteration is shown. “Early” iterations are depicted less thick. Note that the initial trajectory does not satisfy the conditioning.

**6.2.2. Two-dimensional example from [21].** Here we consider an example with 12 two-dimensional landmarks from [21]. In this case we also include the Hamiltonian kernel parameter  $a$  in the sampler, which is endowed with the Pareto distribution. More precisely, we assume  $p(a) = 0.1a^{-2}\mathbf{1}_{[0.1,\infty)}(a)$  (hence the shape and scale parameters are taken 1 and 0.1, respectively). Having  $a$  bounded away from zero is beneficial for numerical stability. Just like the previous example, for the Lagrangian model we fixed  $\gamma = 1/\sqrt{n} \approx 0.29$ , and for the Eulerian model we fixed  $\gamma = 0.1$ .

We ran an MCMC sampler with 20,000 iterations, where one iterate consists of applying one step of Algorithms 5.1, 5.2, and 5.3. We include figures for the Lagrangian model. In Figures 5 and 6 we visualize bridges and shape-evolution. Figure 7 shows traceplots for both coordinates of the momentum at the initial time and final time. From these figures one can deduce, for example, that for landmark number 5 the initial momentum is quickly increased



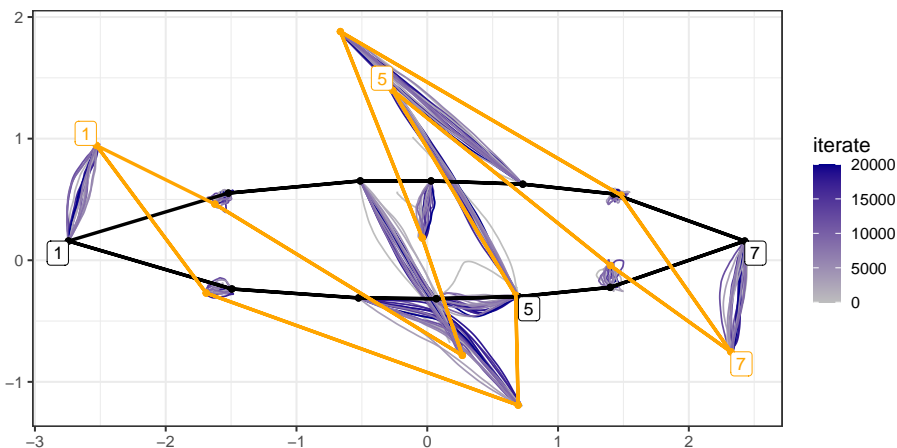
**Figure 3.** Lagrangian model, one-dimensional landmarks example. Traceplots showing iterates of momenta at times 0 and 1. Every 10th iteration is shown.



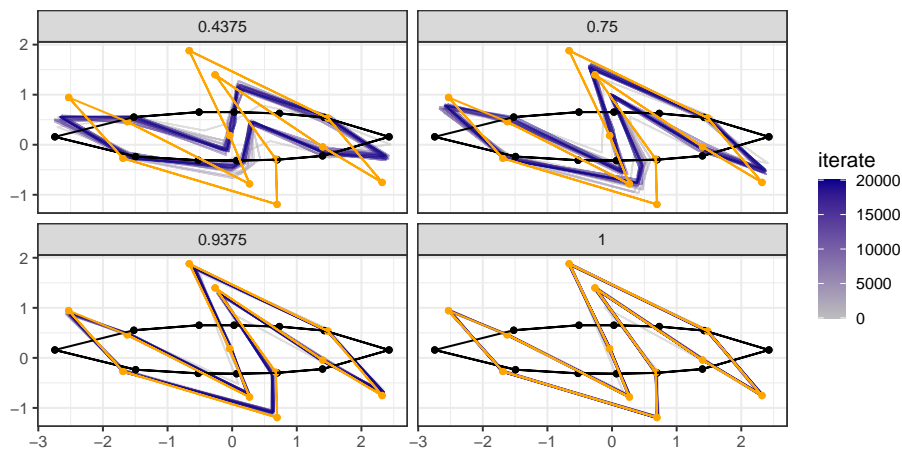
**Figure 4.** Eulerian model, one-dimensional landmarks example. Traceplots showing iterates of momenta at times 0 and 1. Every 10th iterations is shown.

after the initial iterations, to ensure that the bridge path indeed moves in a northward direction. Figure 8 gives a traceplot of the Hamiltonian kernel parameter  $a$ . One can see that it takes about 10,000 iterations for the sampler to stabilize, and the corresponding change in sampled bridges is quite apparent in Figure 5 (switch from grayish to purple-colored paths).

**6.2.3. Corpus callosum.** Here, we consider an example with many more landmarks where we start from an ellipse and transform to landmarks representing points on the outline of a human corpus callosum. In total there are 77 landmarks. Here we used the Lagrangian model with  $a = 0.2$  (Hamiltonian kernel parameter), the average distance between adjacent landmarks on the shape and  $\gamma = 1/\sqrt{n} \approx 0.12$ . In Figure 9 we show sampled bridges over 2,500 iterations, and in Figure 10 correspondingly the deformed shapes at 4 time instances



**Figure 5.** Bridges for the Joshi–Miller example, Lagrangian model. The sampler ran for 20,000 iterations. Every 1,000th iteration is shown.

