

# Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals

Frank van der Meulen

*Delft Institute of Applied Mathematics (DIAM)  
Delft University of Technology  
Mekelweg 4, 2628 CD Delft  
The Netherlands  
e-mail: [f.h.vandermeulen@tudelft.nl](mailto:f.h.vandermeulen@tudelft.nl)*

and

Moritz Schauer

*Mathematical Institute  
Leiden University  
P.O. Box 9512, 2300 RA Leiden  
The Netherlands  
e-mail: [m.r.schauer@math.leidenuniv.nl](mailto:m.r.schauer@math.leidenuniv.nl)*

**Abstract:** Estimation of parameters of a diffusion based on discrete time observations poses a difficult problem due to the lack of a closed form expression for the likelihood. From a Bayesian computational perspective it can be casted as a missing data problem where the diffusion bridges in between discrete-time observations are missing. The computational problem can then be dealt with using a Markov-chain Monte-Carlo method known as data-augmentation. If unknown parameters appear in the diffusion coefficient, direct implementation of data-augmentation results in a Markov chain that is reducible. Furthermore, data-augmentation requires efficient sampling of diffusion bridges, which can be difficult, especially in the multidimensional case.

We present a general framework to deal with with these problems that does not rely on discretisation. The construction generalises previous approaches and sheds light on the assumptions necessary to make these approaches work. We define a random-walk type Metropolis-Hastings sampler for updating diffusion bridges. Our methods are illustrated using guided proposals for sampling diffusion bridges. These are Markov processes obtained by adding a guiding term to the drift of the diffusion. We give general guidelines on the construction of these proposals and introduce a time change and scaling of the guided proposal that reduces discretisation error. Numerical examples demonstrate the performance of our methods.

**MSC 2010 subject classifications:** Primary 62M05, 60J60; secondary 62F15, 65C05.

**Keywords and phrases:** Multidimensional diffusion bridge, data augmentation, discretisation of path integral, linear process, innovation process, non-centred parametrisation, FitzHugh-Nagumo model.

Received July 2016.

## 1. Introduction

In this article we discuss a novel approach for estimating an unknown parameter  $\theta \in \Theta$  of the drift and the diffusion coefficient of a diffusion process

$$dX_t = b_\theta(t, X_t) dt + \sigma_\theta(t, X_t) dW_t, \quad X_0 = u \quad (1.1)$$

which is observed discretely in time. Here  $b_\theta: \mathbb{R} \times \mathbb{R}^d$  denotes the drift function,  $a_\theta = \sigma_\theta \sigma_\theta'$  is the diffusion function, where  $\sigma_\theta: \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d'}$ , and  $W$  is a  $d'$ -dimensional Wiener process. The observation times will be denoted by  $t_0 = 0 < t_1 < \dots < t_n = T$  and the corresponding observations by  $x_i = X_{t_i}$ .

Estimation of  $\theta$  in this setting has attracted much attention during the past decade. Here we restrict attention to estimation within the Bayesian paradigm. From a theoretical perspective, results on posterior consistency have been proved in Van der Meulen and Van Zanten (2013) and Gugushvili and Spreij (2012). The associated computational problem is the object of study here. Two review articles that include many references on this topic are Van Zanten (2013) and Sørensen (2004).

The main difficulty in estimation for discretely observed diffusion processes is the lack of a closed form expression for transition densities, making the likelihood intractable. If the diffusion path is observed continuously, then estimation becomes easier as for a fully observed diffusion path the likelihood is available in closed form (and parameters appearing in the diffusion coefficient can be determined from the quadratic variation of the process). This naturally suggests to study the computational problem within a missing data framework, treating the unobserved path segments between two succeeding observation times as missing data. This setup dates back to at least Pedersen (1995), who used it to obtain simulated maximum likelihood estimates for  $\theta$ . Within the Bayesian computational problem, the resulting Markov-Chain-Monte-Carlo algorithm is known as data-augmentation and was introduced in this context by Eraker (2001), Elerian, Chib and Shephard (2001) and Roberts and Stramer (2001). This algorithm is a special form of the Gibbs sampler which iterates the following steps:

1. draw missing segments, conditional on  $\theta$  and the observed discrete time data;
2. draw from the distribution of  $\theta$ , conditional on the “full data”.

Here, by “full data” we mean the path formed by the drawn segments joined at the observation times. The algorithm can be initialised by either interpolating the discrete time data or choosing an initial value for  $\theta$ . We now discuss the major challenges for the outlined algorithm together with various solutions that have been proposed in the literature.

*Challenge 1: generating “good” proposals for the missing segments.* The problem of simulating diffusion bridges has received a lot of attention over the past 15 years. Vastly different techniques have been proposed, including (i) single site Gibbs updating of the missing segments locally on a discrete grid (Eraker (2001)), (ii) independent Metropolis-Hastings steps using as a proposal a

Laplace approximation to the conditional distribution obtained by Euler approximation (Elerian, Chib and Shephard (2001)), (iii) forward simulated processes derived from representations of the Brownian bridge in discrete time (Durham and Gallant (2002)), (iv) coupling arguments (Bladt and Sørensen (2014) and Bladt and Sørensen (2015)), (v) a constrained sequential Monte Carlo algorithm with a resampling scheme guided by backward pilots (Lin, Chen and Mykland (2010)), and (vi) exact simulation (Beskos et al. (2006)).

Delyon and Hu (2006) extended the work of Durham and Gallant (2002) to a continuous time setup and derived an innovative proposal process taking the drift of the target diffusion into account. In case the diffusion coefficient is constant this proposal was proposed earlier in Clark (1990). The basic idea consists of superimposing an additional term to the drift of the unconditioned diffusion to guide the process towards the endpoint. Such proposals are termed *guided proposals* and have the advantage that only forward simulation of an SDE is required. More precisely, the drift of the proposal that hits  $v \in \mathbb{R}^d$  at time  $T$  equals  $b^\circ(t, x) = \lambda b(t, x) + (v - x)/(T - t)$ , where either  $\lambda = 0$  or  $\lambda = 1$  is chosen. If  $\lambda = 0$ , the guiding term  $(v - x)/(T - t)$  matches with the drift of the SDE for a Brownian bridge, which indeed has drift 0. However, this proposal has the drawback that it is independent of the drift  $b$  of the diffusion. If  $\lambda = 1$ , the guiding term depends on  $b$  and consequently there is a potential mismatch between the drift and guiding term. In both cases (i.e.  $\lambda = 0$  and  $\lambda = 1$ ) there can be a substantial mismatch between the proposals and true bridges, rendering low acceptance rates in an MH-sampler.

In Schauer, Van der Meulen and Van Zanten (2017) a general class of proposal processes for simulating diffusion bridges was introduced. The proposals in Schauer, Van der Meulen and Van Zanten (2017) do take the drift of the target diffusion into account, but in a way different from Delyon and Hu (2006). As a result, these proposals can substantially reduce the mismatch of drift and guiding term, because they allow for more flexibility in choosing an appropriate guiding term to pull the process towards the endpoint in the right manner. An example of the advantage of this approach is given in the introduction of Schauer, Van der Meulen and Van Zanten (2017). General guidelines to exploit the added flexibility are addressed in this paper.

For implementation purposes, any proposal has to be evaluated on a finite number of grid points. As the pulling term added to the drift for guided proposals has a singularity near the endpoint, special care is needed in choosing a discretisation method. More importantly, integrals that appear in the acceptance probability of bridges potentially suffer from this problem as well. In this paper we introduce a time change and scaling of the proposal process that deals with these problems.

*Challenge 2: handling unknown parameters appearing in the diffusion coefficient.* As pointed out by Roberts and Stramer (2001), the data augmentation algorithm degenerates if  $\theta$  appears in the diffusion coefficient as the quadratic variation of the full data  $\int_0^T a_\theta(t, X_t) dt$  forces the conditional distribution for the next iterate for  $\theta$  to be degenerate at the current value. Hence, iterates of

$\theta$  remain stuck at their initial value. The problem was solved in a discretised setting by both Chib, Pitt and Shephard (2004) and Golightly and Wilkinson (2010). Rather than updating  $\theta$  conditional on the discretised diffusion bridge, they proposed updating  $\theta$  conditional on the increments of the Brownian motion driving the discretised diffusion bridge. This decouples the tight dependence between  $\theta$  and the diffusion bridge. However, as Stramer and Bognar (2011) point out “While the promising GW approach can be applied to a large class of diffusions, it is not yet rigorously justified in the literature.” Put differently, whereas the GW (=Golightly-Wilkinson) approach works in the discretised setup, it gives no guarantee that it also works in the limit where the discretisation level tends to zero.

In the continuous-time framework a solution to the aforementioned problem was given in Roberts and Stramer (2001) for one-dimensional diffusions. It was extended to reducible multivariate diffusions (diffusions that can be transformed to have unit diffusion coefficient) in Beskos et al. (2006) and Sermaidis et al. (2013). The basic idea is that the laws of the bridge proposals can be understood as parametrised push forwards of the law of an underlying random process common to all models with different parameters  $\theta$ . This is naturally the case for proposals defined as solutions of stochastic differential equations and the driving Brownian motion can be taken as such an underlying random process. If  $X^*$  denotes a missing segment given that the parameter equals  $\theta$ , the main idea consists of finding a map  $g$  and a process  $Z$  such that  $X^* = g(\theta, Z)$ . In a more general set-up, decouplings of similar forms are discussed under the keyword *non-centred parameterisation* (Papaspiliopoulos, Roberts and Sköld (2003)). The process  $Z$  will be called the “innovation process” (analogous to terminology used in Chib, Pitt and Shephard (2004) and Golightly and Wilkinson (2010)). Whereas in case  $\sigma_\theta = \theta$  the construction is rather easy, in general proving existence of the map  $g$  and process  $Z$  is subtle and this forms an important topic of this paper. We postpone a detailed discussion to Sections 2 and 3.

A first attempt of finding a non-centred parameterisation in continuous time in a general setting was undertaken in Fuchs (2013) (in particular section 7.4). Fuchs (2013) works in the setting of Delyon and Hu (2006), so it is assumed that the diffusion coefficient  $\sigma$  is invertible and the diffusion is time-homogeneous. While the results in Fuchs (2013) are formulated in continuous time, the derivation involves heuristic arguments via the Lebesgue densities of the finite dimensional distributions. A recent work is Papaspiliopoulos, Roberts and Stramer (2013). In their approach the missing data is initially considered in continuous time using Delyon and Hu (2006) bridge proposals, but the degeneracy problem is tackled only *after* discretisation. An important contribution of this paper consists of recognising that bridge simulation is key for building data augmentations algorithms and cancellation of the intractable transition density from the augmented target.

What is the essential structure behind those different approaches and how can the underlying transformations be handled in continuous time without resorting to discretisation? Are these techniques tied to certain proposals, for example the proposal processes in Delyon and Hu (2006), or are they valid for other proposal

processes as well? And can conditions such as invertibility of  $\sigma$  be relaxed and is it essential that the diffusion is time-homogeneous? Part of this paper consists of answering these questions in a rigorous way. As a result of this, in our setting it is evident how to replace the independence sampler for  $Z$  (which updates the diffusion bridges) by a random-walk type update on the process  $Z$  in a straightforward way.

### 1.1. Contribution

In this article we present a general framework for Bayesian estimation of discretely observed diffusion processes that satisfactorily deals with both aforementioned challenges. Our approach reveals the conditions necessary for obtaining an irreducible Markov chain that samples from the posterior (after burnin). We show that the algorithm does not suffer from the degeneracy problem in case unknown parameters appear in the diffusion coefficient, not even in the continuous time setup. The procedure can be seen as extension and unification of previous approaches within a *continuous time* framework. For example the results of the rather complicated heuristics in Section 7.4 of Fuchs (2013) appear as a special case of our work. Specific features of our approach include:

- We use in each data augmentation step “adapted” bridge proposals which take both the drift *and the value of  $\theta$  at that particular iteration* into account. Hence, at each iteration, the pulling term depends on  $\theta$ , a feature which is unavailable using proposals as in Delyon and Hu (2006). Especially in the multivariate case, the additional freedom in devising good proposals is crucial for obtaining a feasible MCMC procedure. The possibility to exploit special features of the drift function to achieve high acceptance rates makes this approach interesting for practitioners. This is illustrated with a practical example in Section 7.2.
- We provide specialised algorithms in case the drift is of the form  $b_\theta = \sum_{i=1}^N \theta_i \varphi_i$  for known functions  $\varphi_1, \dots, \varphi_N$  (Cf. algorithms 2 and 3).
- The innovation process is defined using the proposal process. As a result, in our algorithm (Cf. algorithm 1), the innovations actually never need to be computed. This implies that our method can also cope with the case where  $\sigma$  is not a square matrix (which is not the case for example in Fuchs (2013)).
- We illustrate our work using linear guided proposals as introduced in Schauer, Van der Meulen and Van Zanten (2017) and the proposals introduced in Delyon and Hu (2006). In section 4.4 we give general guidelines on the construction of these proposals. In section 7.2 we show that not taking into account the drift of the diffusion can lead to extremely small acceptance probabilities for bridges.
- Though we derive all our results in a continuous time setup, for implementation purposes integrals in likelihood ratios and solutions to stochastic differential equations need to be approximated on a finite grid. As the drift of our proposal bridges has a singularity near its endpoint, we introduce a

time change and scaling that allows for numerically accurate discretisation and evaluation of the likelihood.

The approach with linear guided proposals can be extended to the case of partially observed diffusions, where for example some components of the diffusion are unobserved. Though the problem becomes much harder, the underlying structure for constructing an algorithm is the same. For details we refer to Van der Meulen and Schauer (2016).

## 1.2. Outline

In Section 2 we clarify the aforementioned difficulties in a toy example. Here, we set some general notation and introduce some key ideas used throughout. In section 3 we precisely state our algorithms and introduce the concept of a feasible proposal. In Section 4 we show that both the proposals from Delyon and Hu (2006) and Schauer, Van der Meulen and Van Zanten (2017) are feasible. In Section 4.4 we give guidelines on constructing a guiding term for the proposals from Schauer, Van der Meulen and Van Zanten (2017). Numerical discretisation issues and the computational complexity of the proposed algorithms are discussed in sections 5 and 6 respectively. Numerical examples are given in Section 7. We end with a short section on summary and discussion. The appendix contains a few postponed proofs.

## 2. A toy problem

In this section we consider a toy example to illustrate some key ideas to solve the aforementioned problems with a simple data-augmentation algorithm. The type of reparameterisation introduced shortly is not new, and has appeared for example in Roberts and Stramer (2001). The goal here is to introduce key ideas and point out some of its potential shortcomings in more complex problems. Furthermore, later on we will deal with more difficult cases and this toy example allows us to sequentially build up an appropriate framework for that. We consider the diffusion process

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

where  $b$  is a known drift function and  $\theta \in \Theta$  an unknown scaling parameter. We assume  $\theta$  is equipped with a prior distribution  $\pi_0(\theta)$  and only one observation  $X_T = v$  at time  $T$  is available. We aim to draw from the posterior  $\pi(\theta | X_T)$ . The diffusion process conditioned on  $X_T = v$  is a diffusion process itself. Denote by  $X^*$  the conditioned diffusion path  $(X_t, t \in (0, T))$  (conditional on  $X_T = v$ ). Suppose we wish to iterate a data-augmentation algorithm and the current iterate is given by  $(X^*, \theta)$ .

*Updating  $X^*$ :* For almost all choices of  $b$ , there is no direct way of simulating  $X^*$ . Instead, one can first generate a proposal bridge  $X^\circ$  and accept with MH-acceptance probability. As an easy tractable example we choose to take

$$dX_t^\circ = \frac{v - X_t^\circ}{T - t} dt + \theta dW_t, \quad X_0^\circ = u \quad (2.1)$$

where  $W$  is a Brownian Motion on  $[0, T]$ .