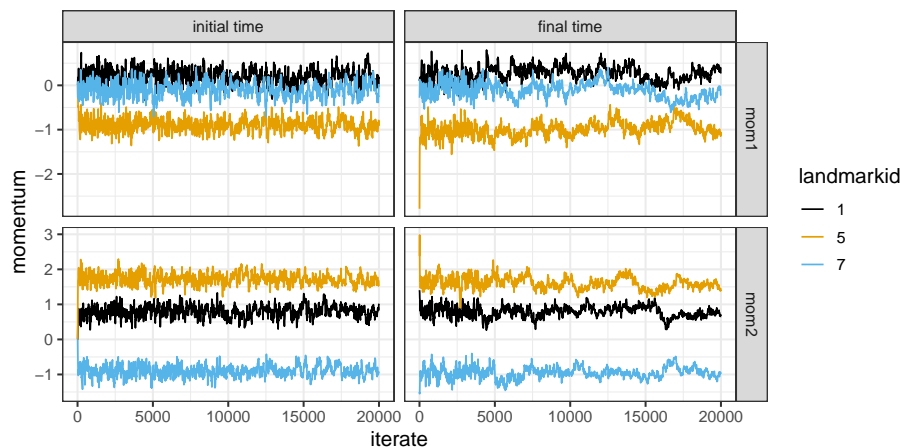


**Figure 6.** Shape evolution for the Joshi–Miller example, Lagrangian model. The sampler ran for 20,000 iterations. Every 1,000th iteration is shown.

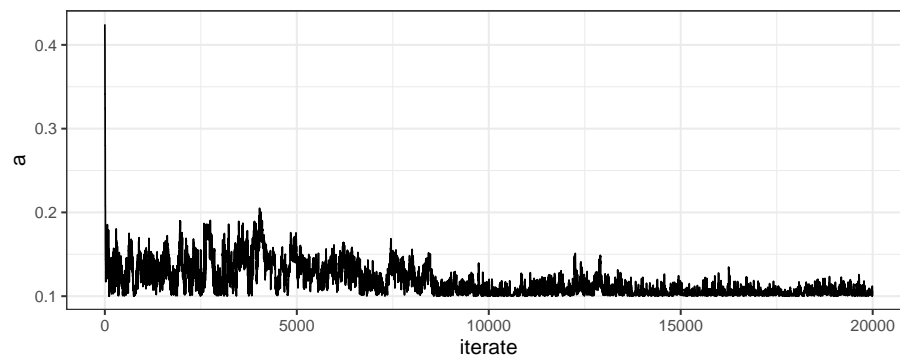
in  $[0, 1]$ . Traceplots of the momenta at times 0 and 1 are shown in Figure 11. This example illustrates that the proposed algorithms can also be used in settings with a large number of landmarks. Note that after about 200 iterations the bridges visually look very similar, whereas the early (light-gray) paths look different.

### 6.3. Examples for template estimation.

**6.3.1. Simulated shapes on an ellipse.** Here, we simulated forward 10 trajectories  $n = 15$  landmarks where the initial configuration is an ellipse. In the MCMC estimation, we took the Lagrangian model and fixed the initial momenta to be equal to zero but include parameter estimation. Hence, each MCMC iteration consists of executing one step of Algorithms 5.1, 5.3, and 5.4. The parameter  $a$  is endowed with the Pareto distribution. More precisely, we assume



**Figure 7.** Joshi–Miller example, Lagrangian model. Traceplots showing iterates of momenta at times 0 and 1. Every 10th iteration is shown.



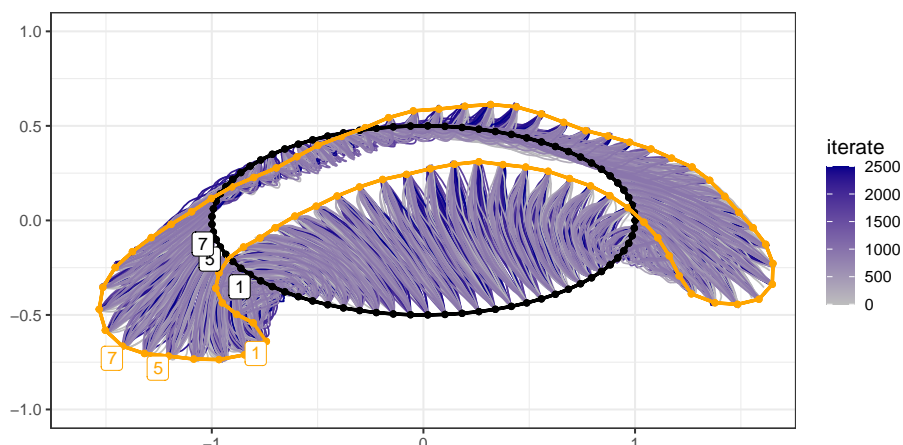
**Figure 8.** Joshi–Miller example, Lagrangian model. Traceplots showing iterates of the Hamiltonian kernel parameter  $a$ .

$p(a) = 0.1a^{-2}\mathbf{1}_{[0.1,\infty)}(a)$  (hence the shape and scale parameters are taken to be 1 and 0.1, respectively). We ran the sampler for 2,500 iterations, saving every 50th configuration. We deliberately initialized the template configuration incorrectly by taking one of the observed configurations and rotating and stretching the shape. This is done to illustrate that the RMMALA steps perform satisfactory. In any practical application one could initialize the template shape by one of the observed shapes.

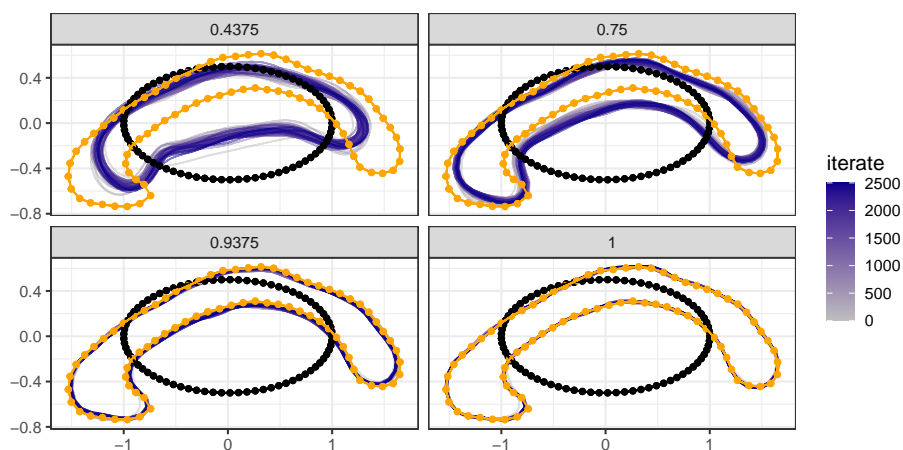
Iterators of the template and parameter  $a$  are shown in Figures 12 and 13, respectively. Clearly, the initial configuration can be recovered quite well, despite there being only 10 observed shapes. From the traceplots of the parameter  $a$  it appears there is mild posterior uncertainty about its value.

**6.3.2. Cardiac data.** In Figure 14 we show the results when applying the algorithm to landmarks on 14 cardiac images of human left ventricles [41]. The manually annotated landmarks are consistently placed over the set of shapes. Based on 22 landmarks for each shape





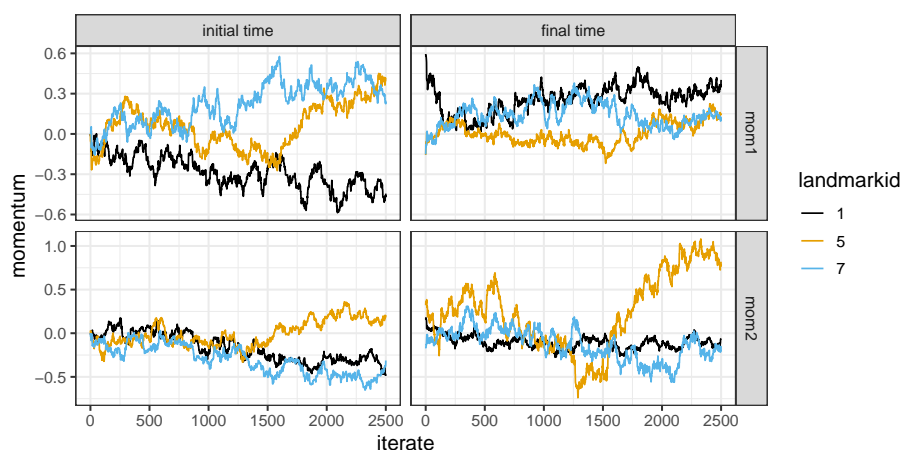
**Figure 9.** Shape evolution for the Corpus callosum data. The sampler ran for 500 iterations. Every 10th iteration is shown.



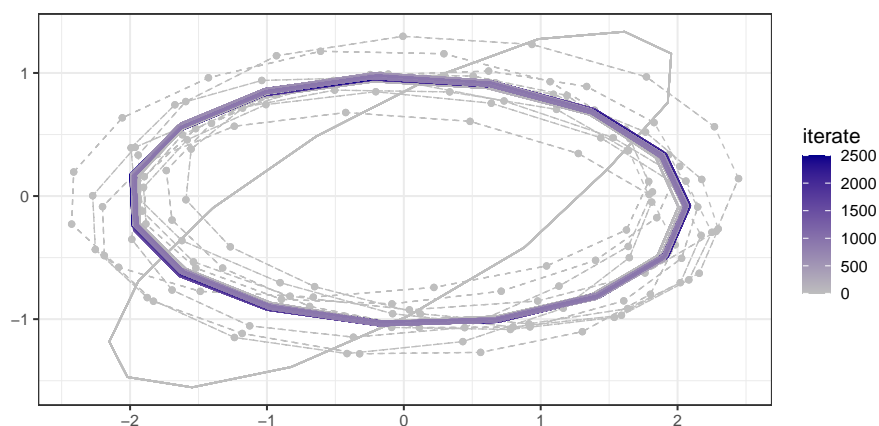
**Figure 10.** Corpus callosum data. Initial shape (black) and final shape (orange) with 77 landmarks. Every 10th iteration from 500 iterations is shown.

we estimate the template shape and Hamiltonian kernel parameter. This experiment extends the template estimation with Brownian motion performed in [40] to the models in focus in this paper. We used parameter updating (assuming a priori  $p(a) = 0.01a^{-2}\mathbf{1}_{[0.01,\infty)}(a)$ ), bridge updating, and initial state updating (using RMMALA), while fixing the initial momenta to zero. Hence, just as in the previous subsection, each MCMC iteration consists of executing one step of Algorithms 5.1, 5.3, and 5.4. We used 2,500 iterations. From Figure 14 it appears that the chain reaches its stationary region within 500 iterations.

**7. Discussion and future work.** We have extended the framework of [38, 8] for sampling diffusion bridges to nonlinear high-dimensional stochastic Hamiltonian systems. For this work, problems in computational anatomy [46] revolving around systems of particles