Denote the laws of  $X^\circ$  and  $X^*$  (viewed as Borel measures on  $C([0, T], \mathbb{R}^d)$ ) by  $\mathbb{P}^\circ$  and  $\mathbb{P}^*$  respectively. We have

$$\frac{d\mathbb{P}_\theta^*}{d\mathbb{P}_\theta^\circ}(X^\circ) = \frac{\tilde{p}_\theta(0, u; T, v)}{p_\theta(0, u; T, v)} \Psi_\theta(X^\circ), \quad (2.2)$$

with

$$\Psi_\theta(X^\circ) = \exp\left(\theta^{-2} \int_0^T b(X_s^\circ) dX_s^\circ - \frac{1}{2} \theta^{-2} \int_0^T b(X_s^\circ)^2 ds\right).$$

Here,  $p$  denotes the transition densities of the process  $X$  and  $\tilde{p}(0, u; T, v) = \varphi(v; u, \theta^2 T)$  (the density of the  $N(u, \theta^2 T)$ -distribution, evaluated at  $v$ ). Absolute continuity is a consequence of Girsanov's theorem applied to the unconditioned processes and the abstract Bayes' formula. Now the MH-step consists of generating a proposal  $X^\circ$  and accepting it with probability  $1 \wedge (\Psi_\theta(X^\circ)/\Psi_\theta(X^*))$  (the ratio of transition densities just acts as a proportionality constant here).

*Updating  $\theta$ :* As explained in the introduction, taking the missing segment as missing data yields the Metropolis-Hastings algorithm reducible. To deal with this problem, note that by equation (2.1), there exists a mapping  $g$  such that  $X^\circ = g(\theta, W)$ . Define the process  $Z$  by the relation

$$X^* = g(\theta, Z). \quad (2.3)$$

Now that  $Z$  is defined, rather than drawing from the distribution of  $\theta$  conditional on  $(X_0 = u, X_T = v, X^*)$  we will sample from the the distribution of  $\theta$  conditional on  $(X_0 = u, X_T = v, Z)$ . This means that we augment the discrete time observations with  $Z$  instead of  $X^*$ . Denote the laws of  $Z$  and  $W$  by  $\mathbb{Z}_\theta$  and  $\mathbb{W}$  respectively. Suppose the current iterate is  $(\theta, Z)$ , where  $Z$  can be extracted from  $\theta$  and  $X^*$  by means of equation (2.3). The following diagram summarises the notation introduced:

$$\begin{array}{c} \text{Process} \\ \hline \mathbb{Z}_\theta \quad \mathbb{P}_\theta^* \quad \mathbb{W} \quad \mathbb{P}_\theta^\circ \end{array} \left\| \begin{array}{c} Z \xrightarrow{g(\theta, \cdot)} X^* \\ W \xrightarrow{g(\theta, \cdot)} X^\circ \end{array} \right. \quad (2.4)$$

For updating  $\theta$  we propose a value  $\theta^\circ$  from some proposal distribution  $q(\cdot | \theta)$  and accept the proposal with probability  $\min(1, A)$ , where

$$A = \frac{\pi_0(\theta^\circ)}{\pi_0(\theta)} \frac{p_{\theta^\circ}(0, u; T, v)}{p_\theta(0, u; T, v)} \frac{d\mathbb{Z}_{\theta^\circ}}{d\mathbb{Z}_\theta}(Z) \frac{q(\theta | \theta^\circ)}{q(\theta^\circ | \theta)}. \quad (2.5)$$

Here, we have implicitly assumed that  $\mathbb{Z}_{\theta^\circ}$  and  $\mathbb{Z}_\theta$  are equivalent, which is indeed the case as we have

$$\frac{d\mathbb{Z}_{\theta^\circ}}{d\mathbb{Z}_\theta}(Z) = \frac{d\mathbb{Z}_{\theta^\circ}}{d\mathbb{W}}(Z) \Big/ \frac{d\mathbb{Z}_\theta}{d\mathbb{W}}(Z) = \frac{d\mathbb{P}_{\theta^\circ}^*}{d\mathbb{P}_\theta^\circ}(g(\theta^\circ, Z)) \Big/ \frac{d\mathbb{P}_\theta^*}{d\mathbb{P}_\theta^\circ}(g(\theta, Z))$$

and thus results from absolute continuity of  $\mathbb{P}_\theta^*$  and  $\mathbb{P}_\theta^\circ$ . By equation (2.2), we now get

$$\frac{dZ_{\theta^\circ}}{dZ_\theta}(Z) = \frac{p_\theta(0, u; T, v) \tilde{p}_{\theta^\circ}(0, u; T, v) \Psi_{\theta^\circ}(g(\theta^\circ, Z))}{p_{\theta^\circ}(0, u; T, v) \tilde{p}_\theta(0, u; T, v) \Psi_\theta(g(\theta, Z))}.$$

Substituting this expression into equation (2.5) yields

$$A = \frac{\pi_0(\theta^\circ) \tilde{p}_{\theta^\circ}(0, u; T, v) \Psi_{\theta^\circ}(g(\theta^\circ, Z))}{\pi_0(\theta) \tilde{p}_\theta(0, u; T, v) \Psi_\theta(g(\theta, Z))} \frac{q(\theta | \theta^\circ)}{q(\theta^\circ | \theta)} \tag{2.6}$$

and all terms containing the unknown transition density cancel. In Section 4 we will show that cancellation of  $p$  from the acceptance probability holds much more generally.

The feasibility and efficiency of this algorithm is crucially determined by choice of the transition kernel  $q$  and proposal process  $X^\circ$ . We focus on an appropriate choice of  $X^\circ$ , though in section 3.3 we give guidelines on appropriate choice of  $q$  if the drift possesses a specific structure with respect to  $\theta$ .

### 3. Proposed MCMC algorithms

Our starting point is that under weak assumptions the target diffusion bridge  $X^*$  from  $u$  at time  $t = 0$  to  $v$  at time  $t = T$  is characterised as the solution to the SDE

$$dX_t^* = b_\theta^*(t, X_t^*) dt + \sigma_\theta(t, X_t^*) dW_t, \quad X_0^* = u, \quad t \in [0, T], \tag{3.1}$$

where

$$b_\theta^*(t, x) = b_\theta(t, x) + a_\theta(t, x) \nabla_x \log p_\theta(t, x; T, v) \tag{3.2}$$

and  $a = \sigma\sigma'$ . Here the transition density of  $X$  is denoted by  $p_\theta$  and  $p_\theta(t, x; T, v)$  is the density of the process starting in  $x$  at time  $t$ , ending in  $v$  at time  $T$ .

#### 3.1. Innovation process

Direct forward simulation of  $X^*$  is hardly ever possible, as  $p$  is intractable. Instead, we propose to simulate a process  $X^\circ$  with induced law that is absolutely continuous with respect to that of  $X^*$ . More precisely, we assume the *proposal process*  $X^\circ$  satisfies the SDE

$$dX_t^\circ = b_\theta^\circ(t, X_t^\circ) dt + \sigma_\theta(t, X_t^\circ) dW_t, \quad X_0^\circ = u, \quad t \in [0, T] \tag{3.3}$$

We now describe a general parametrisation to decouple the dependence between the latent paths of the diffusion between discrete time observations and the parameter  $\theta$ .

The following proposition is key to the definition of the map  $g$ . In its statement we refer to the canonical setup on which an exact SDE can be solved, details are in section V.10 of Rogers and Williams (2000).

**Proposition 3.1.** Assume

- the SDEs for  $X^\circ$  and  $X^*$  are pathwise exact (in the sense of definition V-9.4 of Rogers and Williams (2000));



- there exists a strong solution for the SDE for  $X^\circ$  (in the sense of definition V.10.9 of Rogers and Williams (2000)) jointly measurable with respect to starting point, parameter and path  $W$ ;
- $\mathbb{P}^\circ$  and  $\mathbb{P}^*$  are absolutely continuous.

Then there exists a map  $g$  and a Wiener process  $W$  such that  $X^\circ = g(\theta, W)$  on the canonical setup. Furthermore, there exists a process  $Z$  such that  $X^* = g(\theta, Z)$ . The process  $Z$  satisfies the SDE

$$dZ_t = \mu_\theta(Z_t) dt + dW_t \quad (3.4)$$

where the map  $\mu_\theta$  satisfies

$$\sigma_\theta(t, x)\mu_\theta(t, x) = b_\theta^*(t, x) - b_\theta^\circ(t, x). \quad (3.5)$$

Moreover,

$$dX_t^* = b_\theta^\circ(t, X_t^*) dt + \sigma_\theta(t, X_t^*) dZ_t, \quad X_0^* = u, \quad t \in [0, T]. \quad (3.6)$$

*Proof.* Denote the law of  $W$  by  $\mathbb{W}$ . Existence of  $g$  such that  $X^\circ = g(\theta, W)$  is implied by existence of a strong solution for the SDE for  $X^\circ$ . If  $Y$  satisfies

$$dY_t = b_\theta^\circ(t, Y_t) dt + \sigma_\theta(t, Y_t) dW_t, \quad Y_0 = u, \quad t \in [0, T]$$

then  $Y = g(\theta, W)$ . Define

$$L_\theta = \exp \left( \int_0^T \mu_\theta(t, Y_t) dW_t - \int_0^T \mu_\theta(t, Y_t)^2 dt \right)$$

and assume for the moment that  $E^{\mathbb{W}} L_\theta = 1$ . Define the measure  $\mathbb{Z}_\theta$  by  $d\mathbb{Z}_\theta = L_\theta d\mathbb{W}$ . By Girsanov's theorem, it follows that the process  $Z$  defined by equation (3.4) is a Brownian Motion under the measure  $\mathbb{Z}_\theta$ . If we define  $Y$  by  $Y = g(\theta, Z)$  then

$$dY_t = b_\theta^\circ(t, Y_t) dt + \sigma_\theta(t, Y_t) dZ_t, \quad Y_0 = u, \quad t \in [0, T]$$

under  $\mathbb{Z}_\theta$ . Plugging (3.4) into this equation shows that

$$dY_t = b_\theta^*(t, Y_t) dt + \sigma_\theta(t, Y_t) dW_t, \quad Y_0 = u, \quad t \in [0, T].$$

By pathwise uniqueness  $Y = X^*$  up to indistinguishability under  $\mathbb{W}$  (because  $W$  is a Wiener process). Hence,  $X^* = g(\theta, Z)$  and (3.6) follows. We have

$$\frac{d\mathbb{Z}_\theta}{d\mathbb{W}}(\cdot) = \frac{d\mathbb{P}_\theta^*}{d\mathbb{P}_\theta^\circ}(g(\theta, \cdot)) \quad (3.7)$$

and henceforth existence of  $\mu_\theta$  such that  $E^{\mathbb{W}} L_\theta = 1$  follows from our assumption that  $\mathbb{P}_\theta^*$  and  $\mathbb{P}_\theta^\circ$  are absolutely continuous.  $\square$

We refer to the process  $Z$  as the *innovation process* corresponding to  $X^*$  (by analogy of the terminology of Golightly and Wilkinson (2008) and Chib, Pitt and Shephard (2004)). Clearly,  $X^*$  is related to  $Z$  just like  $X^\circ$  is related to  $W$ . Note however that while the law of  $W$  does not depend on  $\theta$  under  $\mathbb{W}$ , the law of  $Z$  does depend on  $\theta$  under  $\mathbb{W}$ .

In the following we will denote the Radon-Nikodym derivative between  $\mathbb{P}^\circ$  and  $\mathbb{P}^*$  by  $\Phi$ :

$$\Phi_\theta := \frac{d\mathbb{P}_\theta^*}{d\mathbb{P}_\theta^\circ}.$$

### 3.2. Algorithm

In this section we present an algorithm to sample from the posterior of  $\theta$  given the discrete observations  $\mathcal{D} = \{X_0 = u, X_{t_1} = x_1, \dots, X_{t_n} = x_n\}$ . Denote the prior density on  $\theta$  by  $\pi_0$  and let  $q(\theta^\circ | \theta)$  be the density for proposing  $\theta^\circ$  given the current value  $\theta$ . The idea is to define a Metropolis–Hastings sampler on  $(\theta, Z)$  instead of  $(\theta, X^*)$  where  $Z$  is the innovation process from the previous section.

More precisely, we construct a Markov chain for  $(\theta, (Z_i)_{1 \leq i \leq n})$ , where each  $Z_i$  is an innovation process corresponding to the bridge  $X_i^*$  connecting observation  $x_{i-1}$  to  $x_i$ .

#### Algorithm 1.

1. **Initialisation.** Choose a starting value for  $\theta$  and sample  $i = 1, \dots, n$  Wiener processes  $W_i$  and set  $Z_i = W_i$ .
2. **Update  $Z | (\theta, \mathcal{D})$ .** Independently, for  $1 \leq i \leq n$  do
  - (a) Sample a Wiener process  $Z_i^\circ$ .
  - (b) Sample  $U \sim \mathcal{U}(0, 1)$ . Compute

$$A_1 = \frac{\Phi_\theta(g(\theta, Z_i^\circ))}{\Phi_\theta(g(\theta, Z_i))}.$$

Set

$$Z_i := \begin{cases} Z_i^\circ & \text{if } U \leq A_1 \\ Z_i & \text{if } U > A_1 \end{cases}.$$

3. **Update  $\theta | (Z, \mathcal{D})$ .**
  - (a) Sample  $\theta^\circ \sim q(\cdot | \theta)$ .
  - (b) Sample  $U \sim \mathcal{U}(0, 1)$ . Compute

$$A_2 = \frac{q(\theta | \theta^\circ) \pi_0(\theta^\circ)}{q(\theta^\circ | \theta) \pi_0(\theta)} \prod_{i=1}^n \frac{p_{\theta^\circ}(t_{i-1}, x_{i-1}; t_i, x_i) \Phi_{\theta^\circ}(g(\theta^\circ, Z_i))}{p_\theta(t_{i-1}, x_{i-1}; t_i, x_i) \Phi_\theta(g(\theta, Z_i))}$$

Set

$$\theta := \begin{cases} \theta^\circ & \text{if } U \leq A_2 \\ \theta & \text{if } U > A_2 \end{cases}.$$

4. Repeat steps (2) and (3).

Note that in none of these steps we need to compute innovations  $Z$  from  $X^*$ . This is a consequence of adapting the definition of the innovations to the bridge proposals being used.

In step (2) an independent Metropolis-Hastings step is used. Instead, one can also propose  $Z^\circ$  based on the current value of  $Z$  in the following way

$$Z_t^\circ = \sqrt{\rho}Z_t + \sqrt{1-\rho}W_t, \quad (3.8)$$

where  $\rho \in [0, 1)$  and  $W$  is a Wiener process under  $\mathbb{W}$  that is independent of  $Z$ . In this case

$$A_1 = \frac{(dZ_\theta/d\mathbb{W})(Z)}{(dZ_\theta/d\mathbb{W})(Z^\circ)} \frac{dQ_\rho}{dQ_\rho^T}(Z^\circ, Z),$$

where  $Q_\rho(x, y) = Q_\rho^T(y, x)$ . Here we use the general formulation of the Metropolis-Hastings algorithm as explained in Tierney (1998). The second term equals one by symmetry of  $Q(\cdot, \cdot)$ . This implies that the acceptance probability in step 2(b) remains the same.

The proposal in (3.8) is an example of a preconditioned Crank-Nicolson (pCN) scheme (Cf. Cotter et al. (2013) and Beskos et al. (2008)) and seems to have been introduced first in Neal (1999).