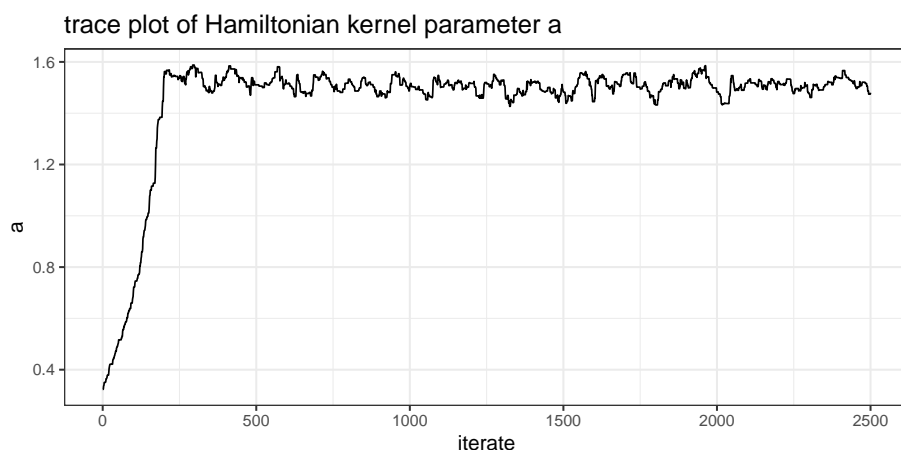
**Figure 11.** *Corpus callosum data. Traceplots showing iterates of momenta at times 0 and 1.*



**Figure 12.** *Initial shape configuration for the example in section 6.3.1 using the Lagrangian model. The dashed gray curves are the 10 observed configurations. Every 20th MCMC iterate is shown. Note that RMMALA quickly “discovers” the deliberately wrong initialization.*

to describe images have been our main motivation. We used these models throughout to illustrate the application of diffusion bridges, specifically in stochastic landmark dynamics. We left open many possible improvements of this method, such as allowing for multiple observations over time, resolving unknown landmark correspondence, including a step towards infinite-dimensional shape matching, and solving the problem of the parametrization of these shapes [6].

This work opens the door to not only a wider use in statistical inference in computational anatomy, but also other related fields, where high-dimensional stochastic Hamiltonian systems are involved in the modeling, such as in biology, data assimilation, swarming, etc. From a more theoretical point of view, various extensions of this work could be considered. The driving Brownian motion in the SDE could, for example, be generalized to a Lévy process or even rough paths.



**Figure 13.** Traceplot for Hamiltonian kernel parameter for the example in section 6.3.1 using the Lagrangian model.

From a numerical perspective, we used classical methods for simplicity, but the presented algorithms could be extended to include more modern numerical integration schemes or MCMC methods, such as geometric integrators [31, 14, 32, 19, 10], to preserve the geometrical structure of the problem, or some more advanced geometrical Hamiltonian Monte Carlo methods [3, 4]. In addition, the use of automatic differentiation, similar to those in [25, 26], has turned out to be beneficial for the implementation of efficient methods, and surely deserves to be more exploited for future works.

**Appendix A. Markov chain Monte Carlo.** A *Markov chain Monte Carlo (MCMC)* method for sampling from a distribution  $\pi$  is any method producing an ergodic Markov chain whose stationary distribution is  $\pi$ . Virtually all such algorithms can be seen as specific instances of the Metropolis–Hastings algorithm. For this paper it suffices to assume that  $\pi$  has a density with respect to Lebesgue measure on  $\mathbb{R}^d$ . We will denote this density by  $\bar{\pi}$ . Suppose  $Q$  is a Markov kernel on  $\mathbb{R}^d$  with density  $q$ . That is, for measurable sets  $B$ ,  $x \mapsto Q(x, B)$  is a measurable mapping, and for fixed  $x$ ,  $B \mapsto Q(x, B) = \int_B q(x, y) dy$  is a probability measure.

**Definition 2.** The Metropolis–Hastings algorithm is the algorithm by which a Markov chain is constructed which evolves  $x_n = x$  to  $x_{n+1}$  by the following steps:

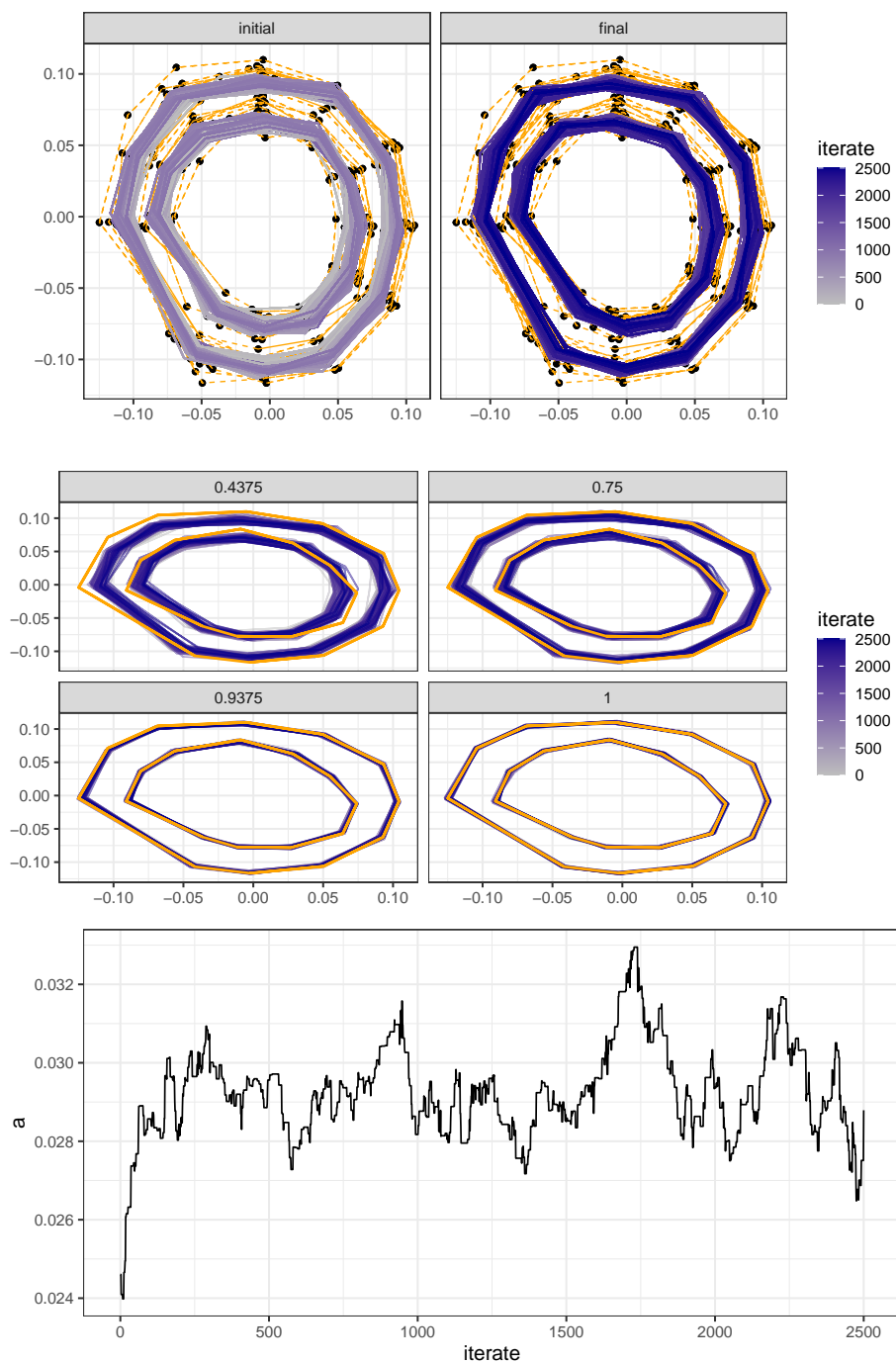
1. Propose  $y$  from the proposal kernel  $Q(x, \cdot)$ .
2. Compute

$$\alpha(x, y) = \min \left( 1, \frac{\bar{\pi}(y) q(y, x)}{\bar{\pi}(x) q(x, y)} \right).$$

3. Set

$$x_{n+1} = \begin{cases} y & \text{with probability } \alpha(x, y), \\ x & \text{with probability } 1 - \alpha(x, y). \end{cases}$$

Under weak assumptions, this algorithm indeed produces an ergodic Markov chain with stationary distribution  $\pi$ . For computing  $\alpha(x, y)$  it suffices to know  $\pi$  up to a proportionality



**Figure 14.** Cardiac data example from section 6.3.2. Top: iterates for template shape (locations of landmarks at time zero). The first 1,250 iterates are shown in the left panel, “initial”; the final 1,250 iterates are shown in the right panel, “final.” Middle: evolution over time of one chosen (fixed) shape (the iterates are shown at times 0.4375, 0.75, 0.9375, and 1.0). In both the top and middle figures the observed landmark locations are depicted by black dots, connected by orange line segments. Bottom: traceplot for the Hamiltonian kernel parameter.

constant. Within Bayesian statistics, this is a very attractive property of the algorithm, as the posterior is typically only known up to an integration constant.

The specification of the algorithm leaves open the choice of the proposal kernel  $Q$ . A symmetric random walk proposal  $q(x, y) = \varphi(y - x)$ , with  $\varphi$  a symmetric density around zero, is a simple choice that may work if the dimension  $d$  is small. Otherwise, a more sophisticated proposal kernel that takes gradient information into account may be computationally more efficient. The Metropolis Adjusted Langevin Algorithm (MALA) is based on the fact that the solution to the stochastic differential equation

$$dX_t = A \nabla \log \bar{\pi}(X_t) dt + \sqrt{2A} dW_t,$$

with  $W$  denoting Brownian motion, has invariant distribution  $\pi$ . Here,  $A$  is a strictly positive definite preconditioning matrix and  $\sqrt{A}$  is the matrix-square root of  $A$ . Upon discretizing this equation with stepsize  $\tau$  (Euler forward discretization) we obtain the proposal density  $q(x, y) = \psi(y; x + A \nabla \log \bar{\pi}(x), 2A)$  with (as before)  $\psi(x; \mu, \Sigma)$  denoting the density of the  $N(\mu, \Sigma)$ -distribution, evaluated at  $x$ . Both  $\tau$  and  $A$  are tuning parameters;  $\tau$  acts as a stepsize. In general, the choice of  $A$  is dictated by the geometrical properties of the density  $\bar{\pi}$ . This is the basic idea of the Riemannian MALA algorithm. If the dimension  $d$  is large, one often uses a proposal where only parts for the current iterate  $x_n = x$  are stochastically perturbed. This is sometimes called “one-at-a-time” sampling, “substitution sampling,” or simply “Gibbs sampling.” Clearly, this is just a special case of the Metropolis–Hastings algorithm.

For illustration purposes, we give a finite-dimensional version corresponding to Algorithm 5.1 (updating of bridges for fixed parameter and initial state). Hence assume a data-generating model of the form  $w \sim N(0, I)$ ,  $x = F(w)$  leading to the observation  $y = h(x) + \epsilon$ ,  $\epsilon$  being a random quantity independent of  $w$ . Assume the likelihood for the parameter  $x$  is given by  $L(x) = L(F(w))$ . Suppose  $L(F(w)) = c\Psi(G(w))$  for fixed maps  $G$  and  $\Psi$ , the constant being independent of  $w$ . Using the proposal density  $q(w^\circ | w)$ , the acceptance probability for the Metropolis–Hastings algorithm is given by  $A \wedge 1$ , where  $A = \frac{q(w|w^\circ) \psi(w^\circ; 0, I) \Psi(G(w^\circ))}{q(w^\circ|w) \psi(w; 0, I) \Psi(G(w))}$ . If  $q$  is chosen according to the Crank–Nicolson rule, i.e.,  $w^\circ | w \sim N(\eta w, (1 - \eta^2)I)$ , then the first two terms in the expression for  $A$  cancel and we end up with  $A = \Psi(x^\circ)/\Psi(x)$ , where  $x^\circ = G(w^\circ)$  and  $x = G(w)$ . This expression remains valid if  $w$  is a Wiener process and  $G = \mathcal{GP}$  is a guided proposal.