**Remark 3.2.** Different proposals can be obtained by varying  $b^\circ$  in (3.3) and it is clear that the mapping  $g$  varies accordingly. A good choice obviously affects the acceptance probability of step 2 in algorithm 1. However, it affects the acceptance probability of step 3 as well as this step is a joint update of  $(\theta, X^*)$ . This implies that a proposal  $\theta^\circ$  in step 3 which is “good” (in the sense of being like a draw from the posterior of  $\theta$ ), may nevertheless be rejected if the mapping  $g$  is such that  $g(\theta^\circ, Z)$  does not resemble a bridge with drift and diffusion coefficient indexed by  $\theta^\circ$ . Ideally, one would take  $g = g_{\text{opt}}$ , where  $g_{\text{opt}}$  is defined by the relation  $X^* = g_{\text{opt}}(\theta, W)$ , with  $W$  denoting a Wiener process.

**Theorem 3.3.** *Suppose  $q(\theta^\circ | \theta)$  is almost everywhere strictly positive on the support of the prior for  $\theta$ . Then the chain induced by algorithm 1 is irreducible.*

*Proof.* Step 2 constitutes a step of a MH-sampler with independent proposals. The expression for  $A_1$  follows directly from equation (3.7). The expression for  $A_2$  in step 3 follows in exactly the same way as equation (2.6) was established in the toy-example (Cf. section 2). The remaining observation needed is the following: As  $\Phi_\theta$  is the Radon-Nikodym derivative between two equivalent distributions, it is almost surely strictly positive and finite. Since the transition densities are strictly positive as well, both  $A_1$  and  $A_2$  are strictly positive and the result follows.  $\square$

At first sight, it may seem that algorithm 1 is not of much practical value. First of all, the mapping  $g$  is unknown. However, as any algorithm derived in continuous time ultimately has to be approximated by discretisation, we can choose a discretisation level and compute  $Y = g(\theta, Z)$  on a fine grid by discretising the stochastic differential equation

$$dY_t = b_\theta^\circ(t, Y_t) dt + \sigma_\theta(t, Y_t) dZ_t.$$

Second, it seems impossible to compute the acceptance probabilities in steps 2 and 3 because  $\Phi_\theta$  depends on  $p$  and  $p$  explicitly pops up in the formula for

$A_2$ . However, it turns out that for many choices of  $b^\circ$  the unknown transition density  $p$  only appears as a multiplicative constant in  $\Phi_\theta$  such that it cancels the  $p$  in the expression for  $A_2$ . For future reference, we introduce the following definition.

**Definition 3.4.** We call a proposal  $X^\circ$  as defined in equation (3.3) *feasible* if  $b_\theta^\circ$  is such that both  $A_1$  and  $A_2$  appearing in algorithm 1 do not depend on the transition density  $p$ .

In section 4 we will give examples of classes of feasible proposals.

### 3.3. Partially conjugate series prior for the drift

In this subsection we study specific cases of algorithm 1 when the drift is of the form

$$b_\vartheta(x) = \sum_{i=1}^N \vartheta_i \varphi_i(x) \tag{3.9}$$

where  $\vartheta = (\vartheta_1, \dots, \vartheta_N)$  is an unknown parameter and  $\varphi_1, \dots, \varphi_N$  are known functions on  $\mathbb{R}^d$ . We assume the diffusion coefficient is parametrised by the parameter  $\gamma$ . We denote the vector of all unknown parameters by  $\theta = (\vartheta, \gamma)$  and assume these are assigned independent priors. With slight abuse of notation we use  $\pi_0(\vartheta)$  and  $\pi_0(\gamma)$  to denote the priors on  $\vartheta$  and  $\gamma$  respectively (the argument in parentheses will clarify which prior is meant). In this case it is convenient to choose a conjugate Gaussian prior for the coefficients,  $\vartheta_i \sim \mathcal{N}(0, \xi_i^2)$ , for positive scaling constants  $\xi_i$ . Priors for the drift obtained by specifying a prior distribution on  $\vartheta$  were previously considered in K uchler and S orensen (1997), Bladt and S orensen (2014) and Van der Meulen, Schauer and Van Zanten (2014). Upon completing the square, it follows that the distribution of  $\vartheta$  conditional on  $\gamma$  and the full path  $Y$  of the diffusion is multivariate normal with mean vector  $W_\gamma^{-1} \mu_\gamma$  and covariance matrix  $W_\gamma^{-1}$ . We define for  $k, \ell \in \{1, \dots, d\}$ ,

$$\begin{aligned} \mu_\gamma[k] &= \int_0^T \varphi_k(Y_t)' a_\gamma^{-1}(Y_t) dY_t \\ \Sigma_\gamma[k, \ell] &= \int_0^T \varphi_k(Y_t)' a_\gamma^{-1}(Y_t) \varphi_\ell(Y_t) dt \\ W_\gamma &= \Sigma + \text{diag}(\xi_1^{-2}, \dots, \xi_N^{-2}). \end{aligned}$$

(For a vector  $x \in \mathbb{R}^n$  we denote the  $i$ -th element by  $x[i]$ . To emphasise the dependence on  $Y$  we sometimes also write  $\mu_\gamma(Y)$ ,  $W_\gamma(Y)$  etc). This leads to a natural adaptation of algorithm 1 from section 3.2.

**Algorithm 2.** Steps 1, 2 and 4 as in algorithm 1. Assume that  $\sigma$  is invertible. Step 3 is given by

3.1 **Update**  $\gamma \mid (\vartheta, Z, \mathcal{D})$ .

- (a) Sample  $\gamma^\circ \sim q(\cdot \mid \gamma)$ .

(b) Sample  $U \sim \mathcal{U}(0, 1)$ . Compute

$$A_3 = \frac{q(\gamma \mid \gamma^\circ) \pi_0(\gamma^\circ)}{q(\gamma^\circ \mid \gamma) \pi_0(\gamma)} \times \prod_{i=1}^n \frac{p_{(\gamma^\circ, \vartheta)}(t_{i-1}, x_{i-1}; t_i, x_i) \Phi_{(\gamma^\circ, \vartheta)}(g((\gamma^\circ, \vartheta), Z_i))}{p_{(\gamma, \vartheta)}(t_{i-1}, x_{i-1}; t_i, x_i) \Phi_{(\gamma, \vartheta)}(g((\gamma, \vartheta), Z_i))}$$

Set

$$\gamma := \begin{cases} \gamma^\circ & \text{if } U \leq A_3 \\ \gamma & \text{if } U > A_3 \end{cases}.$$

**3.2 Update**  $\vartheta \mid (\gamma, Z, \mathcal{D})$ .

- (a) Compute  $\mu_g = \mu_\gamma(g((\vartheta, \gamma), Z))$  and  $W_\gamma = W_\gamma(g((\vartheta, \gamma), Z))$ .
- (b) Sample  $\vartheta^\circ \sim \mathcal{N}(W_\gamma^{-1} \mu_\gamma, W_\gamma^{-1})$ .
- (c) Compute  $Z^\circ$  such that  $g((\vartheta^\circ, \gamma), Z^\circ) = g((\vartheta, \gamma), Z)$ . Set  $\vartheta = \vartheta^\circ$  and  $Z = Z^\circ$ .

Note that computation of  $Z^\circ$  in step 3.2(c) requires invertibility of  $\sigma$ .

*Proof.* Suppose  $(\vartheta, \gamma, Z) \sim \pi$ , where  $\pi$  denotes the posterior distribution. Consider the map  $f: (\vartheta, \gamma, Z) \mapsto (\vartheta, \gamma, X^*)$ , where  $X^* = g((\vartheta, \gamma), Z)$ . We show that step 3.2 preserves  $\pi$ . The distribution of  $(\vartheta, \gamma, X^*)$  is the image measure of the posterior distribution  $\pi$  of the tuple  $(\vartheta, \gamma, Z)$  under  $f$  and coincides with the posterior distribution of  $(\vartheta, \gamma, X^*)$ . Denote the image measure of  $\pi$  under  $f$  by  $\pi \circ f^{-1}$ . In steps 3.2(a) and 3.2(b) we apply the mapping  $f$ , followed by a Gibbs step in which we draw  $\vartheta^\circ$  conditional on  $(\gamma, X^*)$ . The latter preserves  $\pi \circ f^{-1}$ . Hence  $(\vartheta^\circ, \gamma, X^*) \sim \pi \circ f^{-1}$ . In step 3.2(c) we compute  $(\vartheta^\circ, \gamma, Z^\circ)$  as pre-image of  $(\vartheta^\circ, \gamma, X^*)$  under  $f$  (this is possible as we assume  $\sigma$  to be invertible). Hence  $(\vartheta^\circ, \gamma, Z^\circ) \sim \pi$ .  $\square$

A variation of this algorithm is obtained in case the drift is of the form specified in equation (3.9) and the diffusion coefficient depends on both  $\vartheta$  and  $\gamma$ . In this case we can update  $\gamma$  just as in algorithm 2. Updating  $\vartheta$  can be done using a random walk type proposal of the form

$$q(\vartheta^\circ \mid \vartheta) \sim N(\vartheta, \alpha V),$$

with  $\alpha$  a positive tuning parameter. Motivated by the covariance matrix of the prior exploited in the case of partial conjugacy we propose to replace  $V$  by  $W_{(\vartheta, \gamma)}^{-1}$ . By this choice, if two components  $\vartheta_i$  and  $\vartheta_j$  are strongly correlated, the proposed local random walk proposals have the same correlation structure, which can improve mixing of the chain.

**Algorithm 3.** The same algorithm as Algorithm 2 without the invertibility assumptions and Step 3.2 replaced by

**3.2' Update**  $\vartheta \mid (\gamma, Z, \mathcal{D})$ .

- (a) Set  $X^* = g(\vartheta, Z)$ .
- (b) Compute  $W_{(\vartheta, \gamma)}$ .
- (c) Sample  $\vartheta^\circ \sim \mathcal{N}(\vartheta, \alpha^2 W_{(\vartheta, \gamma)}^{-1})$ .
- (d) Compute  $W_{(\vartheta^\circ, \gamma)}$ .
- (e) Sample  $U \sim \mathcal{U}(0, 1)$ . Compute

$$A_4 = \frac{|W_{\vartheta^\circ}|^{1/2}}{|W_\vartheta|^{1/2}} \exp\left(-\frac{1}{2\alpha^2}(\vartheta^\circ - \vartheta)'(W_{\vartheta^\circ} - W_\vartheta)(\vartheta^\circ - \vartheta)\right) \times \frac{\pi_0(\vartheta^\circ)}{\pi_0(\vartheta)} \prod_{i=1}^n \frac{p_{(\gamma, \vartheta^\circ)}(t_{i-1}, x_{i-1}; t_i, x_i)}{p_{(\gamma, \vartheta)}(t_{i-1}, x_{i-1}; t_i, x_i)} \frac{\Phi_{(\gamma, \vartheta^\circ)}(g((\gamma, \vartheta^\circ), Z_i))}{\Phi_{(\gamma, \vartheta)}(g((\gamma, \vartheta), Z_i))}.$$

Set

$$\vartheta := \begin{cases} \vartheta^\circ & \text{if } U \leq A_4 \\ \vartheta & \text{if } U > A_4 \end{cases}.$$

The following argument gives some guidance in the choice of  $\alpha$ . If the target distribution is a  $d$ -dimensional Gaussian distribution  $\mathcal{N}_d(\mu, \Sigma)$  and the proposal is of the form  $\vartheta^\circ \sim q(\vartheta^\circ, \vartheta) \sim \mathcal{N}_d(\vartheta, \alpha^2 \Sigma_q)$ , then optimal choices for  $\alpha$  and  $\Sigma_q$  are given by  $\Sigma_q = \Sigma$  and  $\alpha = 2.38/\sqrt{d}$ , cf. Rosenthal (2011). Hence, we will choose  $\alpha = 2.38/\sqrt{\dim(\vartheta)}$ , which corresponds to an average acceptance probability equal to 0.234. Although this procedure will not be optimal for the examples considered, it provides an automatic choice and avoids tedious pilot runs.

#### 4. Feasible proposals

In this section we discuss examples of proposals that enable application of algorithm 1. First we discuss the prerequisites for this in general. Trivially, we should be able to sample a discretised version of the process  $X^\circ$ . This can be done using a discretisation method for stochastic differential equations, such as the Euler-discretisation. Secondly, it is required that the assumptions of proposition 3.1 are satisfied. Third, we need our proposal to be feasible in the sense of definition 3.4. This requires choosing  $b^\circ$  such that  $\Phi_\theta = d\mathbb{P}_\theta^*/d\mathbb{P}_\theta^\circ$  contains the transition density  $p$  solely as a multiplicative factor in the denominator. As  $\theta$  is fixed throughout this section, we drop it temporarily from our notation. It is not too hard to see why  $p$  would only show up as a multiplicative factor in the denominator. Denote the laws of  $X$ ,  $X^\circ$  and  $X^*$  on  $C[0, t]$  by  $\mathbb{P}^t$ ,  $\mathbb{P}^{\circ, t}$  and  $\mathbb{P}^{*, t}$  respectively. If  $t = T$  we will omit time dependence. We have

$$\frac{d\mathbb{P}^{*, t}}{d\mathbb{P}^{\circ, t}}(X^\circ) = \frac{p(t, X_t^\circ; T, v)}{p(0, u; T, v)} \frac{d\mathbb{P}^t}{d\mathbb{P}^{\circ, t}}(X^\circ)$$

(see for instance the proof of proposition 1 in Schauer, Van der Meulen and Van Zanten (2017)). Hence  $p$  shows up only in the first term on the right-hand-side. Upon taking the limit  $t \uparrow T$  of the expectation on the right-hand-side, the

term  $p(t, X_t^\circ; T, v)$  may vanish, depending on the precise form of  $b^\circ$ . For the proposals of sections 4.1 and 4.2 ahead, a formal proof of this can be found in Delyon and Hu (2006) and Schauer, Van der Meulen and Van Zanten (2017) respectively. In the following we will sketch the argument for the disappearance of  $p(t, X_t^\circ; T, v)$  under  $t \uparrow T$ .

#### 4.1. Proposals by Delyon and Hu

Delyon and Hu (2006) introduced proposals for which

$$b^\circ(t, x) = \lambda b(t, x) + \frac{v - x}{T - t}, \quad (4.1)$$

where  $\lambda \in \{0, 1\}$ . When evaluated for  $x = X_t^\circ$ , the pulling term  $(v - X_t^\circ)/(T - t)$  forces  $X^\circ$  to hit  $v$  at time  $T$ . Sufficient conditions for absolute continuity and expressions for the likelihood ratio of the laws of  $X^\star$  and  $X^\circ$  are derived in Delyon and Hu (2006). However, the proportionality constants in the derived likelihood ratio are missing. Whereas for generating diffusion bridges using a MH-sampler these constants are irrelevant, they do matter for step 3 of algorithm 1 (because the constants depend on  $\theta$ ). In case of a one-dimensional diffusion, the constant in the Radon-Nikodym derivative is derived in Papaspiliopoulos and Roberts (2012). The extension to the multivariate case brings no surprises. Here we consider the case  $\lambda = 0$ . It turns out that the derivative can be obtained by rewriting the expression obtained from applying Girsanov's theorem

$$\begin{aligned} \frac{d\mathbb{P}^{\star, t}}{d\mathbb{P}^{\circ, t}}(X^\circ) &= \exp(J_t(X^\circ)) \times \frac{p(t, X_t^\circ; T, v)}{\varphi(v; X_t^\circ, (T - t)a(t, X_t^\circ))} \\ &\times \frac{1}{p(0, u; T, v)} (2\pi T)^{-d/2} |\det a(t, X_t^\circ)|^{-1/2} \\ &\times \exp\left(-\frac{1}{2T}(v - u)' a(0, u)^{-1}(v - u)\right). \end{aligned} \quad (4.2)$$

Here  $\varphi(x; \mu, a)$  denotes the value of the normal density with mean  $\mu$  and variance  $a$ , evaluated at  $x$  and the functional  $J_t$  is defined by

$$\begin{aligned} J_t(X^\circ) &= \int_0^t b(s, X_s^\circ)' a^{-1}(s, X_s^\circ) dX_s^\circ - \frac{1}{2} \int_0^t b(s, X_s^\circ)' a^{-1}(s, X_s^\circ) b(s, X_s^\circ) ds \\ &- \frac{1}{2} \int_0^t (T - s)^{-1} (v - X_s^\circ)' \diamond da^{-1}(s, X_s^\circ) (v - X_s^\circ), \end{aligned}$$

where the  $\diamond$ -integral is obtained as the limit of sums where the integrand is computed at the right limit of each time interval as opposed to the left limit used in the definition of the Itô integral. It can be shown that all terms are well-behaved under the limit  $t \uparrow T$  and that

$$\Phi(X^\circ) = \exp(J_T(X^\circ)) \frac{\varphi(v; u, a(0, u))}{p(0, u; T, v)} \sqrt{\left| \frac{\det a(0, u)}{\det a(T, v)} \right|}.$$

The term  $p(t, X_t^\circ; T, v)$  appearing in (4.2) is essentially cancelled by  $\varphi(v; X_t^\circ, (T - t)a(t, X_t^\circ))$  in the limit. From the expression for  $\Phi$  we see that the factor  $p(0, u; T, v)$  solely appears as a multiplicative constant in the denominator of the Radon-Nikodym derivative between the target bridge and proposal bridge. Therefore, the proposals derived from (4.1) are feasible.