### Appendix B. Matching conditions on the auxiliary process for exact landmark matching.

For exact matching, absolute continuity will only hold if the parameters of the auxiliary process  $\tilde{X}$  satisfy certain *matching conditions*. In case the diffusion is uniformly elliptic,  $L_T = I_{N \times N}$ , it was shown in [38] that  $\tilde{a} = \sigma\sigma'$  must satisfy  $\tilde{a}(T) = a(T, x_T)$  for absolute continuity. In the more general case considered here, the matching conditions look somewhat more difficult; a precise statement is given in section 2 of [8]. Especially for the Eulerian model, these conditions are hard to verify (mainly due to the fact that the diffusivity is state dependent and not all components of  $X_T$  are observed). However, the results in this paper suggest that  $\mathbb{P}^*$  is absolutely continuous with respect to  $\mathbb{P}^\circ$  if (i) condition (4.9) is satisfied, and (ii) the number of noise sources  $J$  satisfies  $J \geq Nd$  ( $N$  and  $d$  denoting the number of landmarks and their dimension, respectively). This conjecture is numerically confirmed by experiments in section 4.1 of [8]. The second condition that requires sufficiently many noise sources has also been discussed in Remark 4.1 in [2].

**Appendix C. Proof of Proposition 1 on Stratonovich to Itô correction for Eulerian model.** In coordinates, the stochastic equations for the Eulerian model are

$$(C.1) \quad \begin{aligned} dq_i^\alpha &= \frac{\partial h}{\partial p_i^\alpha} dt + \sum_{l=1}^J \sigma_l^\alpha(q_i) \circ dW_t^l, \\ dp_i^\alpha &= -\frac{\partial h}{\partial q_i^\alpha} dt - \sum_{l=1}^J \sum_{\beta} \frac{\partial \sigma_l^\beta(q_i)}{\partial q_i^\alpha} p_i^\beta \circ dW_t^l. \end{aligned}$$

The same process with Itô's integrals has the additional term

$$(C.2) \quad dq_i^\alpha + = \frac{1}{2} \frac{\partial \sigma_l^\alpha(q_i)}{\partial q_i^\beta} \sigma_l^\beta(q_i)$$

in the  $q$  equation, and for the  $p$  equation

$$(C.3) \quad dp_i^\alpha + = \frac{1}{2} p_i^\gamma \frac{\partial \sigma_l^\gamma(q_i)}{\partial q_i^\beta} \frac{\partial \sigma_l^\beta(q_i)}{\partial q_i^\alpha} - \frac{1}{2} p_i^\beta \frac{\partial^2 \sigma_l^\beta(q_i)}{\partial q_i^\alpha \partial q_i^\gamma} \sigma_l^\gamma(q_i).$$

The extra term for  $q^\alpha$  given in (C.2) equals

$$\begin{aligned} \frac{1}{2} \sum_{\ell} \sum_{\beta} \frac{\partial \sigma_{\ell}^\alpha(q)}{\partial q^\beta} \sigma_{\ell}^\beta(q) &= \frac{1}{2} \sum_{\ell} \sum_{\beta} \gamma_\alpha \nabla_\beta \bar{k}_\tau(q - \delta) \gamma_\beta \bar{k}_\tau(q - \delta) \\ &= \frac{1}{2} \sum_{\ell} \langle \nabla \bar{k}_\tau(q - \delta), \gamma \rangle \bar{k}_\tau(q - \delta) \gamma_\alpha, \end{aligned}$$

which gives (4.11). Notice that we write  $\nabla_\beta$  to denote  $(\partial)/(\partial q^\beta)$ .

For  $p^\alpha$  the first term on the right-hand side of (C.3) is given by

$$\begin{aligned} \frac{1}{2} \sum_{\ell} \sum_u \sum_{\beta} p^u \frac{\partial \sigma_{\ell}^u(q)}{\partial q^\beta} \frac{\partial \sigma_{\ell}^\beta(q)}{\partial q^\alpha} &= \frac{1}{2} \sum_{\ell} \sum_u \sum_{\beta} \gamma_u \gamma_\beta \nabla_\beta \bar{k}_\tau(q - \delta) \nabla_\alpha \bar{k}_\tau(q - \delta) p^u \\ &= \frac{1}{2} \sum_{\ell} \nabla_\alpha \bar{k}_\tau(q - \delta) \langle \gamma, p \rangle \langle \gamma, \nabla \bar{k}_\tau(q - \delta) \rangle. \end{aligned}$$

This gives the first term in (4.12). The second term on the right-hand side of (C.3) is given by

$$\begin{aligned} -\frac{1}{2} \sum_{\ell} \sum_u \sum_{\beta} p^\beta \frac{\partial^2 \sigma_{\ell}^\beta(q)}{\partial q^\alpha \partial q^u} \sigma_{\ell}^u(q) &= -\frac{1}{2} \sum_{\ell} \sum_u \sum_{\beta} p^\beta \gamma_\beta [\nabla_\alpha (\nabla_u \bar{k}_\tau(q - \delta))] \gamma_u \bar{k}_\tau(q - \delta) \\ &= -\frac{1}{2} \sum_{\ell} \langle (p, \gamma) \rangle \bar{k}_\tau(q - \delta) \nabla_\alpha (\langle \nabla \bar{k}_\tau(q - \delta), \gamma \rangle). \end{aligned}$$

This gives the second term in (4.12).

**Acknowledgment.** A substantial part of this research was carried out during a sabbatical leave of the second author at the University of Copenhagen (UCPH).

## REFERENCES

- [1] A. ARNAUDON, D. D. HOLM, A. PAI, AND S. SOMMER, *A stochastic large deformation model for computational anatomy*, in *Information Processing in Medical Imaging*, Lecture Notes in Comput. Sci., Springer, 2017, pp. 571–582, [https://doi.org/10.1007/978-3-319-59050-9\\_45](https://doi.org/10.1007/978-3-319-59050-9_45).
- [2] A. ARNAUDON, D. D. HOLM, AND S. SOMMER, *A geometric framework for stochastic shape analysis*, *Found. Comput. Math.*, 19 (2019), pp. 653–701, <https://doi.org/10.1007/s10208-018-9394-z>.
- [3] A. BARP, F.-X. BRIOL, A. D. KENNEDY, AND M. GIROLAMI, *Geometry and dynamics for Markov chain Monte Carlo*, *Annu. Rev. Statist. Appl.*, 5 (2018), pp. 451–471.
- [4] A. BARP, A. KENNEDY, AND M. GIROLAMI, *Hamiltonian Monte Carlo on Symmetric and Homogeneous Spaces via Symplectic Reduction*, preprint, <https://arxiv.org/abs/1903.02699>, 2019.
- [5] F. BAUDOIN, *Conditioned stochastic differential equations: Theory, examples and application to finance*, *Stochastic Process. Appl.*, 100 (2002), pp. 109–145, [https://doi.org/DOI:10.1016/S0304-4149\(02\)00109-6](https://doi.org/DOI:10.1016/S0304-4149(02)00109-6).
- [6] M. BAUER, M. BRUVERIS, AND P. W. MICHOR, *Overview of the geometries of shape spaces and diffeomorphism groups*, *J. Math. Imaging Vis.*, 50 (2014), pp. 60–97.
- [7] A. BESKOS, O. PAPASPILIOPOULOS, G. O. ROBERTS, AND P. FEARNHEAD, *Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion)*, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68 (2006), pp. 333–382, <https://doi.org/10.1111/j.1467-9868.2006.00552.x>.
- [8] J. BIERKENS, F. VAN DER MEULEN, AND M. SCHAUER, *Simulation of partial bridges for hypo-elliptic diffusions*, *Adv. Appl. Probab.*, 52 (2020), pp. 173–212.
- [9] M. BLADT, S. FINCH, AND M. SØRENSEN, *Simulation of multivariate diffusion bridges*, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 78 (2016), pp. 343–369, <https://doi.org/10.1111/rssb.12118>.
- [10] E. BUCKWAR, M. TAMBORRINO, AND I. TUBIKANEC, *Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs*, *Stat. Comput.*, 30 (2020), pp. 627–648.
- [11] R. CAMASSA AND D. D. HOLM, *An integrable shallow water equation with peaked solitons*, *Phys. Rev. Lett.*, 71 (1993), pp. 1661–1664, <https://doi.org/10.1103/PhysRevLett.71.1661>.
- [12] B. DELYON AND Y. HU, *Simulation of conditioned diffusion and application to parameter estimation*, *Stochastic Process. Appl.*, 116 (2006), pp. 1660–1675, <https://doi.org/DOI:10.1016/j.spa.2006.04.004>.
- [13] M. GIROLAMI AND B. CALDERHEAD, *Riemann manifold Langevin and Hamiltonian Monte Carlo methods*, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73 (2011), pp. 123–214, <https://doi.org/10.1111/j.1467-9868.2010.00765.x>.
- [14] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric Numerical Integration: Structure-preserving Algorithms for Ordinary Differential Equations*, Springer Ser. Comput. Math. 31, Springer Science & Business Media, 2006.
- [15] D. D. HOLM, *Variational principles for stochastic fluid dynamics*, *Proc. A*, 471 (2015), 20140963, <https://doi.org/10.1098/rspa.2014.0963>.
- [16] D. D. HOLM AND J. E. MARSDEN, *Momentum maps and measure-valued solutions (peakons, filaments, and sheets) for the EPDiff equation*, in *The Breadth of Symplectic and Poisson Geometry*, Springer, 2005, pp. 203–235.
- [17] D. D. HOLM, J. E. MARSDEN, AND T. S. RATIU, *The Euler-Poincaré equations and semidirect products with applications to continuum theories*, *Adv. Math.*, 137 (1998), pp. 1–81, <https://doi.org/10.1006/aima.1998.1721>.
- [18] D. D. HOLM, J. T. RATNANATHER, A. TROUVÉ, AND L. YOUNES, *Soliton dynamics in computational anatomy*, *NeuroImage*, 23 (2004), pp. S170–S178, <https://doi.org/doi:10.1016/j.neuroimage.2004.07.017>.
- [19] D. D. HOLM AND T. M. TYRANOWSKI, *Stochastic discrete Hamiltonian variational integrators*, *BIT*, 58 (2018), pp. 1009–1048.
- [20] T. JEULIN AND M. YOR, EDS., *Grossissements de filtrations: Exemples et applications*, Papers from the seminar on stochastic calculus held at the Université de Paris VI, Paris, 1982/1983, Lecture Notes in Math. 1118, Springer-Verlag, Berlin, 1985, <https://doi.org/10.1007/BFb0075765>.
- [21] S. JOSHI AND M. MILLER, *Landmark matching via large deformation diffeomorphisms*, *Proc. IEEE Trans. Image Process.*, 9 (2000), pp. 1357–1370.