#### 4.2. Guided proposals

In this section we review a flexible class of proposal processes  $X^\circ$  that was developed and studied in Schauer, Van der Meulen and Van Zanten (2017). We will use this framework in the remainder and provide a recap of the relevant results in this section. For precise statements of these results we refer the reader to Schauer, Van der Meulen and Van Zanten (2017).

The basic idea is to replace the generally intractable transition density  $p$  that appears in the dynamics of the target bridge (see equations (3.1) and (3.2)) by the transition density of a diffusion process  $\tilde{X}$  for which it is known in closed form. Assume  $\tilde{X}$  satisfies the SDE  $d\tilde{X}_t = \tilde{b}(t, \tilde{X}_t) dt + \tilde{\sigma}(t, \tilde{X}_t) dW_t$ . Denote the transition density of  $\tilde{X}$  by  $\tilde{p}(s, x; T, v)$  and set  $\tilde{a} = \tilde{\sigma}\tilde{\sigma}'$ . Define the process  $X^\circ$  as the solution of the SDE (3.3) with

$$b^\circ(t, x) = b(t, x) + a(t, x)\nabla_x \log \tilde{p}(t, x; T, v). \tag{oo}$$

A process  $X^\circ$  constructed in this way is referred to as a *guided proposal* (a guiding term is superimposed on the drift to ensure the process hits  $v$  at time  $T$ ).

We reduce notation by writing  $p(s, x)$  for  $p(s, x; T, v)$ . Define

$$R(s, x) = \log p(s, x), \quad r(s, x) = \nabla R(s, x), \quad H(s, x) = -\Delta R(s, x), \tag{4.3}$$

where  $\nabla$  and  $\Delta$  denote the gradient and Laplacian with respect to  $x$  respectively. Similarly, write  $\tilde{p}(s, x)$  instead of  $\tilde{p}(s, x; T, v)$ , etc. In Schauer, Van der Meulen and Van Zanten (2017) sufficient conditions for absolute continuity of  $\mathbb{P}^*$  and  $\mathbb{P}^\circ$  are established together with a closed form expression for the Radon-Nikodym derivative. It turns out that

$$\frac{d\mathbb{P}^{*,t}}{d\mathbb{P}^{\circ,t}}(X^\circ) = \frac{\tilde{p}(0, u) p(t, X_t^\circ; T, v)}{p(0, u) \tilde{p}(t, X_t^\circ; T, v)} \exp\left(\int_0^t G(s, X_s^\circ) ds\right),$$

where  $G$  is given by

$$G(s, x) = (b(s, x) - \tilde{b}(s, x))' \tilde{r}(s, x) - \frac{1}{2} \text{tr}\left([a(s, x) - \tilde{a}(s, x)] \left[\tilde{H}(s, x) - \tilde{r}(s, x)\tilde{r}(s, x)'\right]\right) \tag{4.4}$$

(Cf. proposition 1 in Schauer, Van der Meulen and Van Zanten (2017)). Upon taking the expectation and the limit  $t \uparrow T$  it is proved in Schauer, Van der Meulen and Van Zanten (2017) that

$$\Phi(X^\circ) = \frac{\tilde{p}(0, u)}{p(0, u)} \exp\left(\int_0^T G(s, X_s^\circ) ds\right). \tag{4.5}$$



This time the term  $p(t, X_t^\circ; T, v)$  is essentially cancelled by  $\tilde{p}(t, X_t^\circ; T, v)$  and henceforth disappears in the limit. From the expression of  $\Phi$  we deduce that guided proposals are feasible.

The class of linear processes,

$$d\tilde{X}_t = \tilde{B}(t)\tilde{X}_t dt + \tilde{\beta}(t) dt + \tilde{\sigma}(t) dW_t, \quad (4.6)$$

is a flexible class with known transition densities and its induced guided proposals satisfy the conditions for absolute continuity derived in Schauer, Van der Meulen and Van Zanten (2017) under weak conditions on  $\tilde{B}$ ,  $\tilde{\beta}$  and  $\tilde{\sigma}$ . Proposal processes  $X^\circ$  derived by choosing a linear process as in (4.6) will be referred to as *linear guided proposals*. One key requirement for absolute continuity of  $X^*$  and  $X^\circ$  is that  $\tilde{\sigma}$  is such that  $\tilde{a}(T) = (\tilde{\sigma}\tilde{\sigma}')(T) = a(T, v)$ . A particularly simple type of guiding proposals is obtained upon choosing  $d\tilde{X}_t = \tilde{\beta}(t) dt + \sigma(T, v) dW_t$ . For this particular choice

$$b^\circ(t, x) = b(t, x) + \frac{a(t, x)a(T, v)^{-1}}{T - t} \left( v - x - \int_t^T \tilde{\beta}(s) ds \right). \quad (4.7)$$

Depending on the precise form of  $b$  and  $\sigma$  it can nevertheless be advantageous to use guided proposals induced for non-zero  $\tilde{B}$ . In section 4.4 we discuss several strategies for choosing the process  $\tilde{X}$ .

**Remark 4.1.** For guided proposals, it is easily seen that the process  $Z$  appearing in proposition 3.1 satisfies (3.5) with  $\mu(t, x) = \sigma'(t, x)(r(t, x) - \tilde{r}(t, x))$ .

**Remark 4.2.** In case  $b$  and  $\sigma$  are of the forms  $b(s, x) = \beta(s) + B(s)x$  and  $\sigma(s, x) = \sigma(s)$ , then we can trivially take  $\tilde{b} = b$  and  $\tilde{\sigma} = \sigma$ . By equation (4.4) it follows that in this case  $\Psi \equiv 1$ . This implies that  $A_2$  in algorithm 1 does not depend on  $\{Z_i, i = 1, \dots, n\}$  and simulating diffusion bridges is unnecessary. That is, step 2 of algorithm 1 can be omitted.

### 4.3. Drift-independent guided proposals

The proposals with  $\lambda = 1$  (appearing in equation (4.1)) provided by Delyon and Hu (2006) are a special case of guided proposals only in case  $\sigma$  is constant. These are recovered upon choosing  $\tilde{b} \equiv 0$  and  $\tilde{\sigma} = \sigma$ . Proposals with  $\lambda = 0$  are a special case when both  $b$  and  $\sigma$  are constant and correspond to choosing  $\tilde{b} = b$  and  $\tilde{\sigma} = \sigma$ . The latter type of proposals enjoys quite some popularity in the literature, especially when discretised with the multiplicative correction term added to the diffusion term introduced by Durham and Gallant (2002) (the resulting discrete time proposal is called the *modified diffusion bridge*, we get back to this in section 5). As such proposals are independent of the drift these can only work satisfactory if the drift is locally constant.

In this article we do not aim to make a formal comparison of guided proposals and Delyon-Hu proposals. Nevertheless, we wish to remark that for the latter class of proposals both in case  $\lambda = 0$  and when  $\lambda = 1$  the resulting bridges may

not resemble true bridges. An illuminating example is given in the introductory section of Schauer, Van der Meulen and Van Zanten (2017) and we refer to that paper for further discussion on this rather subtle issue. In case the reader is uncomfortable with the additional freedom for choosing the process  $\tilde{X}$ , proposals similar (but not equal to) Delyon-Hu proposals can be obtained by taking  $d\tilde{X}_t = \sigma(T, v) dW_t$ , where  $\sigma(T, v)\sigma(T, v)' = a(T, v)$ . In that case we get proposals with

$$b^\circ(t, x) = b(t, x) + a(t, x)a(T, v)^{-1} \frac{v - x}{T - t}.$$

Proposals that ignore the drift completely can be defined by

$$b^\circ(t, x) = a(t, x)a(T, v)^{-1} \frac{v - x}{T - t},$$

We call these *drift-independent guided proposals*. The acceptance probability for drift-independent proposals can easily be obtained from (4.5) and equals

$$\begin{aligned} \Phi(X^\circ) = \frac{\tilde{p}(0, u)}{p(0, u)} \exp & \left( \int_0^T G(s, X_s^\circ) ds + \int_0^T b(s, X_s^\circ)' a^{-1}(s, X_s^\circ) dX_s^\circ \right. \\ & \left. - \frac{1}{2} \int_0^T b(s, X_s^\circ)' a^{-1}(s, X_s^\circ) \left[ b(s, X_s^\circ) + 2a(s, X_s^\circ)a(T, v)^{-1} \frac{v - X_s^\circ}{T - s} \right] ds \right), \end{aligned}$$

where  $G$  is computed with  $\tilde{b} \equiv 0$  and  $\tilde{a} = a(T, v)$ .

#### 4.4. Choice of guided proposals

In this section we discuss the choice of guided proposals. We propose the following strategies:

1. *Linearisation of the drift.* In some examples there is a natural point at which to linearise, as in example 4.3. If this is not the case, one can use a (weighted) regression, as explained in example 4.4.
2. *Solving the dynamical system associated to the SDE.* Suppose  $x(t)$  satisfies the deterministic differential equation

$$dx(t) = b(t, x(t)) dt, \quad x(0) = x_0. \tag{4.8}$$

Then

$$d\tilde{X}_t = b(x(t)) dt + \tilde{\sigma} dW_t. \tag{4.9}$$

is clearly of the form (4.6) with  $\tilde{\beta}(t) = b(x(t))$ ,  $\tilde{B} \equiv 0$  and  $\tilde{\sigma} = \sigma(T, v)$ . This approach is illustrated in example 4.5.

3. *Combined approach.* Approximate  $b(t, X_t)$  with  $b(t, x(t)) + V(t, x(t))(X_t - x(t))$ , where  $V(t, y)$  is the matrix with elements  $V(t, y)_{i,j} = \partial b_i(t, y) / \partial y_j$  for  $y \in \mathbb{R}^d$ . This gives linear guided proposals with

$$\tilde{\beta}(t) = b(t, x(t)) - V(t, x(t))x(t) \quad \text{and} \quad \tilde{B}(t) = V(t, x(t)).$$

This is closely related to the linear noise approximation of the SDE for  $X$  as used in Whitaker et al. (2017).

4. *Iterative linearisation procedures.* A further technique using ideas by Whittaker et al. (2017) is obtained by setting  $\tilde{\beta}(t) = b(t, \mathbb{E}[\tilde{X}^*(t)])$  where  $\tilde{X}^*$  is a tractable diffusion bridge from  $u$  to  $v$  (derived for example from a preliminary linear approximation to  $X$ ).

We will always have  $\tilde{\beta}(0) = b(0, u)$  and  $\tilde{\beta}(T) = b(T, v)$ . Linear interpolation gives

$$\tilde{\beta}(t) = (1 - t/T)b(0, u) + (t/T)b(T, v). \quad (4.10)$$

**Example 4.3.** Let  $X$  be the diffusion process described by the SDE

$$dX_t = (\alpha \arctan(X_t) + \beta) dt + \sigma dW_t. \quad (4.11)$$

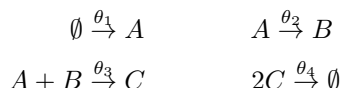
If  $\alpha < 0$ ,  $\frac{\pi}{2}\alpha < \beta < -\frac{\pi}{2}\alpha$  this process is mean reverting to  $\tan(-\beta/\alpha)$ . For  $x \approx \tan(-\beta/\alpha)$

$$b(x) \approx \alpha \cos^2(-\beta/\alpha)(x - \tan(-\beta/\alpha)).$$

So it makes sense to take linear proposals with

$$\tilde{B} = \alpha \cos^2(-\beta/\alpha), \quad \tilde{\beta} = \frac{1}{2}\alpha \sin(2\beta/\alpha) \quad \text{and} \quad \tilde{\sigma} = \sigma.$$

**Example 4.4.** Here we consider a simple example in which the dynamics of a chemical reaction network are approximated by a system of stochastic differential equations. Suppose we have four reactions among chemicals  $A$ ,  $B$  and  $C$ :



The amount of the chemicals  $A$ ,  $B$ ,  $C$  at time  $t$  can be modelled as a pure jump Markov process which can subsequently be approximated by the diffusion process  $X_t \in \mathbb{R}^3$  which solves the *Chemical Langevin Equation* (Fuchs (2013), chapter 4)

$$dX_t = S h_\theta(X_t) dt + S \text{diag}(\sqrt{h_\theta(X_t)}) dW_t \quad (4.12)$$

driven by a  $\mathbb{R}^4$ -valued Brownian motion. Here

$$S = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

is the stoichiometry matrix of the system. Its elements describe the net effect of each reaction on each chemical species. Furthermore,  $h_\theta(x) = \theta \circ h(x)$  is a function describing the hazard for a particular reaction to happen. Here  $\circ$  denotes the Hadamard (or entrywise) product of two vectors and

$$\theta = [\theta_1, \theta_2, \theta_3, \theta_4]' \quad h(x) = [1, x_1, x_1 x_2, x_3(x_3 - 1)/2]'$$

We choose  $\tilde{B}$  and  $\tilde{\beta}$  to depend on  $\theta$  (but not on time) so that  $\tilde{B}x + \tilde{\beta}$  approximates  $b_\theta(x)$ . While it is possible to take different approximations specifically

tailored for each bridge segment, it is computationally advantageous to work with a global approximation to  $b_\theta$  (as we need to evaluate  $\tilde{p}$  in the expression for  $A_2$ , see also the discussion in section 6). To this end, we replace  $h$  by a linear approximation  $\tilde{h}$  which allows for obtaining  $\tilde{B}_\theta$  and  $\tilde{\beta}_\theta$  from the equation

$$\tilde{B}_\theta x + \tilde{\beta}_\theta = S(\theta \circ \tilde{h}(x)).$$

As the first two components of  $h(x)$  are linear, we take  $\tilde{h}_1(x) = h_1(x)$  and  $\tilde{h}_2(x) = h_2(x)$ . We approximate  $h_3$  by  $\tilde{h}_3(x) = c_3 + u_{3,1}x_1 + u_{3,2}x_2$ . Values for  $c_3, u_{3,1}$  and  $u_{3,2}$  are obtained from a weighted linear regression of  $x_1x_2$  on  $x_1$  and  $x_2$ , with weights proportional to  $x_1x_2$ . Similarly, we take  $\tilde{h}_4(x) = c_4 + u_{4,3}x_3$ . Values for  $c_4$  and  $u_{4,3}$  are obtained from a weighted linear regression of  $\frac{1}{2}x_3(x_3 - 1)$  on  $x_3$ , with weights proportional to  $x_3(x_3 - 1)$ . We take a *weighted* regression in this way because for a good proposal the error matters more if the corresponding dispersion component is small. For  $\tilde{\sigma}$  we choose  $\tilde{\sigma} = S \text{diag} \left( \sqrt{\tilde{h}_\theta(x_i)} \right)$  on the segment between times  $t_{i-1}$  and  $t_i$ .

Note that this approach for constructing  $\tilde{B}$  and  $\tilde{\beta}$  can be applied generally to stochastic differential equations arising from chemical reaction networks.

**Example 4.5.** The Lotka-Volterra model with multiplicative noise (cf. Khasminskii and Klebaner (2001)) is given by the Stratonovich stochastic differential equation

$$\begin{aligned} dX_t &= (\theta X_t - X_t Y_t) dt + \sigma X_t \circ dW_t^{(1)}, & X_0 &= x_0 \\ dY_t &= (-\theta Y_t + X_t Y_t) dt + \sigma Y_t \circ dW_t^{(2)}, & Y_0 &= y_0. \end{aligned} \tag{4.13}$$

By Itô's formula,  $(\xi_t, \eta_t) = (\log X_t, \log Y_t)$  satisfies

$$\begin{aligned} d\xi_t &= (\theta - e^{\eta_t}) dt + \sigma dW_t^{(1)}, & \xi_0 &= \log x_0 \\ d\eta_t &= (-\theta + e^{\xi_t}) dt + \sigma dW_t^{(2)} & \eta_0 &= \log y_0. \end{aligned}$$

Proposals for a bridge that hits  $(\xi_T, \eta_T) = (\log X_T, \log Y_T)$  at time  $T$  can be derived from the deterministic dynamical system associated with (4.13). The deterministic system  $(x, y)$  has trajectories  $xye^{-\frac{1}{\theta}(x+y)} = K$  with  $K$  depending on  $x_0, y_0$ . The trajectory can be parametrised by

$$x(z) = \frac{z}{2} \pm \sqrt{z^2 - 4Ke^{z/\theta}}, \quad y(z) = z - x(z),$$

where time is implicit and can be recovered from  $z$  by the equation

$$\theta \sqrt{z^2 - 4Ke^{z/\theta}} dt = \pm dz$$

(Cf. Steiner and Gander (1999)). We obtain guided proposals for  $(\xi_t^\circ, \eta_t^\circ)$  by taking  $\tilde{B} \equiv 0$  and  $\tilde{\beta}(t) = (\theta - x(t), -\theta + y(t))'$ . These proposals can subsequently be transformed to proposals for  $(X_t^\circ, Y_t^\circ)$ .

## 5. Numerical discretisation of guided proposals

Simulation of  $X^\circ$  and numerical evaluation of  $\Psi(X^\circ) := \exp\left(\int_0^T G(s, X_s^\circ) ds\right)$  is numerically cumbersome since the drift of  $X^\circ$  and the integrand  $G$  explode for  $s$  near the endpoint  $T$ .

**Example 5.1.** Suppose  $\sigma$  is constant and we take  $\tilde{X} = \sigma dW_t$ . Then we have  $\tilde{r}(s, x) = \tilde{a}^{-1}(v - x)/(T - s)$ , where  $\tilde{a} = \sigma\sigma'$ . Hence the drift of the SDE for  $X^\circ$  explodes when  $s \uparrow T$ . Furthermore,

$$\log \Psi(X^\circ) = \int_0^T b(s, X_s^\circ)' \tilde{r}(s, X_s^\circ) ds = \int_0^T b(s, X_s^\circ)' \tilde{a}^{-1} \frac{v - X_s^\circ}{T - s} ds,$$

which shows the integrand explodes as well.

In this section we explain how these numerical problems can be dealt with using a time change and scaling of the proposal process. The purpose is not solely obtaining a more accurate discretisation scheme for the SDE, but above all accurate evaluation of the integral appearing in  $\Psi(X^\circ)$ .

For the particular example just given Clark (1990) proposed to perform a time change and scaling of the proposal process to remove the singularities. Define  $\tau^C : [0, \infty) \rightarrow [0, T)$  by  $\tau^C(s) = T(1 - e^{-s})$  and  $U_s^C = e^{s/2}(v - X_{\tau^C(s)}^\circ)$ . Then  $U^C$  satisfies the stochastic differential equation

$$dU_s^C = -Te^{-s/2}b(T(1 - e^{-s}), v - e^{-s/2}U_s^C) ds - \frac{1}{2}U_s^C ds - \sqrt{T}\sigma dW_s,$$

which behaves like a zero-mean mean-reverting Ornstein-Uhlenbeck process as  $s \rightarrow \infty$ . Furthermore,

$$\log \Psi(X^\circ) = \int_0^\infty e^{-s/2}b(\tau^C(s), v - e^{-s/2}U_s^C)' T\tilde{a}^{-1}U_s^C ds$$

(note that there are some minor typographical errors in Clark (1990)). Clearly, if  $b$  is bounded, this removes the singularity near  $T$ , but at the cost of having to deal with an infinite integration interval. For this reason, we propose a different time-change and scaling.

The time change and scaling due to Clark (1990) is a special case obtained from considering the process  $U_s = m(s)\left(v(\tau(s)) - X_{\tau(s)}^\circ\right)$ , where  $s \mapsto \tau(s)$  is nondecreasing. The choice by Clark (1990) corresponds to  $\tau(s) = T(1 - e^{-s})$  and  $m(s) = e^{s/2}$ . In the following we denote the time derivatives of  $m$  and  $\tau$  by  $\dot{m}$  and  $\dot{\tau}$  respectively. The time changed process  $U = (U_s, s \in [0, T))$  satisfies the stochastic differential equation

$$dU_s = \left( \frac{\dot{m}(s)}{m(s)}U_s - m(s)\dot{\tau}(s)b^\circ(\tau(s), v - U_s/m(s)) \right) ds - m(s)\sqrt{\dot{\tau}(s)}\sigma(\tau(s), v - U_s/m(s)) dW_s$$

Using the setting of example 5.1, we motivate another choice of  $\tau$  and  $m$  for improving numerical accuracy. For the example, the drift of  $U$  is given by

$$-m(s)\dot{\tau}(s)b(\tau(s), v - U_s/m(s)) + \left( \frac{\dot{m}(s)}{m(s)} - \frac{\dot{\tau}(s)}{T - \tau(s)} \right) U_s$$

and  $\log \Psi(X^\circ)$  can be expressed in terms of  $U$  as follows

$$\int_{\tau(0)}^{\tau(T)} b(\tau(s), v - U_s/m(s)) \tilde{a}^{-1} \frac{U_s}{m(s)} \frac{\dot{\tau}(s)}{T - \tau(s)} ds. \tag{5.1}$$

As shown in Schauer, Van der Meulen and Van Zanten (2017), up to a logarithmic term,  $v - X_s^\circ \sim \sqrt{T - s}$  for  $s$  close to  $T$ . Therefore,  $U_s \sim m(s)\sqrt{T - \tau(s)}$  which implies that the possibly exploding part of the integral in (5.1) satisfies

$$\frac{U_s}{m(s)} \frac{\dot{\tau}(s)}{T - \tau(s)} \sim \frac{\dot{\tau}(s)}{\sqrt{T - \tau(s)}}.$$

To make this constant, we take  $\tau(s) = s(2 - s/T)$ . Furthermore, we choose  $m(s) = 1/(T - s)$  (see section 5.2 for a justification). With these choice of  $\tau$  and  $m$ ,  $U$  satisfies the SDE

$$dU_s = -\frac{2}{T}b(\tau(s), v - (T - s)U_s) ds - \frac{1}{T - s}U_s ds - \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T - s}}\sigma dW_s,$$

with  $U_0 = (v - u)/T$ . Compared to the original SDE for  $X^\circ$ , we see that an additional exploding factor appears in the diffusion coefficient. At first sight, this may seem like we have worsened the numerical problems. Note however that the integral we wish to evaluate ( $\log \Psi(X^\circ)$ ) behaves much better now. For  $s \approx T$ , the process  $U$  behaves like a mean-zero stationary Ornstein-Uhlenbeck process, with balanced increased mean-reversion and diffusivity. The process  $U^C$  proposed by Clark (1990) behaves like an Ornstein-Uhlenbeck process for large times as well, and we see that with our choice of  $\tau$  we speed up time to run through this process much faster, preventing us from evaluating an integral over an unbounded integration region.

**5.1. Time changing and scaling of linear guided proposals**

Based on the motivational derivations of the preceding section, we define a convenient time change and scaling in this section. To do this, we need a few more results from Schauer, Van der Meulen and Van Zanten (2017). If  $\tilde{X}$  is a linear process (satisfying equation (4.6)), then

$$\tilde{r}(s, x) = \tilde{H}(s)(v(s) - x), \tag{5.2}$$

where

$$v(s) = F(s, T)v - \int_s^T F(s, z)\tilde{\beta}(z) dz \tag{5.3}$$

( $\tilde{r}$  and  $\tilde{H}$  are defined in equation (4.3)). Here  $F(t, s) = F(t)F(s)^{-1}$  with  $F(t)$  the fundamental  $d \times d$  matrix that satisfies

$$F(t) = I + \int_0^t \tilde{B}(z)F(z) dz. \quad (5.4)$$

Define the process  $U$  by

$$U_s := \frac{v(\tau(s)) - X_{\tau(s)}^\circ}{T - s}. \quad (5.5)$$

This implies

$$X_{\tau(s)}^\circ = v(\tau(s)) - (T - s)U_s =: \Gamma(s, U_s). \quad (5.6)$$

**Lemma 5.2.** *The time changed and scaled process  $U = (U_s, s \in [0, T])$  satisfies the stochastic differential equation*

$$\begin{aligned} dU_s &= \frac{2}{T} \dot{v}(\tau(s)) ds - \frac{2}{T} b(\tau(s), \Gamma(s, U_s)) ds \\ &\quad + \frac{1}{T - s} \left( I - 2a(\tau(s), \Gamma(s, U_s))J(s) \right) U_s ds \\ &\quad - \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T - s}} \sigma(\tau(s), \Gamma(s, U_s)) dW_s, \quad U_0 = \frac{v - u}{T} \end{aligned} \quad (5.7)$$

where  $W$  is a Brownian motion and  $J$  defined by

$$J(s) = \tilde{H}(\tau(s))(T - \tau(s)) \quad (5.8)$$

satisfies  $\lim_{s \uparrow T} \tilde{a}(s)J(s) = I$ . Moreover,

$$\begin{aligned} \int_0^T G(s, X_s^\circ) ds &= 2 \int_0^T (b - \tilde{b})'(\tau(s), \Gamma(s, U_s))J(s)U_s ds \\ &\quad - \int_0^T \text{tr} \left[ \frac{(a - \tilde{a})(\tau(s), \Gamma(s, U_s))}{T - s} J(s) (I - T U_s U_s' J(s)) \right] ds. \end{aligned}$$

If we simulate  $U$  on an equidistant grid we can recover  $X^\circ$  on a non-equidistant grid from equation (5.6). This implies  $X^\circ$  is evaluated on an increasingly finer grid as  $s$  increases to  $T$ . In our implementation, all computations are done in time-changed/scaled domain, and the mapping  $g$  is in fact defined by setting  $U = g(\theta, Z^\circ)$ , where  $Z^\circ$  is the driving Brownian Motion for  $U$ .

## 5.2. Motivation for the scaling

Consider the SDE for  $\tilde{X}$  as defined in equation (4.6). The corresponding fundamental matrix is given in equation (5.4). Define the process  $\tilde{X}^*$  as the process  $\tilde{X}$ , conditioned on  $\tilde{X}_T = v$ . Then  $\tilde{X}^*$  is a linear process itself with drift  $\tilde{b}^*(t, x) = \tilde{B}(t)x + \tilde{\beta}(t) + \tilde{a}(t)\tilde{H}(t)(v(t) - x)$  and diffusion coefficient  $\tilde{\sigma}^*(t) = \sigma(t)$ . Denote the corresponding fundamental matrix by  $F^*$ . Hence  $F^*$  satisfies

$$\frac{d}{dt} F^*(t) = \left( \tilde{B}(t) - \tilde{a}(t)\tilde{H}(t) \right) F^*(t), \quad F^*(0) = I.$$

**Theorem 5.3.** Fix a nondecreasing differentiable mapping  $\tau : [0, T] \rightarrow [0, \infty)$ . If we define the scaling matrix  $m$  by  $m(s) = (T - s)F^*(\tau(s))^{-1}$ . then the process  $U$  defined by

$$U_s = m(s) \left[ v(\tau(s)) - X_{\tau(s)}^\circ \right]$$

(with  $s \mapsto v(s)$  as defined in equation (5.3)) satisfies the SDE

$$dU_s = \left( -\frac{U_s}{T-s} - m\dot{\tau} \left[ b(\tau, \Gamma) - \tilde{b}(\tau, \Gamma) + (a(\tau, \Gamma) - \tilde{a}(\tau))\tilde{H}(\tau)m^{-1}U_s \right] \right) ds - m\sqrt{\tilde{\tau}}\sigma(\tau, \Gamma) dW_s,$$

where  $\Gamma \equiv \Gamma(s, U_s) = v(\tau(s)) - m(s)^{-1}U_s$ . To lighten the notation we have written  $\tau, \dot{\tau}$  and  $m$  to denote  $\tau(s), \dot{\tau}(s)$  and  $m(s)$  respectively.

The proof is deferred to the appendix (section B).

**Corollary 5.4.** Let  $\bar{U}_{t_i}$  denote the Euler approximation at time  $t_i$  of  $U$ . If  $a(t, x) \equiv a(T, v) = \tilde{a}$  and  $b(t, x) = \tilde{b}(t, x) = \tilde{B}(t)x + \tilde{\beta}(t)$ , then

$$\mathbb{E} [U_{t_i} | U_{t_{i-1}} = u] = \mathbb{E} [\bar{U}_{t_i} | \bar{U}_{t_{i-1}} = u].$$

*Proof.* In this case

$$dU_s = -\frac{U_s}{T-s} ds - m\sqrt{\tilde{\tau}}\sigma(\tau, \Gamma) dW_s.$$

Hence

$$\mathbb{E} [U_{t_i} | U_{t_{i-1}} = u] = \frac{T - t_i}{T - t_{i-1}} u.$$

It is easy to see that this coincides with  $\mathbb{E} [\bar{U}_{t_i} | \bar{U}_{t_{i-1}} = u]$ . □

This shows that if we use linear guided proposals and use the scaling matrix  $m$  defined in theorem 5.3, then the Euler approximation of the process  $U$  has the correct conditional expectation when  $X$  itself is a linear process. Note that this is not necessarily the case without applying the scaling.

In case  $\tilde{\beta} = 0, \tilde{B} = 0$  and  $\tilde{\sigma}(t) = \tilde{\sigma}$ , we have  $F^*(t) = I/(T - t)$  and  $m(s) = 1/(T(T - s))I$ . This means that we should have  $m(s) = O(1/(T - s))$  for  $s \approx T$ .