- [22] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, 2nd ed., Grad. Texts in Math. 113, Springer-Verlag, New York, 1991, <https://doi.org/10.1007/978-1-4612-0949-2>.
- [23] D. G. KENDALL, *Shape manifolds, procrustean metrics, and complex projective spaces*, Bull. London Math. Soc., 16 (1984), pp. 81–121, <https://doi.org/10.1112/blms/16.2.81>.
- [24] F. C. KLEBANER, *Introduction to Stochastic Calculus with Applications*, World Scientific, 2012.
- [25] L. KÜHNEL AND S. SOMMER, *Computational anatomy in Theano*, in Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, Springer, 2017, pp. 164–176.
- [26] L. KÜHNEL, S. SOMMER, AND A. ARNAUDON, *Differential geometry and stochastic dynamics with deep learning numerics*, Appl. Math. Comput., 356 (2019), pp. 411–437.
- [27] J.-L. MARCHAND, *Conditioning Diffusions with Respect to Partial Observations*, preprint, <https://arxiv.org/abs/1105.1608>, 2011.
- [28] B. MARKUSSEN, *Large deformation diffeomorphisms with application to optic flow*, Comput. Vis. Image Underst., 106 (2007), pp. 97–105, <https://doi.org/10.1016/j.cviu.2005.09.006>.
- [29] S. MARSLAND AND T. SHARDLOW, *Langevin equations for landmark image registration with uncertainty*, SIAM J. Imaging Sci., 10 (2017), pp. 782–807, <https://doi.org/10.1137/16M1079282>.
- [30] M. MIDER, M. SCHAUER, AND F. VAN DER MEULEN, *Continuous-discrete smoothing of diffusions*, Electron. J. Stat., 15 (2021), pp. 4295–4342, <https://arxiv.org/abs/1712.03807>.
- [31] G. N. MILSTEIN, Y. M. REPIN, AND M. V. TRETYAKOV, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal., 40 (2002), pp. 1583–1604, <https://doi.org/10.1137/S0036142901395588>.
- [32] X. NIU, J. CUI, J. HONG, AND Z. LIU, *Explicit pseudo-symplectic methods for stochastic Hamiltonian systems*, BIT, 58 (2018), pp. 163–178, <https://doi.org/10.1007/s10543-017-0668-7>.
- [33] B. ØKSENDAL, *Stochastic differential equations*, in Stochastic Differential Equations, Springer, 2003, pp. 65–84.
- [34] O. PAPASPILIOPOULOS AND G. ROBERTS, *Importance sampling techniques for estimation of diffusion models*, in Statistical Methods for Stochastic Differential Equations, Monographs on Statistics and Applied Probability, Chapman and Hall, 2012, pp. 311–337.
- [35] J. REVELS, M. LUBIN, AND T. PAPAMARKOU, *Forward-Mode Automatic Differentiation in Julia*, preprint, <https://arxiv.org/abs/1607.07892>, 2016.
- [36] G. O. ROBERTS AND O. STRAMER, *On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm*, Biometrika, 88 (2001), pp. 603–621, <https://doi.org/10.1093/biomet/88.3.603>.
- [37] L. C. G. ROGERS AND D. WILLIAMS, *Diffusions, Markov Processes, and Martingales. Vol. 2, Itô Calculus*, reprint of the 2nd (1994) edition, Cambridge University Press, Cambridge, UK, 2000.
- [38] M. SCHAUER, F. VAN DER MEULEN, AND H. VAN ZANTEN, *Guided proposals for simulating multi-dimensional diffusion bridges*, Bernoulli, 23 (2017), pp. 2917–2950, <https://doi.org/10.3150/16-BEJ833>.
- [39] S. SOMMER, *Probabilistic approaches to geometric statistics: Stochastic processes, transition distributions, and fiber bundle geometry*, in Riemannian Geometric Statistics in Medical Image Analysis, X. Pennec, S. Sommer, and T. Fletcher, eds., Academic Press, 2020, pp. 377–416, <https://doi.org/10.1016/B978-0-12-814725-2.00018-2>.
- [40] S. SOMMER, A. ARNAUDON, L. KÜHNEL, AND S. JOSHI, *Bridge simulation and metric estimation on landmark manifolds*, in Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics, Lecture Notes in Comput. Sci., Springer, 2017, pp. 79–91, [https://doi.org/10.1007/978-3-319-67675-3\\_8](https://doi.org/10.1007/978-3-319-67675-3_8).
- [41] M. B. STEGMANN, R. FISHER, AND B. K. ERSBØLL, *Extending and applying active appearance models for automated, high precision segmentation in different image modalities*, Scandinavian Conference on Image Analysis, 2001, pp. 90–97.
- [42] A. TROUVE AND F.-X. VIALARD, *Shape splines and stochastic shape evolutions: A second order point of view*, Quart. Appl. Math., 70 (2012), pp. 219–251, <https://doi.org/10.1090/S0033-569X-2012-01250-4>.
- [43] F. VAN DER MEULEN AND M. SCHAUER, *Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals*, Electron. J. Stat., 11 (2017), pp. 2358–2396, <https://doi.org/10.1214/17-EJS1290>.



- [44] F. VAN DER MEULEN AND M. SCHAUER, *BridgeLandmarks* 0.4, 2021, <https://doi.org/doi:10.5281/zenodo.5101751>.
- [45] F.-X. VIALARD, *Extension to infinite dimensions of a stochastic second-order model associated with shape splines*, *Stochastic Process. Appl.*, 123 (2013), pp. 2110–2157, <https://doi.org/10.1016/j.spa.2013.01.012>.
- [46] L. YOUNES, *Shapes and Diffeomorphisms*, Springer, 2010.
- [47] L. YOUNES, F. ARRATE, AND M. I. MILLER, *Evolutions equations in computational anatomy*, *NeuroImage*, 45 (2009), pp. S40–S50, <https://doi.org/10.1016/j.neuroimage.2008.10.050>.