### 5.3. Numerical illustrations

In this section we use simulation to assess the effect of discretisation when using guided proposals with dynamics

$$dX_t^\circ = b(t, X_t^\circ) dt + a(t, X_t^\circ)\tilde{r}(t, X_t^\circ) dt + \sigma(t, X_t^\circ) dW_t \tag{5.9}$$

(Cf. section 4.2). We consider combinations of drift functions  $b(x) \equiv 0$  and  $b(x) = \arctan(x)$  and diffusion functions  $\sigma = 1$  and  $\sigma(x) = 1 + 0.3 \sin(3x)$ . We consider two types of guided proposals  $X^\circ$ :



1. proposals generated by choosing  $d\tilde{X}_t = \sigma(T, v) dW_t$  which gives pulling term

$$\tilde{r}(t, x) = \sigma(T, v)^{-2}(v - x)/(T - t) \quad (\text{BM-pull})$$

2. proposals generated by choosing  $d\tilde{X}_t = -\beta\tilde{X}_t dt + \sigma(T, v) dW_t$ , which gives pulling term

$$\tilde{r}(t, x) = \frac{2\beta}{\sigma^2(T, v)} e^{-\beta(T-t)} \frac{v - xe^{-\beta(T-t)}}{1 - e^{-2\beta(T-t)}}. \quad (\text{OU-pull})$$

We aim to compare various discretisation schemes of the SDE (5.9). In all cases, we use the equidistant grid by imputing  $m - 1$  points on  $[0, T]$  for discretisation. Define  $h = T/m$  and set  $t_j = jh$ ,  $j = 0, \dots, m$ . The discretisation schemes considered are

1. **Euler**: Euler discretisation of the SDE for  $X^\circ$ .
2. **Mdb**: The Modified diffusion bridge (Mdb) discretisation introduced in Durham and Gallant (2002). This discretisation is obtained by applying Euler discretisation to the SDE for  $X^\circ$  and adding a correction term to the diffusion coefficient. This gives the scheme  $\{\check{X}_{t_j}^\circ\}$  where

$$\check{X}_{t_{j+1}}^\circ = \check{X}_{t_j}^\circ + b^\circ(t_j, \check{X}_{t_j}^\circ)(t_{j+1} - t_j) + \sigma(t_j, \check{X}_{t_j}^\circ) \sqrt{\frac{T - t_{j+1}}{T - t_j}} (W_{t_{j+1}} - W_{t_j}) \quad (5.10)$$

3. **Euler-U**: Euler discretisation of the SDE for  $U$ .
4. **Euler-V**: Euler discretisation of the SDE for the time-changed process using  $\tau$ , but *without* the scaling. This means that we apply Euler discretisation to the SDE

$$dV_s = b^\circ(\tau(s), V_s) \dot{\tau}(s) ds + \sqrt{\dot{\tau}(s)} \sigma(\tau(s), V_s) dW_s, \quad V_0 = u$$

where  $V_s = X_{\tau(s)}^\circ$ .

The first two of these schemes have gained quite some popularity in the literature; the third one is what we propose and the fourth one is included to assess the effect of including a scaling (on top of the time-change).

In the simulation study we are interested in accurate discretisation of the likelihood given in equation (4.5) which appears in the acceptance probabilities of the algorithms of Section 3.2 (whether the considered pulling terms are good choices is of minor importance for that purpose). More precisely, we evaluate the discretisation of the path-integrals  $I(X^\circ) = \int_0^T G(s, X_s^\circ) ds$  in case of discretisation of the SDE for  $X^\circ$ ,  $I(V) = \int_0^T G(\tau(s), V_s) \dot{\tau}(s) ds$  in case of discretisation of the SDE for  $V$  and  $I(U) = \int_0^T G(\tau(s), v(\tau(s) - (T - s)U_s)) \dot{\tau}(s) ds$  in case of discretisation of the SDE for  $U$ .

At the finest discretisation level, we divide  $[0, T]$  into  $2^L$  intervals of equal length. If  $h = T/2^L$ , then  $t_j = jh$ ,  $j = 0, \dots, 2^L$ . We start by simulating on the finest grid. Next we redo the simulation on the grid of length  $2^{L-1}$  using the

same Wiener process increments. This can be continued iteratively (until there are only 2 intervals of equal length). The simulation study was run as follows:

1. Generate the sequence  $\{t_j\}$  with  $h = T/2^L$ .
2. Generate Wiener increments on the generated grid.
3. Simulate a realisation  $X^{\circ,L}$  of the diffusion bridge with the generated Wiener increments. Compute and store  $I(X^{\circ,L})$ .
4. for  $k=L-1$  downto 2
  - Coarsen the grid by removing the 2nd, 4th, 6th, etc point from the grid and aggregate the Wiener-increments. Simulate a realisation  $X^{\circ,k}$  of the diffusion bridge with these Wiener increments.
  - Compute the error  $e_k = I(X^{\circ,k}) - I(X^{\circ,L})$ .
5. Repeat for  $B$  times steps 1 up till 4 and compute the Root Mean Squared Error of all errors using an equal number of grid-points.

We chose  $L$  sufficiently large such that the approximation for  $I$  is virtually the same for all discretisation methods. As quadrature rule we used the midpoint rule, where the integrand is evaluated at the left-point.

In the simulations, we simulated bridges starting in  $u = 0$  at time 0 and ending in  $v = 3$  at time  $T = 1$ . The results of the simulations are in figure 1. By definition, there is no error in the lower-left panel. From the simulation results we see that for various combined choices of drift, diffusion coefficient and pulling term, our approach performs best. We have run simulations with other values for  $v$ ,  $T$ ,  $b$  and  $\sigma$  leading qualitatively to the same conclusion.

The beneficial effect of the time change and scaling is further illustrated in the examples of Sections 7.1 and 7.2.

#### 5.4. Order of convergence

Ideally, one would derive a result on the order of convergence of each of the discussed discretisation methods for approximating  $I$ . We feel that this is outside the scope of this paper. As noted in Papaspiliopoulos, Roberts and Stramer (2013) (page 676): “Quantitative results on the relative efficiency of discretisation schemes are scarce in the literature.” In case  $b = 0$  and  $\sigma$  is constant (which is the simplest case to consider), Papaspiliopoulos, Roberts and Stramer (2013) show that the strong order of convergence of the Euler scheme is  $O(\sqrt{\delta})$  at  $T - \delta$ . This shows that the usual higher  $O(\delta)$  strong order (which holds for diffusions with additive noise) is lost due to the exploding behaviour of the drift. Along similar lines as in Papaspiliopoulos, Roberts and Stramer (2013) one can prove that the strong order  $\delta$  is maintained if the time-change and scaling is used. Admittedly, this is a rather weak result since (i) the case  $b = 0$  and  $\sigma$  constant is very specific, (ii) the focus is on accurate evaluation of a path integral of the proposal bridge and not solely the process at specified points. The concept of strong order is not really needed here: we are interested in almost sure convergence of Euler approximation pathwise. Under local Lipschitz conditions on the drift and diffusion coefficients, the pathwise convergence rate of the Euler scheme coincides up to an arbitrarily small  $\varepsilon > 0$  with its strong convergence

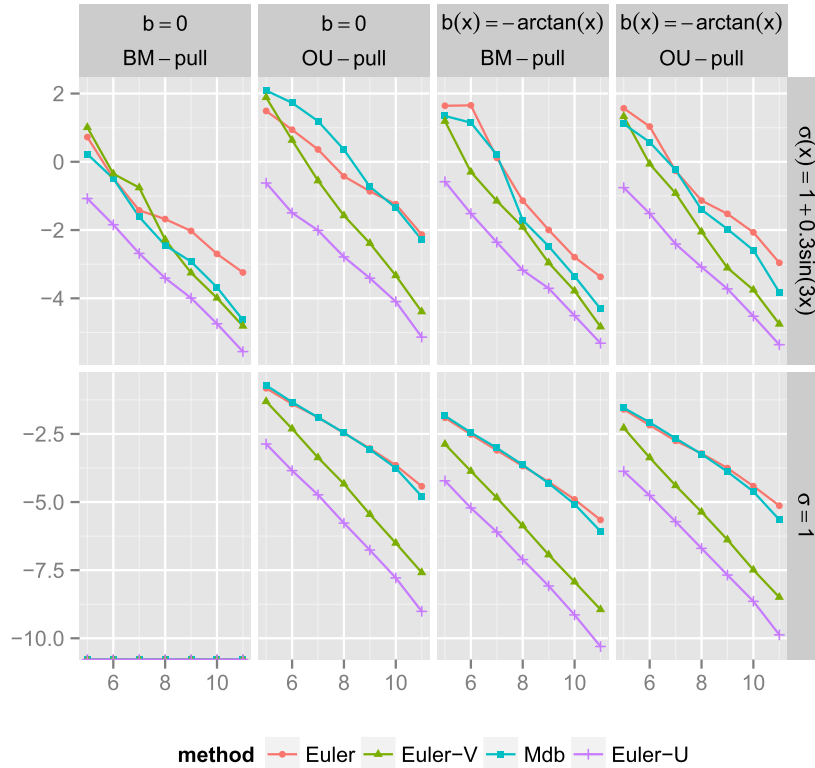


FIG 1.  $\log_2(RMSE)$  of  $I(\cdot)$  versus  $\log_2(nr \text{ of segments})$ . **Euler**: Euler discretisation of the SDE for  $X^\circ$ . **Mdb**: The Modified diffusion bridge discretisation introduced in Durham and Gallant (2002). **Euler-U**: Euler discretisation of the SDE for  $U$ . **Euler-V**: Euler discretisation of the SDE for the time-changed process using  $\tau$ , but without the scaling.

rate  $1/2$  (Cf. Gyöngy (1998)). We expect the same pathwise convergence rate to hold for the integrals, when approximated using the proposed time-change and scaling. In this sense, it is not unexpected that the lower panel in figure 1 shows lines with slopes close to either  $1/2$  (**Euler**, **Mdb**) or  $1$  (**Euler-V**, **Euler-U**).

### 6. Computational costs and implementation

In this section we discuss the computational cost of using guided proposals. For comparison, we add the computational cost of Delyon-Hu type proposals. Here we only consider the cost of imputation by diffusion bridges (including the computation of their acceptance probabilities). Let

- $K$  denotes the number of iterations of the data-augmentation algorithm;
- $M$  denotes the number of segments for imputations (so  $M+1$  is the number of discrete-time observations);
- $N$  denotes the number of Euler-step applied to each segment.

TABLE 1  
 Overview of computational cost for simulating proposals.

	$\int_t^T \beta_\theta ds$	$\exp(sB)$	$B\lambda + \lambda B' = \tilde{a}$
Delyon-Hu	0	0	0
$\tilde{b}(t, x) = \beta_\theta(t)$	$NMK$	0	0
$\tilde{b}(t, x) = Bx + \beta_\theta(t)$	$NMK$	$N$	$M$
$\tilde{b}(t, x) = B_\theta x + \beta_\theta(t)$	$NMK$	$NK$	$MK$

The computational costs of simulating proposals are summarised in table 1. We give some elucidation on this table.

1. Applying guided proposals with  $\tilde{B} \equiv 0$  gives minor additional computations compared to Delyon-Hu type proposals. One merely needs to compute  $\int_t^T \beta_\theta(s) ds$  on the whole augmented grid during all simulations. If  $\beta$  does not depend on  $\theta$  this computation needs to be carried out only once on the whole grid.
2. If  $\tilde{B} \neq 0$ , then simulation of  $U$  as defined in equation (5.7) requires evaluation of both  $\dot{v}$  and  $J$ , where  $v$  and  $J$  are defined in equations (5.3) and (5.8) respectively. As  $\dot{v}(s) = \tilde{B}(s)v(s) + \tilde{\beta}(s)$ , evaluating  $\dot{v}$  requires evaluation of  $v$ . This in turn requires evaluation of matrix exponentials. For evaluating  $J$ , we first compute  $\lambda$  as the solution to continuous time Lyapunov equation  $B\lambda + \lambda B' = -\tilde{a}$ . Using  $\lambda$  we can evaluate  $J$  using

$$J(s) = z(s) \left( e^{-\tilde{B}_\theta z(s)} \lambda e^{-\tilde{B}'_\theta z(s)} - \lambda \right)^{-1},$$

where  $z(s) = T - \tau(s)$ . These functions need to be computed on the whole augmented grid in each iteration. In case  $\tilde{B}$  does not depend on  $\theta$ , both  $J$  and  $\dot{v}$  can be precomputed on a grid in advance to the MCMC-algorithm, preventing multiple expensive matrix exponential computations.

Besides simulation of the proposals, an acceptance probability needs to be computed. This requires evaluation of certain integrals of the proposal. A potential disadvantage of Delyon-Hu type proposals is that inverses appear. Moreover, stochastic integrals need to be approximated.

## 7. Examples

The source code of the examples is available online.<sup>1</sup> It is written in the programming language Julia (Bezanson et al. (2012)).

<sup>1</sup>See <https://github.com/mschauer/BayesEstDiffusion.jl>.

### 7.1. Example for one-dimensional diffusion

In this section we discuss example 4.3, where we considered the SDE

$$dX_t = (\alpha \arctan(X_t) + \beta) dt + \sigma dW_t.$$

Cf. equation (4.11). The goal is twofold: (i) to show that the proposed algorithm does not deteriorate when increasing the number of imputed points, (ii): to show that the discretisation scheme of section 5 reduces discretisation error.

We take the diffusion process with dynamics of (4.11). Assume that we observe  $X$  at times points  $t = 0, 0.3, 0.6 \dots, T = 30$  and wish to estimate  $(\alpha, \beta, \sigma)$ . As true values we took  $\alpha = -2, \beta = 0$  and  $\sigma = 0.75$ . For generating the discrete time data we simulated the process on  $[0, T]$  at 400 001 equidistant time points using the Euler scheme and take a subsample.

For  $\alpha$  and  $\beta$  we chose apriori independently a  $\mathcal{N}(0, \xi^2)$ -distribution with variance  $\xi^2 = 5$ . For  $\log \sigma$  we used an uninformative flat prior. We applied algorithm 2 with  $\rho = 0$  in (3.8) with random walk proposals for  $q(\sigma^\circ | \sigma)$  of the form  $\log \sigma^\circ := \log \sigma + u$  with  $u \sim \mathcal{U}(-0.1, 0.1)$ .

We initialised the sampler with  $\alpha = -0.1, \beta = -0.1$  and  $\sigma = 2$  and varied the number of imputed points over  $m = 10, 100$  and  $1000$ . Acceptance rates for proposed bridges were in all cases between 94% and 95% and for  $\sigma$  between 72% and 73%.

Figures 2 and 3 illustrate the results of running the MCMC chain for 10.000 iterations using  $m = 10, 100, 1000$  imputed points respectively for each bridge (including endpoints), both with time change and without. Two things stand out: firstly, increasing the number of imputed points  $m$  does not worsen the mixing of the chain and secondly the vastly reduced bias when using discretisation of  $U$  (especially when  $m$  is small).

### 7.2. FitzHugh-Nagumo model

The stochastic FitzHugh-Nagumo model for spike generation in squid axons is based on a two dimensional diffusion process with drift and diffusion coefficient parametrised as

$$b(x) = \begin{bmatrix} \vartheta_1(-x_1^3 + x_1 - x_2 + 1/2) \\ \vartheta_2 x_1 - x_2 + \vartheta_3 \end{bmatrix} \quad \sigma = \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix}.$$

The first coordinate  $X^{(1)}$  represents the axon membrane potential and  $X^{(2)}$  is a recovery variable. Parameter estimation for the FitzHugh-Nagumo model is discussed in Jensen et al. (2012) and extensively in the work of Jensen (2014). In this example we consider three type of proposals: the modified diffusion bridge (which is of Delyon-Hu type with  $\lambda = 0$ ), the modified diffusion bridge with random-walk type updates on the innovations and guided-proposals with random-walk type updates on the innovations. In both cases we took  $\rho = 0.5$  in equation (3.8). We used time-change guided proposals as in (5.5) with  $\tilde{\sigma}$  con-

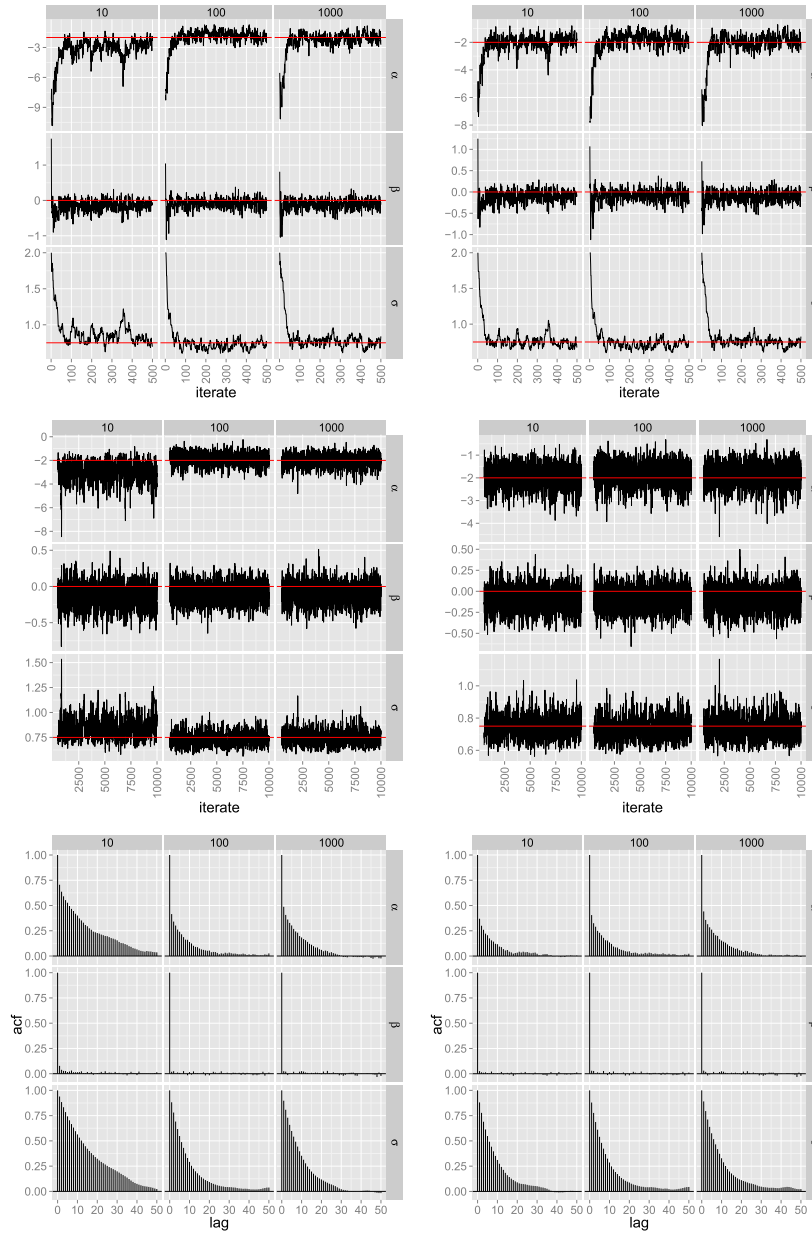


FIG 2. Panels comparing different numbers of imputed points ( $m = 10, 100, 1000$ ). Left: without time change. Right: with time change. Top: first 500 iterates. Middle: iterates 501-10.000. Bottom: ACF-plots based on iterates 501-10.000.

stant,  $\tilde{B} \equiv 0$  and  $\tilde{\beta}$  as in equation (4.10). This is a simple default choice. We discretise (5.7) as follows: suppose the current iterate is  $U_s$ . We have

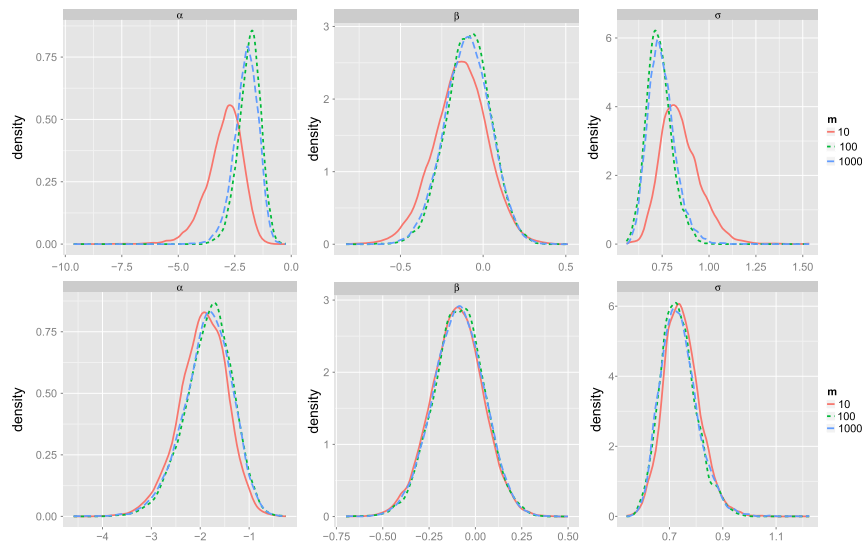


FIG 3. Kernel density estimates of the parameters based on iterates 501-10.000. Top: non time-changed. Bottom: time-changed. The bias from using only a small number of imputed points ( $m = 10$ , red curve) is clearly smaller for the time changed process. For a high number of imputed points ( $m = 1000$ , blue dashed curve) both methods agree.

$$dU_s = \frac{ds}{T} (\tilde{b}(\tau(s)) - b(\tau(s), \Gamma(s, U_s))) + dR_s \tag{7.1}$$

with  $dR_s = (T - s)^{-1} (I - 2aJ(s)) U_s ds - \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T-s}} \sigma dW_s$  (where we have used the relation  $\dot{v}(s) = \tilde{\beta}(s)$ ). Define

$$\bar{R}(h) = (T - s)^{-1} (I - 2aJ(s)) U_s - \sqrt{\frac{2}{T}} \frac{1}{\sqrt{T-s}} \sigma \frac{W_{s+h} - W_s}{h}.$$

To obtain an approximation  $u(s + h)$  for  $U_{s+h}$  we discretise the ordinary differential equation

$$du(s) = \left( \frac{1}{T} \tilde{b}(\tau(s)) - \frac{1}{T} b(\tau(s), \Gamma(s, u(s))) + \bar{R}(h) \right) ds, \quad u(s) = U_s$$

using the Runge-Kutta-4 method with step size  $h$ . We propose this discretisation scheme since by corollary 5.4,  $E[R_{s+h} - R_s | R_s] = hE[\bar{R}(h) | R_s]$ .

We simulated the process with parameters  $\vartheta_1 = 1.4$ ,  $\vartheta_2 = 1.5$ ,  $\vartheta_3 = 10$ ,  $\gamma_1 = 0.25$ ,  $\gamma_2 = 0.2$  on the time interval from 0 to  $T = 300$  using the Euler scheme with discretisation step 0.0004, starting in  $[0 \ 1]'$ , retaining 400 equidistant observations and the starting point. With these parameters this process presents a challenging estimation problem due to the strong nonlinear dynamics in the drift.

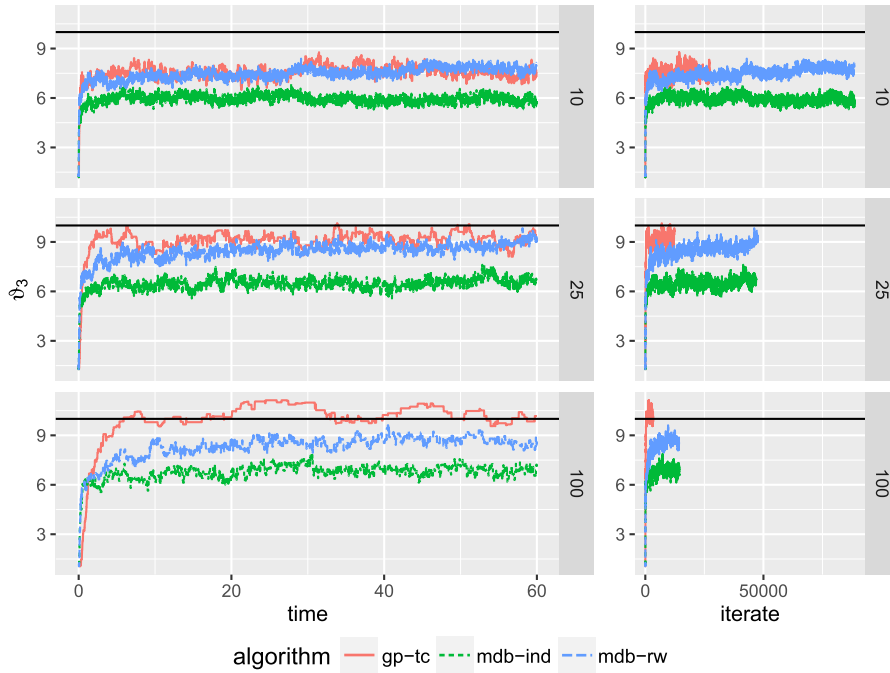


FIG 4. Trace-plots for  $\vartheta_3$ . Left: with respect to computing time in minutes. Right: with respect to iterate number. The three different panels correspond to  $m = 10$ ,  $m = 25$  and  $m = 100$ . “gp-tc” for guided proposals time-changed; “mdb-rw” for random-walk type modified diffusion bridges; “mdb-ind” for modified diffusion bridges.

We chose independent centred Gaussian priors with variance 50 for the parameters  $\vartheta_1, \vartheta_2, \vartheta_3$  and a product  $\text{InvGamma}(0.002, 0.002)$  prior on  $(\gamma_1^2, \gamma_2^2)$ .

We used Metropolis-Hastings steps for updating  $\gamma_i$  by setting  $\log \gamma_i^\circ = \log \gamma_i + 0.02Z_i$  ( $i = 1, 2$ ), where  $Z_i \sim N(0, 1)$ . For  $j = 1, 2, 3$  we took  $\vartheta_j^\circ = \vartheta_j + \nu_j Y_j$ , with  $Y_j$  independent Uniform random variables on  $[-1, 1]$ ,  $\nu_1 = \nu_2 = 0.03$  and  $\nu_3 = 0.15$ .

We estimated the joint posterior of the unobserved path and parameters  $\vartheta, \gamma$  using algorithm 1.

We ran the algorithm for the three different proposals with  $m = 10$  and  $m = 25$  and  $m = 100$ . Each simulation was stopped after 1 hour. The simulations were done on a computer equipped with 4 core Xeon CPU clocked at 3.40GHz with 30 GiB memory. In figure 4 trace-plots with respect to both computing time and iterate number for  $\vartheta_3$  are shown for the three samplers when  $m \in \{10, 25, 100\}$ . While iterates for the guided-proposals are more costly, the algorithm with these proposals does reach the stationary region way faster than the two variants of the modified diffusion bridge, especially when  $m = 100$ . However, solely examining trace-plots for the parameters can be misleading as illustrated by figure 5. Here, we plotted the average acceptance probability for



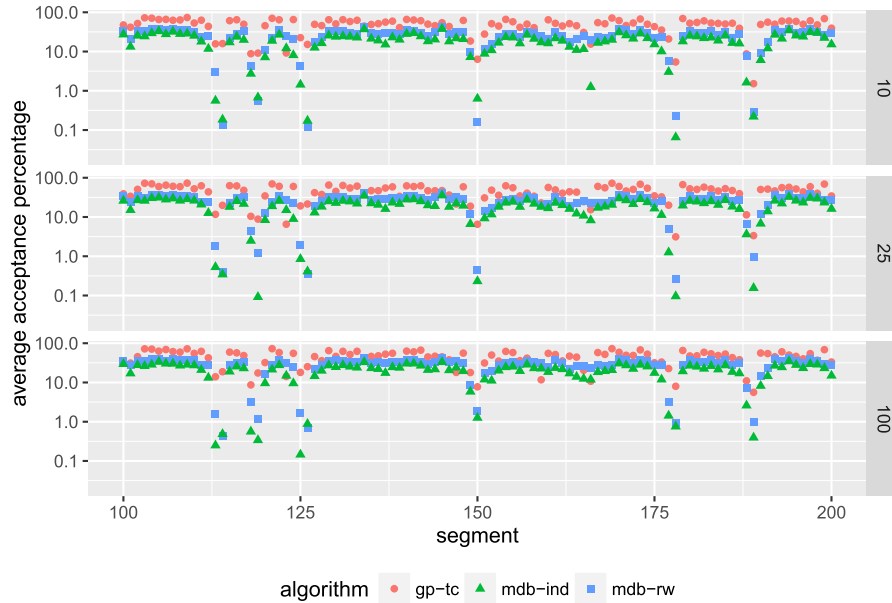


FIG 5. Average acceptance percentages for proposed bridges on segments 100 up to 200. The three different panels correspond to  $m = 10$ ,  $m = 25$  and  $m = 100$ .

bridge proposals (on a log10-scale) for the segments in between the 99-th and 200-th observations (the picture is representative for all segments). At certain segments the acceptance probabilities differ by several magnitudes. These segments correspond precisely to observations during an excursion from the meta-stable region. In these excursions the diffusion path follows closely the strong nonlinear drift dynamics, unlike in the meta-stable region. Small acceptance probabilities manifest themselves in slow convergence of the chain.

In addition we ran the algorithm for a longer time (16 hours) with  $m = 200$ . In this case, we simply used the Euler-approximation for the time-changed guided proposals. Actually, the Runge-Kutta-4 method is due to the stiffness of the SDE and only necessary in the case of a few imputed points. Trace-plots for  $\vartheta_3$  and  $\gamma_1$  are shown in figure 6. The trace-plot of  $\gamma_1$  against iteration number clearly shows that the guided proposals chain mixes better. The posterior means obtained by these methods were then considered to be the “true” posterior mean. These were used in computing the error values in table 2.

## 8. Summary and discussion

In this paper we have shown that the bridge proposals of Delyon and Hu (2006) and Schauer, Van der Meulen and Van Zanten (2017) can be used in a data-augmentation algorithm that is derived in a continuous time setting. An advantage of the latter type of proposals is that the pulling term

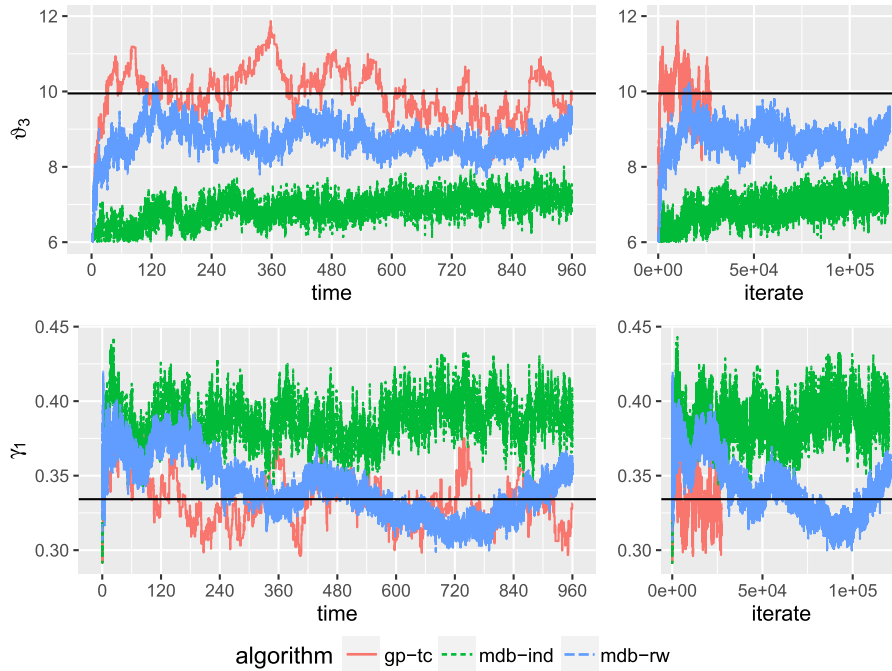


FIG 6. Trace-plots for  $\vartheta_3$  and  $\gamma_1$ . Left: with respect to computing time in minutes. Right: with respect to iterate number. In all cases  $m = 200$ . “gp-tc” is guided proposals time-changed; “mdb-rw” is random-walk type modified diffusion bridge; “mdb-ind” is modified diffusion bridge. The initial 735 iterates of  $\vartheta_3$  are smaller than 6 and not shown.

is adapted to the present value of  $\theta$  at that particular iteration. The innovations scheme allows for dependent Metropolis-Hastings updating of diffusion bridges.

Clearly, in any numerical implementation these bridge proposals need to be discretised. We have shown that this discretisation should be done carefully near the end point of the bridge. Instead of applying Euler discretisation directly on the bridge process, we advocate to first time change the bridge process, next scale it, and subsequently apply Euler discretisation.

All results have been derived under 2 assumptions: (i) the process is fully observed discretely in time; (ii) the diffusion is uniformly elliptic. In Van der Meulen and Schauer (2016) it is shown how condition (i) can be relaxed to assuming that at each discrete time a known linear combination of the components of the diffusion is observed with error. Assumption (ii) implies that our methods as stated here do not apply to hypo-elliptic diffusions. However, preliminary results indicate that the guided proposals from Schauer, Van der Meulen and Van Zanten (2017) can be used in this case as well, albeit with some more carefully chosen auxiliary process  $\tilde{X}$  in their construction. This is part of ongoing research.

TABLE 2

Estimation results FitzHugh-Nagumo model based on all iterates (no burn-in was considered). For each parameter we report its relative error with respect to the posterior mean obtained from the 16-hour run with  $m = 200$  imputed points using the time-changed guided proposals. Furthermore, we report the Root Relative Squared Error (RRSE) and the multivariate Effective Sample Size (mESS) from Vats, Flegal and Jones (2015).  $K$  is the total number of iterations executed within one hour.

m = 10	RRSE	$\vartheta_1$	$\vartheta_2$	$\vartheta_3$	$\gamma_1$	$\gamma_2$	mESS	K
mdb-rw	0.3	0.96	0.95	0.75	0.94	0.94	598.4	88485
gp-tc	0.29	1.03	1.04	0.76	1.02	0.99	333.67	27366
mdb-ind	0.49	0.9	0.88	0.6	1.08	1.05	761.25	88685
m = 25	RRSE	$\vartheta_1$	$\vartheta_2$	$\vartheta_3$	$\gamma_1$	$\gamma_2$	mESS	K
mdb-rw	0.18	0.95	0.94	0.85	1.06	0.96	439.12	47576
gp-tc	0.1	1	1	0.91	1.05	0.99	177.78	12359
mdb-ind	0.42	0.91	0.89	0.65	1.16	1.09	505.77	47006
m = 100	RRSE	$\vartheta_1$	$\vartheta_2$	$\vartheta_3$	$\gamma_1$	$\gamma_2$	mESS	K
mdb-rw	0.2	0.94	0.93	0.84	1.09	0.99	216.48	14259
gp-tc	0.01	0.99	0.98	1.01	1.08	1.01	85.47	3365
mdb-ind	0.39	0.89	0.87	0.68	1.19	1.1	246.4	14441

## Appendix A: Proof of Lemma 5.2

For ease of notation we will write  $\tau$  and  $\dot{\tau}$  instead of  $\tau(s)$  and  $\dot{\tau}(s)$ . If  $X$  satisfies the SDE

$$dX_s = b(s, X_s) ds + \sigma(s, X_s) d\widetilde{W}_s$$

and we are given a smooth function  $\tau = \tau(s)$ ,  $\tau: [0, T] \rightarrow \mathbb{R}_+$  with positive derivative  $\dot{\tau}$ , then

$$dX_\tau = \dot{\tau}b(\tau, X_\tau) ds + \sqrt{\dot{\tau}}\sigma(\tau, X_\tau) dW_s, \quad (\text{A.1})$$

where  $W$  is a different Brownian motion on the same probability space as  $\widetilde{W}$ .

Applying this to  $X^\circ$  (defined in equation (3.3)) gives

$$dX_\tau^\circ = 2(1 - s/T)[b(\tau, X_\tau^\circ) + a(\tau, X_\tau^\circ)\tilde{r}(\tau, X_\tau^\circ)] ds + \sqrt{2(1 - s/T)}\sigma(\tau, X_\tau^\circ) dW_s$$

by Itô's formula

$$\begin{aligned} dU_s &= d\left(\frac{v(\tau) - X_\tau^\circ}{T - s}\right) \\ &= \frac{2}{T}\dot{v}(\tau) ds + \frac{v(\tau)}{(T - s)^2} ds - \frac{X_\tau^\circ}{(T - s)^2} ds \\ &\quad - \frac{2}{T}[b(\tau, X_\tau^\circ) + a(\tau, X_\tau^\circ)\tilde{r}(\tau, X_\tau^\circ)] ds - \frac{\sqrt{2/T}}{\sqrt{T - s}}\sigma(\tau, X_\tau^\circ) dW_s \\ &= \frac{2}{T}\dot{v}(\tau) ds + \frac{U_s}{T - s} ds - \frac{2}{T}[b(\tau, X_\tau^\circ) + a(\tau, X_\tau^\circ)\tilde{r}(\tau, X_\tau^\circ)] ds \end{aligned}$$

$$-\frac{\sqrt{2/T}}{\sqrt{T-s}}\sigma(\tau, X_\tau^\circ) dW_s.$$

By equation (5.2),

$$\tilde{r}(\tau, X_\tau^\circ) = \tilde{H}(\tau)(v(\tau) - X_\tau^\circ) = \tilde{H}(\tau)(T - s)U_s = J(s)\frac{T}{T-s}U_s.$$

The result now follows from substituting this expression and using the relation  $X_{\tau(s)}^\circ = \Gamma(s, U_s)$  (see (5.6)).

The statement on  $J$  is a consequence of  $\tilde{a}(s)J(s) = \tilde{H}(\tau(s))(T - \tau(s))$  together with  $\lim_{t \rightarrow T} \tilde{H}(t)(T - t) = \tilde{a}^{-1}$  (see (Schauer, Van der Meulen and Van Zanten, 2017, Lemma 8)).

The expression for the integral follows upon the substitution  $s := \tau(s)$  and using relation (5.2).

**Appendix B: Proof of Theorem 5.3**

*Proof.* By straightforward calculus, the process  $U$  satisfies a SDE with drift coefficient

$$m\dot{v}(\tau)\dot{\tau} + \dot{m}m^{-1}U_s - m\dot{\tau} \left[ b(\tau, \Gamma) + a(\tau, \Gamma)\tilde{H}(\tau)m^{-1}U_s \right]$$

and diffusion coefficient as given in the theorem. We can rewrite the drift coefficient using specific properties of  $m$ .

For the first term in the drift, note that

$$\tilde{b}(\tau, \Gamma) = \tilde{B}(\tau) (v(\tau) - m^{-1}U_s) + \tilde{\beta}(\tau) = \dot{v}(\tau) - \tilde{B}(\tau)m^{-1}U_s,$$

where we have used the relation  $\dot{v}(s) = \tilde{B}(s)v(s) + \tilde{\beta}(s)$  at the second equality. Multiplying by  $m\dot{\tau}$  we get

$$m\dot{\tau}\dot{v}(\tau) = m\dot{\tau}\tilde{b}(\tau, \Gamma) + m\dot{\tau}\tilde{B}(\tau)m^{-1}U_s. \tag{B.1}$$

Next, we rewrite the second term appearing in the drift. Using

$$dA^{-1}/dt = -A^{-1} (dA/dt) A^{-1}$$

for an invertible matrix  $A$ , we obtain that

$$\dot{m}m^{-1} = -(T - s)^{-1}I - \dot{\tau} \left( \tilde{B}(\tau) - \tilde{a}(\tau)\tilde{H}(\tau) \right). \tag{B.2}$$

Substituting (B.1) and (B.2) into the drift and reordering terms shows that the drift of  $U$  equals

$$\begin{aligned} &-(T - s)^{-1}U_s - m\dot{\tau} \left( b(\tau, \Gamma) - \tilde{b}(\tau, \Gamma) \right) \\ &+ m\dot{\tau} \left( \tilde{B}(\tau) - a(\tau, \Gamma)\tilde{H}(\tau) \right) m^{-1}U_s - \dot{\tau} \left( \tilde{B}(\tau) - \tilde{a}(\tau)\tilde{H}(\tau) \right) U_s. \end{aligned}$$

This can be simplified to the form given in the theorem by using that

$$\tilde{B}(\tau) - \tilde{a}(\tau)\tilde{H}(\tau) = m \left( \tilde{B}(\tau) - \tilde{a}(\tau)\tilde{H}(\tau) \right) m^{-1}$$

which follows from the defining relation of  $F^*$ . □

## Acknowledgement

This work was partly supported by the Netherlands Organisation for Scientific Research (NWO) under the research programme “Foundations of nonparametric Bayes procedures”, 639.033.110 and by the ERC Advanced Grant “Bayesian Statistics in Infinite Dimensions”, 320637.

## References

- BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 333–382. With discussions and a reply by the authors. [MR2278331](#)
- BESKOS, A., ROBERTS, G., STUART, A. and VOSS, J. (2008). MCMC Methods for diffusion bridges. *Stochastics and Dynamics* **08** 319–350. [MR2444507](#)
- BEZANSON, J., KARPINSKI, S., SHAH, V. B. and EDELMAN, A. (2012). Julia: A Fast Dynamic Language for Technical Computing. *CoRR* [abs/1209.5145](#). [MR3605826](#)
- BLADT, M. and SØRENSEN, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli* **20** 645–675. [MR3178513](#)
- BLADT, M. and SØRENSEN, M. (2015). Simulation of multivariate diffusion bridges. *To appear in Journal of the Royal Statistical Society, series B*. [MR3454200](#)
- CHIB, S., PITT, M. K. and SHEPHARD, N. (2004). Likelihood based inference for diffusion driven models Economics Papers No. 2004-W20, Economics Group, Nuffield College, University of Oxford.
- CLARK, J. M. C. (1990). The simulation of pinned diffusions. In *Decision and Control, 1990., Proceedings of the 29th IEEE Conference on* 1418–1420. IEEE.
- COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statist. Sci.* **28** 424–446. [MR3135540](#)
- DELYON, B. and HU, Y. (2006). Simulation of conditioned diffusion and application to parameter estimation. *Stochastic Processes and their Applications* **116** 1660–1675. [MR2269221](#)
- DURHAM, G. B. and GALLANT, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** 297–338. With comments and a reply by the authors. [MR1939904](#)
- ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. [MR1839375](#) (2002e:62078)
- ERAKER, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19** 177–191. [MR1939708](#) (2003j:60107)

- FUCHS, C. (2013). *Inference for diffusion processes*. Springer, Heidelberg  
With applications in life sciences, With a foreword by Ludwig Fahrmeir.  
[MR3015023](#)
- GOLIGHTLY, A. and WILKINSON, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52** 1674–1693. [MR2422763](#)
- GOLIGHTLY, A. and WILKINSON, D. J. (2010). *Learning and Inference in Computational Systems Biology* Markov chain Monte Carlo algorithms for SDE parameter estimation, 253–276. MIT Press.
- GUGUSHVILI, S. and SPREIJ, P. (2012). Parametric inference for stochastic differential equations: a smooth and match approach. *ALEA Lat. Am. J. Probab. Math. Stat.* **9** 609–635. [MR3069378](#)
- GYÖNGY, I. (1998). A note on Euler’s approximations. *Potential Anal.* **8** 205–216. [MR1625576](#)
- JENSEN, C. ANDERS (2014). *Statistical Inference for Partially Observed Diffusion Processes*. Ph.d. Thesis University of Copenhagen.
- JENSEN, A. C., DITLEVSEN, S., KESSLER, M. and PAPASPILIOPOULOS, O. (2012). Markov chain Monte Carlo approach to parameter estimation in the FitzHugh-Nagumo model. *Phys. Rev. E* **86** 041114.
- KHASHMINSKII, R. Z. and KLEBANER, F. C. (2001). Long term behavior of solutions of the Lotka-Volterra system under small random perturbations. *Ann. Appl. Probab.* **11** 952–963. [MR1865029 \(2002h:34106\)](#)
- KÜCHLER, U. and SØRENSEN, M. (1997). *Exponential families of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York. [MR1458891 \(98k:62150\)](#)
- LIN, M., CHEN, R. and MYKLAND, P. (2010). On generating Monte Carlo samples of continuous diffusion bridges. *J. Amer. Statist. Assoc.* **105** 820–838. [MR2724864 \(2012b:62109\)](#)
- NEAL, R. M. (1999). Regression and classification using Gaussian process priors. In *Bayesian statistics, 6 (Alcoceber, 1998)* 475–501. Oxford Univ. Press, New York. [MR1723510](#)
- PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2003). Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian statistics, 7 (Tenerife, 2002)* 307–326. Oxford Univ. Press, New York With a discussion by Alan E. Gelfand, Ole F. Christensen and Darren J. Wilkinson, and a reply by the authors. [MR2003180](#)
- PAPASPILIOPOULOS, O. and ROBERTS, G. (2012). Importance sampling techniques for estimation of diffusion models. In *Statistical Methods for Stochastic Differential Equations. Monographs on Statistics and Applied Probability* 311–337. Chapman and Hall. [MR2976985](#)
- PAPASPILIOPOULOS, O., ROBERTS, G. O. and STRAMER, O. (2013). Data Augmentation for Diffusions. *J. Comput. Graph. Statist.* **22** 665–688. [MR3173736](#)
- PEDERSEN, A. R. (1995). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1** 257–279. [MR1363541 \(97d:62188\)](#)

- ROBERTS, G. O. and STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* **88** 603–621. [MR1859397 \(2002g:62122\)](#)
- ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov processes, and martingales. Vol. 2. Cambridge Mathematical Library*. Cambridge University Press, Cambridge. Itô calculus, Reprint of the second (1994) edition. [MR1780932 \(2001g:60189\)](#)
- ROSENTHAL, J. S. (2011). *Handbook of Markov Chain Monte Carlo (Chapman & Hall/CRC Handbooks of Modern Statistical Methods)*, 1 ed. Optimal proposal distributions and adaptive MCMC. Chapman and Hall/CRC. [MR2858446](#)
- SCHAUER, M. R., VAN DER MEULEN, F. H. and VAN ZANTEN, J. H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23** 2917–2950. [MR3648050](#)
- SERMAIDIS, G., PAPASPILIOPOULOS, O., ROBERTS, G. O., BESKOS, A. and FEARNHEAD, P. (2013). Markov chain Monte Carlo for exact inference for diffusions. *Scand. J. Stat.* **40** 294–321. [MR3066416](#)
- SØRENSEN, H. (2004). Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey. *Internat. Statist. Rev.* **72** 337–354.
- STEINER, A. and GANDER, M. J. (1999). Parametrische Lösungen der Räuber-Beute-Gleichungen im Vergleich. *Il Volterrano* **7** 32–44.
- STRAMER, O. and BOGNAR, M. (2011). Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach. *Bayesian Anal.* **6** 231–258. [MR2806243 \(2012g:60233\)](#)
- TIERNEY, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. [MR1620401](#)
- VAN DER MEULEN, F. H., SCHAUER, M. and VAN ZANTEN, J. H. (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.* **71** 615–632. [MR3131993](#)
- VAN DER MEULEN, F. H. and SCHAUER, M. R. (2016). Bayesian estimation of incompletely observed diffusions. *ArXiv e-prints*.
- VAN DER MEULEN, F. H. and VAN ZANTEN, J. H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. [MR3019485](#)
- VAN ZANTEN, J. H. (2013). Nonparametric Bayesian methods for one-dimensional diffusion models. *Mathematical biosciences* **243** 215–222. [MR3065207](#)
- VATS, D., FLEGAL, J. M. and JONES, G. L. (2015). Multivariate Output Analysis for Markov chain Monte Carlo. *ArXiv e-prints*.
- WHITAKER, G. A., GOLIGHTLY, A., BOYS, R. J. and SHERLOCK, C. (2017). Improved bridge constructs for stochastic differential equations. *Statistics and Computing* **27** 885–900. [MR3627552](#)